Journal of Financial Econometrics Advance Access published August 22, 2007

Journal of Financial Econometrics, 2007, 1–35

Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation

GREGORY R. DUFFEE AND RICHARD H. STANTON
Haas School of Business, University of California at Berkeley

ABSTRACT
The high persistence of interest rates has important implications for the preferred method used to estimate term structure models. We study the finite-sample properties of two standard dynamic simulation methods—efficient method of moments (EMM) and indirect inference—when they are applied to an first order autoregressive (AR[1]) process with Gaussian innovations. When simulated data are as persistent as interest rates, the finite-sample properties of EMM differ both from their asymptotic properties and from the finite-sample properties of indirect inference and maximum likelihood. EMM produces larger confidence bounds than indirect inference and maximum likelihood, yet is much less likely to contain the true parameter value. This is primarily because the population variance of the data plays a much larger role in the EMM conditions than in the moment conditions for either indirect inference or maximum likelihood. These results suggest that, under Gaussian assumptions, indirect inference (if practical) is preferable to EMM when working with persistent data such as interest rates. EMM's emphasis on the population variance strongly enforces stationarity on the underlying process, so this same reasoning suggests that EMM may be preferable in settings where stability and stationarity are important and difficult to impose.

KEYWORDS: AR process, unit root, EMM, indirect inference

Term structure estimation poses serious econometric challenges. The models are highly parameterized—empirical work often uses models with more than 30 free parameters. In addition, in all but the simplest settings, the likelihood

We thank Eric Renault, George Tauchen (the editor) and Hao Zhou for helpful discussions, and two anonymous referees for helpful comments. We are grateful for financial support from the Fisher Center for Real Estate and Urban Economics. Address correspondence to Richard H. Stanton, Haas School of Business, U.C. Berkeley, 545 Student Services Building #1900 Berkeley, CA 94720-1900, or
e-mail: stanton@haas.berkeley.edu.

doi:10.1093/jjfinec/nbm014

functions of bond yields are intractable. Thus, models are often estimated using dynamic simulation methods, such as the method of simulated scores (including efficient method of moments, or EMM), indirect inference, or simulation-based approximations to maximum likelihood.

Although dynamic simulation, and in particular EMM, has been successfully applied to a variety of economic models, Monte Carlo evidence casts some doubt on the finite-sample performance of EMM in a few specific term structure settings, such as the square-root diffusion model studied in Zhou (2001) and two-factor affine models studied in Duffee and Stanton (2004). Unfortunately, we do not know how much weight to put on this evidence. Finite-sample properties have been calculated only for models that are simultaneously both too complicated and too simplistic. Their complex structures make it impossible to determine which model features are difficult for EMM to capture, yet practical term structure estimation focuses on even more sophisticated (and thus more highly parameterized) models. Because we do not know which features of the relatively simplistic models create difficulties, we do not know if they carry over to models of greater practical relevance.

Evaluation of the finite-sample properties of dynamic simulation methods thus requires the study of either much simpler or much more complicated models. This paper takes the first approach, focusing on the effects of the high persistence exhibited by interest rates. We ask how accurately EMM and indirect inference estimate a highly persistent Gaussian first order autoregressive (AR[1]) process, compared with a benchmark of maximum likelihood (ML).[1] We find that, for reasonable sample sizes, statistical inference using EMM with confidence intervals based on the estimation criterion function is less reliable than using either indirect inference or maximum likelihood.

This evidence needs to be tempered by the fact that our setting is so simple. In more complicated models, EMM is likely to be much more tractable than either ML or indirect inference, and it may have better small sample properties due to its heavy penalty for deviations from stationarity. Moreover, EMM has been updated as a Bayesian method based on a Laplace likelihood (Gallant and Tauchen, 2007). It is known from Sims and Uhlig (1991) that, under a Bayesian implementation, near unit roots do not cause problems with the criterion function. We defer comparison of Bayesian-EMM to other methods in more complicated models to future research.

Our conclusion is surprising for two reasons. First, Gallant and Tauchen (1996) note that, if the probability distribution of the data implied by an auxiliary model is close to the distribution implied by the structural model, then simulation-based estimates should be close to those obtained using maximum likelihood.[2] Second,

[1] Somewhat conflicting definitions of these techniques can be found in the literature. We defer a precise description of what we mean by EMM and indirect inference to Section 1.

[2] Previous Monte Carlo evidence generally finds that, in many practical settings, the finite-sample performance of these estimators compares favorably to tractable alternatives. Relevant research includes Andersen et al. (1999), Chumacero (1997, 2001), Michaelides and Ng (2000), and Monfardini (1998).

the methodologies that underlie EMM and indirect inference are similar. EMM matches the score vector of an auxiliary model, while indirect inference matches its parameters.

The simple structure of the AR(1) model allows us to identify analytically key features of EMM that drive these results. The most important feature is that when the data are highly autocorrelated, the EMM criterion function is much more asymmetric than both the corresponding full ML criterion function (which is also asymmetric) and the indirect inference criterion function (which is symmetric). The intuition underlying the extreme asymmetry is straightforward. Given the parameters of an auxiliary model, the scale of the score vector depends on the variability of the data. This is a general property of auxiliary models. The EMM estimator evaluates the variability of the data using the data's implied population variance. For persistence parameters close to one, the population variance is an extremely asymmetric function of the persistence parameter. By contrast, the indirect inference estimator uses the data's sample variance instead of the implied population variance, eliminating the asymmetry.

To illustrate our results, assume that an econometrician has 1000 weekly observations of an AR(1) process. The true half-life of a shock is six years, which is similar to the persistence of Treasury bond yields in the postwar period. Using a conditional Gaussian auxiliary model, the median length of the EMM-based 95% confidence interval percent confidence interval for the autoregression coefficient is more than 20 times larger than the median length of the ML-based confidence interval. Yet, even though the EMM confidence intervals are large, they are much less likely to contain the true coefficient. With these 95th percentile bounds, the empirical rejection rate of truth is about 10% for ML and about 75% for EMM. In contrast, indirect inference generally performs as well as, or better than, maximum likelihood. For this example, the empirical rejection rate for indirect inference using the same auxiliary model is almost identical to the asymptotic rejection rate of five percent.

Although the simplicity of our setting allows us to draw analytic inferences about the estimators, it also necessarily limits the generality of our conclusions. We point out specific limitations at various points in the paper. Two are worth highlighting here. First, we specify the auxiliary models based on theoretical considerations. In practice, these specifications are often based on a data-driven search for an accurate in-sample fit, such as a semi-nonparametric (SNP) specification search. Our approach allows us to disentangle the finite-sample properties of a dynamic simulation from the finite-sample properties of the search for an appropriate auxiliary model. But it also prevents us from studying the combined properties of a specification search and dynamic simulation.

Second, in many practical problems, extreme asymmetry of the EMM criterion function in the neighborhood of nonstationarity can be an advantage of the EMM estimator. For example, as discussed by Tauchen (1998), asymmetry effectively forces the parameter estimates to lie in the stationary region. In an AR(1) setting, the econometrician knows the stationary region and can impose it directly on the parameter space (as we do). This is also a property of standard affine term structure models, thus from our perspective it is an attractive feature of our setting. But for econometric problems where stability of the process is desirable but analytically intractable to impose directly on the parameter space, our results may paint a biased picture of the attractiveness of EMM.

The next section describes the data-generating process studied throughout the paper. It also reviews the estimation techniques. Sections 2 and 3 describe the finite-sample properties of estimating the autoregression coefficient of an AR(1) process. The estimation techniques examined in Section 2 use the full likelihood function, both in ML estimation and as the auxiliary model, while the techniques examined in Section 3 use more highly parameterized auxiliary models. Concluding remarks are offered in Section 4.

1 OVERVIEW OF THE ANALYSIS

This section describes the data-generating process examined in this paper, sets up notation, and briefly reviews the relevant estimation techniques.

1.1 The Data-Generating Process

The true data generating process is

$$y_t = \rho_0 y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0,1), \tag{1}$$

where $|\rho_0| < 1$. Interest rate data are often studied at high frequencies relative to, say, data on macroeconomic performance. Accordingly, we interpret (1) as a process generating weekly observations of y_t. Tauchen (1998) uses the same formal process to investigate some features of EMM estimation, although his focus differs from ours.

The values of ρ_0 we consider are motivated by the properties of interest rates. We use three different values. The smallest is $\rho_0 = 0.8522$. This is a fairly low level of persistence. With weekly observations, this choice implies that shocks have a half life of only one month. Thus the estimation properties for this value are a good benchmark with which to compare the properties for highly persistent processes. The second value of ρ_0 is 0.9868, which corresponds to a half-life of one year. The most persistent process uses $\rho = 0.9978$, which corresponds to a half-life of six years. This value is chosen based on the behavior of the five-year Treasury bond yield over the sample period 1960 through 2003.[3]

The observed data are a sequence of interest rates $Y_T \equiv \{y_1, \ldots, y_T\}$. Again, we use three different values of T. The sample sizes are 1000, 2000, and 10,000 observations. With weekly data, these values correspond to about 19 years, 38 years, and 192 years respectively. The shortest period is similar to those used to study the behavior of interest rates after the disinflation of the early 1980s. Research that spans multiple interest rate regimes often use samples that span 40

[3] An AR(1) regression using monthly observations of the CRSP five-year bond yield from 1960:1 through 2003:12 produces an estimate of 0.98974, which corresponds to a half life of 5.6 years.

Table 1 Power of Dickey-Fuller test for nonstationarity.

	ρ_0		
T	0.8522	0.9868	0.9978
1000	1	0.936	0.157
2000	1	1	0.315
10000	1	1	1

This table summarizes results from 5,000 Monte Carlo simulations. The true data-generating process is

$$y_t = \rho_0 y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0,1)$$

For each simulation a sample of length T is generated. A Dickey-Fuller test (case 1) of the hypothesis that $\rho = 1$ is performed. This table reports the fraction of simulations for which the null hypothesis of nonstationarity is rejected at the five percent level.

or 50 years, although the data are typically observed at a lower frequency than weekly.

1.2 The Estimated Model

The econometrician estimates the true process (1), but with an unknown parameter ρ replacing ρ_0:

$$y_t = \rho y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0,1). \tag{2}$$

The econometrician follows the dynamic term structure literature by assuming that interest rates are stationary. Although statistical tests often cannot reject the hypothesis of a unit root, stationarity is almost universally imposed in the empirical literature on dynamic term structure models. The assumption is motivated in part by its theoretical plausibility, as noted by Clarida et al. (2000). It is also a convenient restriction to impose in these models because the parameter restrictions that correspond to stationarity are straightforward. Here, of course, the restriction is simply $|\rho| < 1$. This restriction maintained on the admissible parameter space distinguishes our approach from that of Tauchen (1998).

For many of the combinations of true persistence parameters ρ_0 and sample sizes T that we consider, stationarity is easily verified statistically. Table 1 summarizes Monte Carlo simulations that illustrate this point. Samples of length T are generated, then a Dickey and Fuller (1979) test is constructed to test the null hypothesis of a unit root in the data-generating process. The Dickey-Fuller test appropriate for (2) is their Case 1, which does not include a constant term. The table reports the fraction of simulations for which the null is rejected at the five percent level. Stationarity is statistically ambiguous only for $\rho_0 = 0.9978$ with the sample sizes $T = 1000$ or $T = 2000$.

Estimation is performed with ML, EMM and indirect inference. The remainder of this section reviews these techniques.

1.3 Full Maximum Likelihood

We use the full (i.e., unconditional) likelihood function. Denote the mean log-likelihood of Y_T given a candidate parameter ρ as $Q_T(Y_T, \rho)$. The ML estimate of ρ is

$$\hat{\rho}_{L,T} = \underset{\rho \in (-1,1)}{\operatorname{argmax}} Q_T(Y_T, \rho). \tag{3}$$

Denote the derivative of Q_T with respect to ρ as $h_T(Y_T, \rho)$. This is the mean score vector (which here has a single element). ML test statistics considered in this paper are based on the mean outer product of the score, denoted $\hat{S}_T$. The Generalized Method of Moments (GMM) criterion function associated with the ML problem is

$$J_{L,T}(\rho) = T h_T (Y_T, \rho)' \hat{S}_T^{-1} h_T (Y_T, \rho) . \tag{4}$$

The null hypothesis $\rho = \rho_0$ can be tested using this criterion function. Asymptotically,

$$J_{L,T}(\rho_0) \xrightarrow{d} \chi^2(1). \tag{5}$$

This is the standard score test in ML. It is also the GMM version of the likelihood ratio test.

1.4 Auxiliary Likelihoods

Both simulated scores and indirect inference estimation choose ρ to make sample properties of an auxiliary function close to the expected properties of the auxiliary function, where the expected properties are calculated using ρ. The econometrician chooses the auxiliary function to be both tractable and to fit the important dynamics of the data. With simulated scores, the auxiliary function is a likelihood function that is typically easier to work with than the true likelihood. Although indirect inference does not emphasize the use of likelihood functions as auxiliary functions, for comparability with simulated scores we restrict the focus of this paper to auxiliary functions that are likelihoods.

We use auxiliary likelihoods associated with two auxiliary models. The first is the structural model (2), replacing the structural parameter ρ with the auxiliary parameter β. For this model we use the full likelihood as the auxiliary likelihood.

The second auxiliary model is the more general AR(1) model

$$y_t = \beta_0 + \beta_1 y_{t-1} + \xi_t, \quad \xi_t \sim N(0, \beta_2), \quad \beta = (\beta_0\ \beta_1\ \beta_2)'. \tag{6}$$

For this model we use the likelihood conditioned on the first observation.

For notational convenience, the parameter vector for each model is a vector β, with length $p_a = 1$ for the first auxiliary model and $p_a = 3$ for the second. For a given auxiliary model, denote the mean auxiliary log-likelihood of Y_T given β by

$Q_{a,T}(Y_T, \beta)$. Denote the mean score vector by $h_{a,T}(Y_T, \beta)$. The estimate of β is

$$\hat{\beta}_T = \underset{\beta \in B}{\operatorname{argmax}}\, Q_{a,T}(Y_T; \beta). \tag{7}$$

Denote by $\hat{d}_{a,T}$ the second derivative of the auxiliary log-likelihood at this estimate.

As T gets large, the auxiliary likelihood no longer depends on the particular sample Y_T. It is simply a function of the true structural parameter ρ_0,

$$\lim_{T \to \infty} Q_{a,T}(Y_T; \beta) = Q_{a,\infty}(\rho_0, \beta). \tag{8}$$

In this limit, the auxiliary parameters that maximize the auxiliary likelihood converge to β_0, defined as

$$\beta_0 = \underset{\beta \in B}{\operatorname{argmax}}\, Q_{a,\infty}(\rho_0, \beta). \tag{9}$$

Similarly, $\hat{d}_{a,T}$ converges to $d_{a,0}$.

Denote the covariance of the mean score vector evaluated at β_0 (and scaled by T) by Ω_T. Formally,

$$\Omega_T(Y_T) = T \text{ Var}\left(h_{a,T}(Y_T, \beta_0)\right). \tag{10}$$

As T gets large, this converges to a fixed matrix Ω. Then, from the Central Limit Theorem, the probability density of $\hat{\beta}_T$ converges in distribution to

$$\sqrt{T}\left(\hat{\beta}_T - \beta_0\right) \xrightarrow{d} N\left(0, d_{a,0}^{-1} \Omega d_{a,0}^{-1}\right). \tag{11}$$

In (11), $d_{a,0}$ and Ω are unknown. The former can be replaced with $\hat{d}_{a,T}$, which is a consistent estimate of $d_{a,0}$. Similarly, Ω can be replaced with a consistent estimator. Throughout this paper we use the mean outer product of the auxiliary score vector to estimate Ω. The outer product is a consistent estimator of Ω if the auxiliary score vector is asymptotically serially uncorrelated. Both the auxiliary likelihoods we consider nest the true likelihood function, thus this property of the score vector is satisfied.

1.5 EMM

EMM is developed in Bansal et al. (1993, 1995) and Gallant and Tauchen (1996). Our implementation differs from the standard procedure, so we need to explain our approach precisely, and how it differs from the standard procedure.

The standard implementation of EMM is a two-step process. First, a data-driven search process chooses a reduced-form auxiliary model that best fits the data (by an appropriate metric). The reduced-form model is often in the semi-nonparametric family. Second, the structural model is estimated using the method of simulated scores, which chooses the parameter vector to set the expectation of the auxiliary likelihood score vector as close to zero as possible. Gallant and Tauchen (1996) use the term EMM to describe the method of simulated scores when the search process asymptotically produces an auxiliary likelihood that converges to the true likelihood.

We do not perform the first step. Instead, we use pre-specified auxiliary models that nest the true model. This choice is motivated by two goals. First, we want to understand the finite-sample properties of the simulated score procedure without introducing complications associated with the finite-sample properties of the auxiliary model search. Second, we want to give simulation-based estimators the greatest opportunity to succeed. Poor performance on the part of an estimator cannot be blamed on a poorly-chosen auxiliary likelihood. We use the term EMM because our auxiliary likelihoods converge to the true likelihood, as in the standard implementation of EMM. Nonetheless, the finite-sample properties of our procedure will certainly differ from those of the two-step procedure, and it is conceivable that properties of the latter procedure are better than those of the former.

A minor difference between the standard implementation and our procedure concerns the type of auxiliary likelihood. In the former case, the auxiliary likelihood function is a conditional likelihood function. Because we want to allow for the use of both full and conditional auxiliary likelihoods, our description of the simulated score methodology differs slightly from that in Gallant and Tauchen (1996). Denote the expectation of $h_{a,T}(Y_T, \beta)$ as

$$H_{a,T}(\rho, \beta) \equiv E\left(h_{a,T}(Y_T(\rho), \beta)\right). \tag{12}$$

This expectation is taken over the density of Y_T, thus it depends on ρ. In practice, $H_{a,T}$ is computed by simulating the true data-generating process using ρ, unless the combination of auxiliary model and true process admits an analytic solution. The data generating process studied in this paper is sufficiently tractable that analytic solutions are typically available.

The subscript T on $H_{a,T}$ is necessary only if a full auxiliary likelihood is used. In this case, the expectation of the auxiliary score of the first (and unconditional) observation differs from the expectation of the auxiliary score of all other (conditional) observations. Therefore the expectation of the mean score depends on the number of conditional observations relative to the single unconditional observation. If a conditional auxiliary likelihood is used, all observations have the same expected score, hence $H_{a,T}$ does not depend on T.

The GMM criterion function associated with EMM estimation is

$$J_{E,T}(\rho) = T H_{a,T}\left(\rho, \hat{\beta}_T\right)' \hat{S}_{a,T}^{-1} H_{a,T}\left(\rho, \hat{\beta}_T\right). \tag{13}$$

In (13), $\hat{S}_{a,T}$ plays the role of a consistent estimator of Ω. The EMM estimate of ρ, denoted by $\hat{\rho}_{E,T}$, minimizes this function:

$$\hat{\rho}_{E,T} = \underset{\rho \in (-1,1)}{\operatorname{argmin}} J_{E,T}(\rho). \tag{14}$$

The parameter vector is overidentified if the length of β exceeds one. Then asymptotically,

$$J_{E,T}(\hat{\rho}_{E,T}) \xrightarrow{d} \chi^2(p_a - 1). \quad (15)$$

This is a test of general model misspecification. The null hypothesis $\rho = \rho_0$ can be tested with

$$J_{E,T}(\rho_0) - J_{E,T}(\hat{\rho}_{E,T}) \xrightarrow{d} \chi^2(1). \quad (16)$$

This is the EMM counterpart of the ML test (5).

An interesting special case is when the auxiliary likelihood is identical to the true likelihood. Then by the consistency property of ML, the EMM estimate $\hat{\rho}_{E,T}$ equals the ML estimate $\hat{\rho}_{L,T}$ (and also equals $\hat{\beta}_T$). Equality of $\hat{\rho}_{L,T}$ and $\hat{\rho}_{E,T}$ does not correspond to equality of the ML and EMM likelihood ratio statistics (5) and (16). In other words, the criterion functions (4) and (13) are guaranteed to equal each other only at the ML estimate of ρ, where both functions equal zero.

1.6 Indirect Inference

Indirect inference is developed in Smith (1990, 1993), Gouriéroux et al. (1993), and Gouriéroux and Monfort (1996).[4] The setting here is more restrictive because only likelihood functions are considered. Estimation chooses ρ to minimize the difference between the sample estimate of the auxiliary model's parameters and the expected parameter estimates. This difference can be expressed in terms of the binding function

$$b(\rho) = \underset{\beta \in B}{\operatorname{argmax}}\, Q_{a,\infty}(\rho, \beta). \quad (17)$$

Asymptotically, the auxiliary model estimate $\hat{\beta}_T$ approaches the value of the binding function evaluated at ρ_0. More precisely,

$$T\left(\hat{\beta}_T - b(\rho_0)\right)' d_{a,0}\Omega^{-1} d_{a,0}\left(\hat{\beta}_T - b(\rho_0)\right) \xrightarrow{a} \chi^2(p_a). \quad (18)$$

This result suggests an estimator for ρ,

$$\hat{\rho}_{I,T} = \underset{\rho \in (-1,1)}{\operatorname{argmin}}\, J_{I,T}(\rho), \quad (19)$$

where the criterion function $J_{I,T}$ is

$$J_{I,T}(\rho) = T\left(\hat{\beta}_T - b(\rho)\right)' \hat{d}_{a,T}\hat{S}_{a,T}^{-1}\hat{d}_{a,T}\left(\hat{\beta}_T - b(\rho)\right). \quad (20)$$

In (20), the unknown parameters $d_{a,0}$ and Ω are replaced with consistent estimates. Gouriéroux and Monfort (1996) prove that indirect inference is asymptotically

[4] Although EMM is one of several estimators called ''indirect inference'' by Gouriéroux et al. (1993), we follow standard usage by distinguishing between EMM and indirect inference.

equivalent to EMM (when the same auxiliary function is used, as it is here), thus it necessarily achieves the same asymptotic efficiency.

For the purposes of comparing the finite-sample properties of indirect inference with those of EMM, note that we can interpret the criterion function (20) as a quadratic form in a vector of moments $\hat{d}_{a,T}(\hat{\beta}_T - b(\rho))$. The weighting matrix is the inverse of $\hat{S}_{a,T}$, which is the same weighting matrix used with EMM estimation. Thus comparing the properties of $H_{a,T}(\rho, \hat{\beta}_T)$ with those of $\hat{d}_{a,T}(\hat{\beta}_T - b(\rho))$ will help us understand the differences in the finite-sample properties of the estimators.

As with EMM, we can construct a test of model misspecification and a test of the hypothesis that $\rho = \rho_0$. Asymptotically,

$$J_{I,T}(\hat{\rho}_{I,T}) \stackrel{d}{\to} \chi^2(p_a - 1), \tag{21}$$

and

$$J_{I,T}(\rho_0) - J_{I,T}(\hat{\rho}_{I,T}) \stackrel{d}{\to} \chi^2(1). \tag{22}$$

The difficulty with indirect inference is in computing the binding function $b(\rho)$. In some applications it is known analytically. One case is where the auxiliary likelihood is identical to the true likelihood. Then, by the consistency property of ML, $b(\rho) = \rho$. The indirect inference estimator is then $\hat{\rho}_{I,T} = \hat{\beta}_T$. In the absence of an analytic expression of the binding function, it is approximated with simulations. The indirect inference literature uses two alternative simulation approaches, one of which requires a slight modification to the derivation above. The existing literature uses the term indirect inference to refer to both approaches, but we modify this term to help distinguish between them.

The first approach, which we call *asymptotic indirect inference*, defines a simulated version of the binding function,

$$b^*(\tilde{Y}_\tau(\rho)) = \underset{\beta \in B}{\text{argmax}}\, Q_{a,\tau}(\tilde{Y}_\tau(\rho), \beta), \tag{23}$$

where $\tilde{Y}_\tau(\rho)$ is a length-τ simulated sequence of data generated by the true model with parameter ρ. As τ approaches infinity, this simulated binding function approaches $b(\rho)$. Thus we also refer to the case where $b(\rho)$ has an analytic expression as asymptotic indirect inference.

The second approach, which we call *finite-sample indirect inference*, defines a finite-sample version of the binding function. The function is

$$b_T^H(\tilde{Y}_T^1(\rho), \ldots, \tilde{Y}_T^H(\rho)) = \frac{1}{H} \sum_{h=1}^{H} \underset{\beta \in B}{\text{argmax}}\, Q_{a,T}(\tilde{Y}_T^h(\rho), \lambda), \tag{24}$$

where $Y_T^h(\rho)$ is a length-T simulated sequence of data generated by the true model with parameter ρ. The sequences are independent. To simplify notation, denote this function by $\tilde{b}_T^H(\rho)$. As H approaches infinity, this function no longer depends on the particular randomly generated samples, so we write

$$\lim_{H\to\infty} \tilde{b}_T^H(\rho) = b_T(\rho). \tag{25}$$

For finite H, the criterion function must be modified to account for the randomness introduced by simulating the binding function. The adjusted criterion function is

$$J_{F,T}(\rho) = T\frac{1}{1+1/H}\left(\hat{\beta}_T - \tilde{b}_T^H(\rho)\right)' \hat{d}_{a,T}\hat{S}_{a,T}^{-1}\hat{d}_{a,T}\left(\hat{\beta}_T - \tilde{b}_T^H(\rho)\right). \tag{26}$$

The subscript F refers to the finite-sample version of indirect inference. The finite-sample indirect inference estimator is

$$\hat{\rho}_{F,T} = \underset{\rho\in(-1,1)}{\operatorname{argmin}} J_{F,T}(\rho). \tag{27}$$

The test statistics (21) and (22) also apply to finite-sample indirect inference, with $J_{F,T}$ replacing $J_{I,T}$.

Note that finite-sample indirect inference does not require a large value of H. For example, $H = 1$ is valid. Higher values of H reduce the uncertainty in the estimate of ρ because the variance of $\hat{\beta}_T - \tilde{b}_T^H(\rho_0)$ is proportional to $(1 + 1/H)$.

1.7 Tradeoffs Among These Techniques

Gouriéroux et al. (1993) show that EMM and the two indirect inference estimators described above all have the same asymptotic properties. However, they differ markedly in their computational needs. Both indirect inference estimators require the auxiliary model to be reestimated for every set of simulated data. In contrast, with EMM, the parameters of the auxiliary model need to be estimated only once, on the original sample. For each simulated dataset, EMM requires only the computation of the auxiliary model's score vector at the original set of auxiliary parameter values. This speed advantage is the primary reason why researchers use EMM more often than they use either indirect inference estimator.

There is also a speed tradeoff with the two forms of the indirect inference estimator. The asymptotic indirect inference estimator requires only a single estimation of the auxiliary model's parameters for a given structural-model parameter vector. The finite-sample version requires estimation of H sets of auxiliary model parameters.[5]

Advocates of finite-sample indirect inference focus on the bias in point estimates. The use of simulated datasets of the same length as the original sample means that finite-sample indirect inference corrects some of the small-sample bias present in both EMM and asymptotic indirect inference. Indeed, Gouriéroux et al. (2000) show that, as H approaches infinity, the finite-sample indirect inference estimator provides the same second-order bias correction in the point estimates

[5] For the AR(1) model we examine, asymptotic indirect inference has an even larger speed advantage, because there is an analytic expression for the asymptotic indirect inference binding function.

as the bootstrap estimator of Efron (1979). While this result does not hold for finite H (unlike with a finite number of bootstrap replications), Gouriéroux et al. (2000) nevertheless show in simulations that the finite-sample indirect inference estimator reduces bias in point estimates as well as the exact median-unbiased estimator of Andrews (1993) for AR(1) models, and as well as the approximate median-unbiased estimators of Rudebusch (1992) and Andrews and Chen (1994) for AR(p) models.

In the simulations that follow we confirm the bias reduction associated with finite-sample indirect inference. However, statistical inference depends on both point estimates and estimated standard errors. One of the surprising conclusions of this paper is that the small-sample properties of test statistics produced by finite-sample indirect inference are generally inferior to those produced by asymptotic indirect inference.

2 ESTIMATION USING THE FULL LIKELIHOOD FUNCTION

This section focuses on estimation of ρ in (2) using the full likelihood function, both directly in ML estimation and indirectly as the auxiliary likelihood function. Although the calculations are straightforward, we go into a little detail to illustrate the differences among the estimation techniques.

2.1 Setup

The mean score vector is

$$h_T(Y_T;\rho) = \frac{1}{T}\left[\rho\left(y_1^2 - (1-\rho^2)^{-1}\right) + (T-1)\left(\overline{y_t y_{t-1}} - \rho\overline{y_{t-1}^2}\right)\right]. \tag{28}$$

The ML estimator, $\hat{\rho}_{L,T}$, sets $h_T(Y_T;\rho)$ to zero. We solve for this estimate using the Matlab root-finder "fzero." The outer product of the score vector is

$$\hat{S}_T = \frac{1}{T}\left[\hat{\rho}_{L,T}^2\left(y_1^2 - (1-\hat{\rho}_{L,T}^2)^{-1}\right)^2 + \sum_{t=2}^{T}\left(y_t y_{t-1} - \hat{\rho}_{L,T} y_{t-1}^2\right)^2\right]. \tag{29}$$

The second derivative of the mean log-likelihood is

$$\hat{d}_T = \frac{1}{T}\left[y_1^2 - \left(1-\hat{\rho}_{L,T}^2\right)^{-2}\left(1+\hat{\rho}_{L,T}^2\right) - (T-1)\overline{y_{t-1}^2}\right]. \tag{30}$$

Note that the unconditional variance of y_t conditional on a candidate ρ is

$$\text{Var}(y_t|\rho) = E(y_t^2|\rho) = (1-\rho^2)^{-1}. \tag{31}$$

The auxiliary likelihood is the full likelihood of the true model, with the parameter ρ replaced by β. The expectation of $h_{a,T}$ is thus identical to the expectation of h_T and is given by

$$H_{a,T}(\rho, \beta) = \frac{1}{T}\left[\beta\left(\left(1-\rho^2\right)^{-1} - \left(1-\beta^2\right)^{-1}\right) + (T-1)(\rho-\beta)(1-\rho^2)^{-1}\right]. \tag{32}$$

Inspection of (32) reveals that it is identically zero for $\rho = \beta$. The asymptotic indirect inference binding function is $b(\rho) = \rho$. Therefore the estimate of ρ for both of these estimators is $\hat{\rho}_{L,T}$. The finite-sample indirect inference binding function does not have an analytic expression. We approximate it with $H = 4$. This choice is sufficient to illustrate the advantages and disadvantages of finite-sample indirect inference. The Matlab root-finder ''fzero'' is used to find the finite-sample indirect inference estimate of ρ.

2.2 Criterion Functions and Confidence Intervals

For this setup, the GMM criterion functions for ML, EMM, asymptotic indirect inference, and finite-sample indirect inference are, respectively,

$$J_{L,T}(\rho) = Th_T(Y_T;\rho)^2/\hat{S}_T; \tag{33}$$

$$J_{E,T}(\rho) = TH_{a,T}(\rho;\hat{\rho}_{L,T})^2/\hat{S}_T; \tag{34}$$

$$J_{I,T}(\rho) = T(\hat{d}_T(\hat{\rho}_{L,T} - \rho))^2/\hat{S}_T; \tag{35}$$

and

$$J_{F,T}(\rho) = T\frac{1}{1+1/4}(\hat{d}_T(\hat{\rho}_{L,T} - \tilde{b}^4_T(\rho)))^2/\hat{S}_T. \tag{36}$$

We follow the recommendation of Gallant and Tauchen (1998), and construct confidence intervals for ρ by inverting the criterion functions.[6] Hansen et al. (1996) conclude, based on Monte Carlo studies of various GMM estimators, that Wald-type confidence intervals typically perform poorly. They find that confidence intervals based on inverting GMM criterion functions are more reliable.

Define c_ζ as the probability-ζ critical value of a $\chi^2(1)$ distribution:

$$\text{Prob}(\chi^2(1) > c_\zeta) = 1 - \zeta. \tag{37}$$

For example, with $\zeta = 0.95$, the probability is 5% that a random variable distributed as $\chi^2(1)$ exceeds c_ζ. A confidence interval for ρ based on maximum likelihood is

$$\zeta\text{-level ML confidence interval} = \left\{\rho : J_{L,T}(\rho) < c_\zeta\right\}. \tag{38}$$

6 In particular, we do not calculate confidence intervals from asymptotic standard errors, including those calculated, as in recent implementations of EMM, using the MCMC methodology of Chernozhukov and Hong (2003).

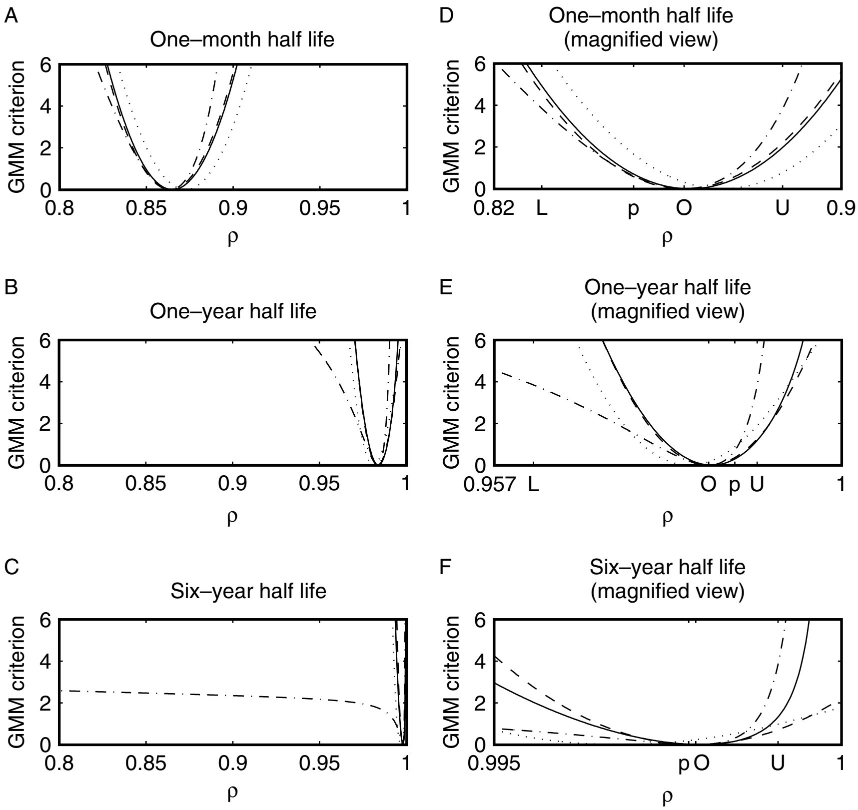

Figure 1 Criterion functions for estimation of an autoregression coefficient.

Confidence intervals for the other estimation methods are constructed similarly. Because these intervals are constructed from criterion functions instead of inverting asymptotic standard errors, they are not necessarily symmetric around the ML point estimate.[7] In principle, this asymmetry is a nice feature of the confidence intervals, because it reflects the cost (in terms of the likelihood) of values of ρ near the boundary of stationarity. However, one of the main problems we document with the EMM criterion function is that it is excessively asymmetric when the data are highly persistent.

Examples of the criterion functions and associated confidence intervals are displayed in Figure 1. We generate three samples of 1000 weekly observations for the different values of ρ_0 discussed in Section 1.1.[8] Panels A, B, and C contain criterion functions for $\rho_0 = 0.8522$ (one-month half life of shocks), $\rho_0 = 0.9868$ (one

[7] The intervals are not necessarily symmetric around the point estimate even if the criterion function is symmetric, because the intervals do not extend beyond the boundary of the stationary region $|\rho| < 1$.

[8] The samples are constructed with the same underlying random numbers, generated using the Matlab random number generator randn. The seed for the sequence is zero.

year), and $\rho_0 = 0.9978$ (six years) respectively. Panels D, E, and F are magnified views of the corresponding panel on the left. Criterion functions for full ML, asymptotic indirect inference, finite-sample indirect inference, and EMM are illustrated with solid lines, dashed lines, dotted lines, and dashed-dotted lines respectively. In the magnified panels, the point estimate of ρ is marked with an "O" and the true value is marked with a "p". The points "L" and "U" are the lower and upper 95th percentile bounds on ρ constructed using the EMM criterion function.

The criterion functions for ML, EMM, and asymptotic indirect inference all reach a minimum of zero at the same point—the ML estimate of ρ. The point that minimizes the finite-sample indirect inference criterion function is random because of the finite number of simulations used to construct $\tilde{b}_T^H(\rho)$.

2.3 The Asymmetry of the EMM Criterion Function

The most interesting feature of these plots is the asymmetry in the EMM criterion function relative to the other criterion functions. The relative asymmetry increases significantly as the persistence of the data increases. A detailed discussion of the functions for which analytic expressions are available (i.e., not finite-sample indirect inference) helps explain this pattern. Begin with the asymptotic indirect inference criterion (35). Inspection of this equation reveals the function is symmetric in $(\hat{\rho}_{L,T} - \rho)$, and in particular the function is quadratic. A comparison of the EMM criterion function (34) with (35) is simplified by ignoring the role of the first observation in the construction of both $\hat{d}_T$ and $H_{a,T}$. Then the two functions are approximately given by

$$J_{E,T}(\rho) \approx T\left((\hat{\rho}_{L,T} - \rho)(1 - \rho^2)^{-1}\right)^2 / \hat{S}_T \tag{39}$$

and

$$J_{I,T}(\rho) \approx T\left((\hat{\rho}_{L,T} - \rho)\overline{y_{t-1}^2}\right)^2 / \hat{S}_T. \tag{40}$$

Both functions scale the difference between the ML estimate and the candidate ρ by a measure of the variance of the data. For the EMM criterion function, the variance measure is the population variance (31). For the asymptotic indirect inference criterion function, the variance measure is the sample estimate. When $\rho > \hat{\rho}_{L,T}$, EMM scales the deviation $(\hat{\rho}_{L,T} - \rho)$ by a larger variance than when $\rho < \hat{\rho}_{L,T}$. This asymmetry is more pronounced when $\hat{\rho}_{L,T}$ is close to one, because (31) is more sensitive to ρ at such points.

A similar approximation illustrates the relation between the ML criterion function (33) and the EMM criterion function. The conditional maximum likelihood estimate of ρ (i.e., dropping the first observation) is $\overline{y_t y_{t-1}}/\overline{y_{t-1}^2}$. Ignoring the difference between this estimate and the full ML estimate, the ML criterion function is approximately

$$J_{L,T}(\rho) \approx \left[\frac{1}{T}\rho(y_1^2 - (1-\rho^2)^{-1}) + \frac{T-1}{T}(\hat{\rho}_{L,T} - \rho)\overline{y_{t-1}^2}\right]^2 / \hat{S}_T. \tag{41}$$

Aside from the treatment of the first observation, this criterion function is identical to (40). The population variance (31) appears in the derivative of the log-likelihood of the first observation. Therefore the ML criterion function is asymmetric for the same reason that the EMM criterion function is asymmetric. But because the population variance affects only one of the T observations, the magnitude of the asymmetry is smaller than in (39).

For the one-month half life, the asymmetry of the EMM criterion function is noticeable but small. In Panel D, the distance between the EMM lower 95th percentile bound L and the point estimate $\hat{\rho}_{L,T}$ is 0.0326. The distance between $\hat{\rho}_{L,T}$ and the EMM upper 95th percentile bound is 0.0223, or about two-thirds the length of the lower bound. The asymmetry is stronger with the one-year half life in Panel E, where the lower and upper distances between the bounds and $\hat{\rho}_{L,T}$ are 0.0213 and 0.0059, respectively. By contrast, the ML and asymptotic indirect inference criterion functions are almost indistinguishable in this panel. With a six-year half life, the asymmetry in the ML criterion function relative to the asymptotic indirect inference criterion function is clear (Panel F), but it is swamped by the asymmetry of the EMM criterion function. In fact, the lower bound of the EMM 95th percentile bound cannot be displayed in either Panel C or Panel F. It is $\rho = 0.38$.

When ρ is close to one, the population variance (31) is both large and highly nonlinear in ρ. The asymmetry in the criterion function is created by the latter effect, not the former. Although our data-generating process does not allow us to distinguish between these effects, consider the more general DGP

$$y_t = \rho_0 y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2). \tag{42}$$

The magnitude of the population variance depends on both ρ_0 and σ, yet σ has no effect on any of our results. To see this, define $y_t^* = y_t/\sigma$ and $\epsilon_t^* = \epsilon_t/\sigma$. Dividing (42) by σ produces dynamics for y_t^* that are identical to (1). Thus all calculations and conclusions regarding the estimation of ρ are unaltered aside from the substitution of y_t/σ for y_t.

Asymptotically, the asymmetry of the EMM criterion function has no implications for test statistics based on the criterion function because the ML estimate converges to ρ_0. In other words, all that matters is the behavior of the functions over the tiny range that encompasses both $\hat{\rho}_{L,T}$ and ρ_0. But in finite samples, the asymmetries can have important effects. For concreteness, consider the functions for the one-year half life displayed in Panels B and E. At the ML estimate, the second derivative of the EMM criterion function (34) is close to those of (33) and (35). (By construction, the latter two have identical second derivatives at the ML estimate). Their ratio is 1.2. However, the third derivative of the ML criterion function at $\hat{\rho}_{L,T}$ is much closer to zero than is the third derivative of the

EMM criterion function. To simplify the interpretation of the third derivatives, divide them by the corresponding second derivatives. With this scaling, the third derivative of the ML function (33) is 23, while the third derivative of the EMM function (34) is 362. Thus, although the true value of ρ is very close to the ML estimate (0.9868 versus 0.9836), the ratio of second derivative of the EMM function to the second derivative of the asymptotic indirect inference function at ρ_0 is 4.1. The corresponding ratio for the ML and asymptotic indirect inference functions is only 1.1.

In summary, the evidence in Figure 1 suggests that with highly persistent data, the finite-sample properties of EMM confidence bounds and test statistics are likely to differ substantially from the corresponding finite-sample properties of both ML and indirect inference. The next subsection uses Monte Carlo simulations to document this claim. It also shows that the finite-sample properties of ML and asymptotic indirect inference are much closer to their asymptotic properties than are the finite-sample properties of EMM.

2.4 Finite Sample Properties

Bias in estimated autocorrelation coefficients, particularly when the data are persistent, is a well-known problem for most estimation techniques, not just those that are simulation-based (see, for example, Orcutt (1948), Quenouille (1949), Marriott and Pope (1954), Kendall (1954), or, for a more recent example involving non-linear processes, Ball and Torous (1996)). Because our focus is on persistent processes, it is not a surprise that the techniques we consider exhibit deviations between finite-sample and asymptotic properties. Our goal is to examine how and why these deviations differ across estimation methods. Relevant evidence from Monte Carlo simulations is reported in Tables 2 and 3.

Table 2 reports means, medians, and root mean squared errors of point estimates for both maximum likelihood and finite-sample indirect inference. (The remainder of the table is discussed in Section 3.3.) The two conclusions to draw from these statistics are unsurprising. First, the ML estimate is downward biased. Second, the finite-sample indirect inference estimate is less biased than the ML estimate. These statements apply to each combination of ρ_0 and T.

Table 3 reports median lengths, across 5000 simulations, of 95% confidence intervals for ρ. This length is defined by (38) for ML and defined by similar equations for EMM and indirect inference. The table also reports empirical rejection rates for these bounds (i.e., for what fraction of the simulations is ρ_0 contained in the range defined in that simulation (38)).

We draw two important conclusions from the results of Table 3. The first concerns EMM and the second concerns finite-sample indirect inference.

2.4.1 Hypothesis testing with EMM. When the data-generating process is highly persistent, the finite-sample properties of EMM differ sharply from those of the other estimators. The EMM confidence intervals are much larger than the ML and indirect inference confidence intervals, yet the empirical rejection rates for

Table 2 Estimates of an autocorrelation coefficient.

T	ρ_0	ML	F-II using true like	EMM w/cond Gaussian	A-II w/cond Gaussian	F-II w/cond Gaussian	OLS
1000	0.8522	0.8504	0.8518	0.8467	0.8504	0.8536	0.8486
		(0.8513)	(0.8524)	(0.8475)	(0.8511)	(0.8543)	(0.8494)
		[0.0165]	[0.0184]	[0.0183]	[0.0166]	[0.0186]	[0.0171]
2000	0.8522	0.8515	0.8522	0.8496	0.8515	0.8531	0.8506
		(0.8517)	(0.8525)	(0.8500)	(0.8517)	(0.8535)	(0.8508)
		[0.0119]	[0.0133]	[0.0125]	[0.0120]	[0.0134]	[0.0121]
10000	0.8522	0.8520	0.8521	0.8516	0.8520	0.8523	0.8518
		(0.8521)	(0.8522)	(0.8518)	(0.8521)	(0.8524)	(0.8519)
		[0.0053]	[0.0060]	[0.0053]	[0.0053]	[0.0060]	[0.0053]
1000	0.9868	0.9849	0.9859	0.9733	0.9848	0.9883	0.9825
		(0.9859)	(0.9874)	(0.9809)	(0.9859)	(0.9893)	(0.9836)
		[0.0060]	[0.0062]	[0.0314]	[0.0062]	[0.0066]	[0.0053]
2000	0.9868	0.9858	0.9863	0.9827	0.9858	0.9876	0.9847
		(0.9863)	(0.9871)	(0.9841)	(0.9863)	(0.9881)	(0.9853)
		[0.0040]	[0.0043]	[0.0089]	[0.0041]	[0.0045]	[0.0047]
10000	0.9868	0.9866	0.9867	0.9862	0.9866	0.9870	0.9864
		(0.9867)	(0.9868)	(0.9863)	(0.9867)	(0.9870)	(0.9865)
		[0.0016]	[0.0018]	[0.0019]	[0.0017]	[0.0019]	[0.0017]
1000	0.9978	0.9962	0.9972	0.9451	0.9961	0.9975	0.9929
		(0.9972)	(0.9983)	(0.9781)	(0.9970)	(0.9982)	(0.9940)
		[0.0034]	[0.0031]	[0.0890]	[0.0037]	[0.0027]	[0.0069]
2000	0.9978	0.9969	0.9974	0.9748	0.9969	0.9980	0.9954
		(0.9974)	(0.9981)	(0.9930)	(0.9973)	(0.9984)	(0.9960)
		[0.0021]	[0.0020]	[0.0435]	[0.0022]	[0.0019]	[0.0034]
10000	0.9978	0.9976	0.9977	0.9966	0.9976	0.9980	0.9974
		(0.9977)	(0.9979)	(0.9972)	(0.9977)	(0.9981)	(0.9975)
		[0.0007]	[0.0008]	[0.0047]	[0.0008]	[0.0008]	[0.0009]

This table reports means, medians (in parentheses), and RMSEs (in brackets) from 5,000 Monte Carlo simulations. The true data-generating process is

$$y_t = \rho_0 y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0,1).$$

For each simulation, a sample of length T is generated. The parameter ρ is treated as unknown, and is estimated with full maximum likelihood (ML), efficient method of moments (EMM), asymptotic indirect inference (A-II), finite-sample indirect inference (F-II) where the binding function is approximated with four simulations, and OLS. The auxiliary likelihood used with EMM and indirect inference is either the true likelihood or a conditional Gaussian likelihood. The OLS regression includes a constant term.

Table 3 Hypothesis tests of an autoregression coefficient.

		Median length of 95% conf bound				Actual rejection rate of 95% conf bound			
T	ρ_0	ML	EMM	A-II	F-II	ML	EMM	A-II	F-II
1000	0.8522	0.0646	0.0670	0.0646	0.0725	0.051	0.059	0.051	0.049
2000	0.8522	0.0457	0.0466	0.0458	0.0512	0.055	0.059	0.056	0.055
10000	0.8522	0.0205	0.0206	0.0205	0.0229	0.052	0.052	0.052	0.055
1000	0.9868	0.0190	0.0383	0.0197	0.0217	0.058	0.145	0.058	0.079
2000	0.9868	0.0140	0.0193	0.0141	0.0158	0.056	0.108	0.057	0.068
10000	0.9868	0.0063	0.0067	0.0064	0.0071	0.051	0.066	0.051	0.054
1000	0.9978	0.0066	0.5671	0.0066	0.0055	0.106	0.256	0.087	0.182
2000	0.9978	0.0050	0.1629	0.0054	0.0050	0.075	0.213	0.071	0.129
10000	0.9978	0.0025	0.0038	0.0026	0.0029	0.058	0.121	0.059	0.067

This table summarizes results from 5,000 Monte Carlo simulations. The true data-generating process is

$$y_t = \rho_0 y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0,1).$$

For each simulation, a sample of length T is generated. The parameter ρ is treated as unknown, and is estimated with full maximum likelihood (ML). Confidence bounds and test statistics are based on GMM criterion functions for ML, EMM, asymptotic indirect inference (A-II), and finite-sample indirect inference (F-II) where the binding function is approximated with four simulations. The true full likelihood is the auxiliary likelihood for EMM and indirect inference. Confidence bounds are constructed by inverting the criterion functions. Tests of the hypothesis that $\rho = \rho_0$ are GMM versions of the likelihood ratio test. Under the null, all statistics have asymptotic $\chi^2(1)$ distributions.

EMM are much higher than those of the other estimators. For example, with 2000 weeks of data and $\rho_0 = 0.9978$, the median length of the EMM confidence interval is more than 30 times the median length of the ML or indirect inference confidence intervals. The empirical rejection rate at the 95th percentile level exceeds 20 percent for EMM, compared with between 7.5 and 7 percent for ML and asymptotic indirect inference. Differences between EMM and the other estimators are relatively large even with extremely long data samples. With 10,000 observations (192 years), the median length of the EMM confidence intervals is more than 50 percent larger than the median length of the ML and asymptotic indirect inference confidence intervals, and empirical rejection rates are twice as large.

The problem with EMM is inherent in the sharp asymmetry in the EMM criterion function induced by the role that the population variance function (31) plays in the EMM moment condition. This problem is explained easily in the context of Figure 1. With highly persistent data, the EMM criterion function is extremely asymmetric. This asymmetry produces a confidence interval that covers a wide range of ρ below $\hat{\rho}_{L,T}$, but covers too little of the range of ρ above $\hat{\rho}_{L,T}$. Given this asymmetry, it is not surprising that almost all of the EMM rejections of the hypothesis $\rho = \rho_0$ occur when the point estimate $\hat{\rho}_{L,T}$ is less than ρ_0. (This fact is not reported in any table.)

We emphasize that our point is not simply that EMM has poor finite-sample properties in this AR(1) setting. It is that EMM has poor properties relative to the alternative simulation technique of asymptotic indirect inference. In fact, the test statistic for indirect inference has better finite-sample properties than does the test statistic for ML. Moreover, EMM breaks down in cases where the hypothesis of a unit root is typically rejected. For example, with $\rho_0 = 0.9868$ and $T = 2000$, the empirical rejection rate of the hypothesis $\rho = \rho_0$ is almost twice as high with EMM as it is with ML and asymptotic indirect inference. Yet, as mentioned in Section 1.2, the power of a Dickey-Fuller test is extremely high for this combination of ρ_0 and T. (The test rejects the hypothesis of a unit root at the 95 percent confidence level in 4999 of 5000 simulations.) Thus the finite-sample properties of EMM are poor relative to asymptotic indirect inference when the underlying process is highly persistent, but not necessarily statistically close to a unit root.

2.4.2 Hypothesis testing with finite-sample indirect inference. When the data-generating process is highly persistent, asymptotic indirect inference outperforms finite-sample indirect inference in testing the hypothesis $\rho = \rho_0$. For example, with 1000 weeks of data and $\rho_0 = 0.9978$, the empirical rejection rates are 9 percent and 18 percent respectively. At first glance, this is a surprising result because finite-sample indirect inference point estimates of ρ are less biased than those of asymptotic indirect inference. We might be tempted to attribute this result to a low value of H, which creates uncertainty in the calculation of the finite-sample binding function. However, this poor performance is actually attributable to the smaller bias in the point estimate.

To understand how the smaller bias affects statistical inference, note that the indirect inference test statistics (35) and (36) are derived from the following asymptotic results:

$$\lim_{T\to\infty} \frac{\hat{\rho}_{L,T} - \rho_0}{\widehat{SE}(\rho_{L,T})} \xrightarrow{d} N(0,1) \tag{43}$$

and

$$\lim_{T\to\infty} \left(\frac{1}{1+1/H}\right)^{1/2} \frac{\hat{\rho}_{L,T} - \tilde{b}_T^H(\rho_0)}{\widehat{SE}(\hat{\rho}_{L,T})} \xrightarrow{d} N(0,1), \tag{44}$$

where the robust estimate of the standard error of $\hat{\rho}_{L,T}$ is

$$\widehat{SE}(\hat{\rho}_{L,T}) = \sqrt{\hat{S}_T/(T\hat{d}_T^2)}. \tag{45}$$

The statistics (35) and (36) are squared versions of (43) and (44). According to (43) and (44), ML estimates of ρ are asymptotically distributed symmetrically around ρ_0 or $\tilde{b}_T^H(\rho)$ respectively. (Asymptotically, $\tilde{b}_T^H(\rho)$ converges to ρ_0.) Hence if we use, say, a 95% confidence level for the $\chi^2(1)$ tests (35) and (36), then asymptotically

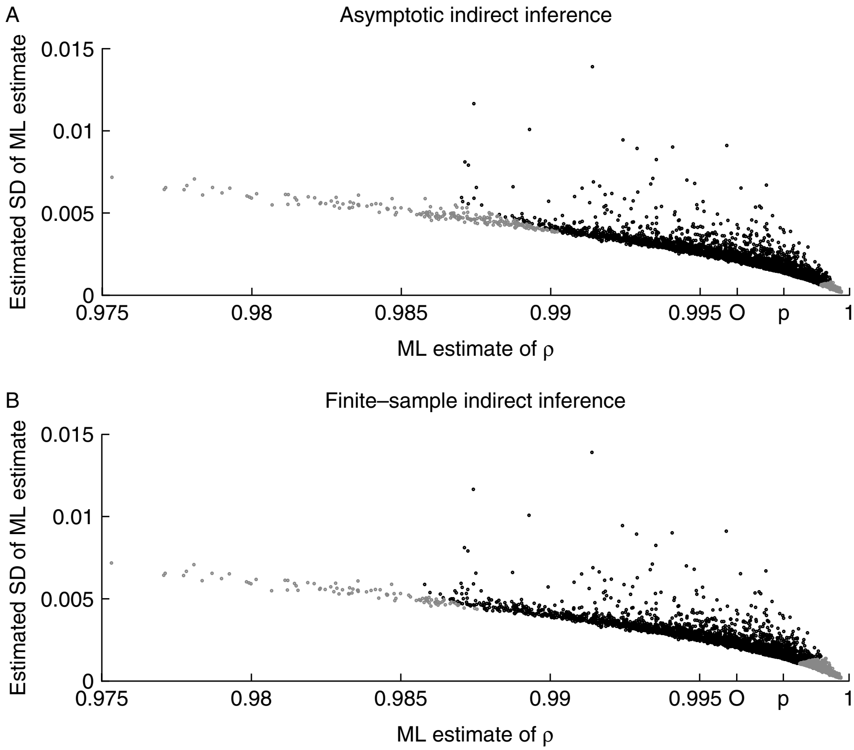

Figure 2 Hypothesis testing with indirect inference.

half of the rejections should occur when the estimate of ρ exceeds ρ_0 (or $\tilde{b}_T^H(\rho)$) and half should occur when the estimate of ρ is less than ρ_0 (or $\tilde{b}_T^H(\rho)$). Of course, in finite samples this symmetry need not be observed.

Visual evidence of this symmetry, or lack thereof, is in Figure 2. It is constructed using 5,000 simulations produced with the high-persistence parameter $\rho_0 = 0.9978$ and $T = 1000$. Both panels A and B display the same scatter plot of ML point estimates and the corresponding estimate of its standard error. The mean point estimate is marked with an "O" on the horizontal axis. The true value ρ_0 is marked with a "p." It is the square root of $(1 - \rho_0^2)/T$. For the points in Panel A labeled in grey, the asymptotic indirect inference test (35) rejects the hypothesis of $\rho = \rho_0$ at the 95 percent confidence level. The test is not rejected at the points in black.

Panel A illustrates three features of the ML estimates that affect the symmetry of the outcome of the asymptotic indirect inference test (35). First, the ML estimate is biased down; point "O" is less than "p." All else equal, this bias generates more rejections in the region $\hat{\rho}_T < \rho_0$ and fewer in the region $\hat{\rho}_T > \rho_0$, simply because the ML estimates are less likely to occur in the latter range. Second, the ML estimates are negatively skewed—the distribution has a long left tail. All else equal, this skewness also generates more rejections in the region $\hat{\rho}_T < \rho_0$, because the typical numerator in (43) is larger in magnitude in this region than in the region $\hat{\rho}_T > \rho_0$. Third, there is a strong inverse relation between the point estimate of ρ and its estimated standard error. All else equal, this relation generates more rejections in the region $\hat{\rho}_T > \rho_0$, because the denominator in (43) is lower in this region than in the region $\hat{\rho}_T < \rho_0$.

In combination, these three features produce empirical rejections of the hypothesis $\rho = \rho_0$ that are almost evenly divided between estimates of ρ that exceed ρ_0 and estimates that are less than ρ_0. Using a 95% critical value, these ranges contain 4.5% and 4.2% of the total number of simulations, for a total rejection rate of 8.7%.

Panel B contains similar evidence for finite-sample indirect inference. Rather than approximate the binding function at ρ_0 with $H = 4$, we can use the Monte Carlo simulations to approximate the function with $H = 5000$. The point labeled "O" is not only the mean ML point estimate; it is also $\tilde{b}_T^{5000}(\rho_0)$. Thus we can calculate the statistic (36) for $H = 5000$. For this test, the grey and black dots in Panel B have the same interpretation as the dots in Panel A. Rejections of the null hypothesis at the 95% critical value are no longer evenly divided. The upper and lower ranges contain 15.6% and 2.6% of the total number of simulations, for a total rejection rate of 18.2%.

The difference between Panels A and B is that in Panel B, the effect of the biased ML estimate is removed. The test statistic is based on the difference between the ML estimate and the expected ML estimate, not the difference between the ML estimate and ρ_0. As a consequence, the inverse relation between the point estimate of ρ and its estimated standard error dominates the distribution of the test statistic. Rejections thus increase in the region $\hat{\rho}_T > \rho_0$. Because the probability density of $\hat{\rho}_T$ is high in this region, the null hypothesis is rejected for a relatively large fraction of the total sample.

We do not know if the patterns in Panel A are a happy accident, or if there is something fundamental about the combination of the three features that affect the distribution of the asymptotic indirect inference test statistics. A more robust conclusion is that there is no theoretical reason to believe that finite-sample indirect inference produces more accurate test statistics than asymptotic indirect inference. The motivation for finite-sample indirect inference lies exclusively in correcting for bias in point estimates, not the size of statistical tests. In the setting here, correcting for biased point estimates has unfortunate consequences for the size of statistical tests.

3 ESTIMATION USING A CONDITIONAL GAUSSIAN LIKELIHOOD

Section 2 documents relatively poor finite-sample performance of statistical inference with EMM. The auxiliary likelihood used in that section−full ML−provides the cleanest way to explain differences among the criterion functions. However, it does not allow us to compare point estimation properties, because EMM and asymptotic indirect inference point estimates are identical to ML estimates. Nor does it permit an examination of the finite-sample properties of overidentifying restrictions. Thus in this section we use the auxiliary conditional likelihood of (6). We refer to this likelihood as the ''conditional Gaussian'' likelihood.

The conditional Gaussian likelihood is less precise than the full ML likelihood. Put differently, it is a more flexible, and thus more highly parameterized, likelihood. By using it, we move in the direction of auxiliary likelihoods more frequently used in practice. Econometricians use auxiliary likelihoods to circumvent the problem of an unknown or intractable true likelihood. Gallant and Tauchen (1996) and Bansal et al. (1995) argue in favor of a data-based approach to choosing the auxiliary likelihood in order to detect misspecification of the structural model. In practice, this means using a flexible functional form that can capture arbitrarily complicated dynamics (at least asymptotically). This flexibility typically results in an auxiliary likelihood that is more highly parameterized than is the true data-generating process. Gallant and Tauchen recommend the SNP family of conditional likelihood functions described in Gallant and Tauchen (1992). For the case of a scalar time series, the simplest SNP specification is a conditional Gaussian likelihood.

It is important to recognize, however, that we are not conducting a search for the best SNP specification. Instead, we use a common specification across all data samples. Although we use an auxiliary likelihood function that nests the true likelihood and is asymptotically equivalent to the true likelihood, it is not necessarily the best in-sample SNP specification for every sample.

3.1 Setup

The derivative of the mean auxiliary log-likelihood (6) is

$$h_{a,T}(Y_T,\beta) = \beta_2^{-1}\begin{pmatrix} e_t \\ e_t y_{t-1} \\ (1/2)(e_t^2/\beta_2 - 1) \end{pmatrix} \tag{46}$$

where

$$e_t = y_t - \beta_0 - \beta_1 y_{t-1}. \tag{47}$$

The optimal parameter vector $\hat{\beta}_T = (\hat{\beta}_{0,T}\,\hat{\beta}_{1,T}\,\hat{\beta}_{2,T})'$ is equivalent to the combination of the coefficients of an OLS regression of y_t on y_{t-1} and the mean of squared regression residuals. At this point, the Hessian matrix is

$$\hat{d}_{a,T} = -\hat{\beta}_{2,T}^{-1}\begin{pmatrix} 1 & \overline{y_{t-1}} & 0 \\ \overline{y_{t-1}} & \overline{y_{t-1}^2} & 0 \\ 0 & 0 & \frac{1}{2}\hat{\beta}_{2,T}^{-1} \end{pmatrix}. \tag{48}$$

The mean outer product of the score is

$$\hat{S}_{a,T} = \hat{\beta}_{2,T}^{-2} \begin{pmatrix} \hat{\beta}_{2,T} & \overline{e_t^2 y_{t-1}} & \frac{1}{2}\overline{e_t^3}\hat{\beta}_{2,T}^{-1} \\ \overline{e_t^2 y_{t-1}} & \overline{e_t^2 y_{t-1}^2} & \frac{1}{2}\left(\overline{e_t^3 y_{t-1}}\hat{\beta}_{2,T}^{-1} - \overline{e_t y_{t-1}}\right) \\ \frac{1}{2}\overline{e_t^3}\hat{\beta}_{2,T}^{-1} & \frac{1}{2}\left(\overline{e_t^3 y_{t-1}}\hat{\beta}_{2,T}^{-1} - \overline{e_t y_{t-1}}\right) & \frac{1}{4}(\overline{e_t^4}\hat{\beta}_{2,T}^{-2} - 1) \end{pmatrix}. \tag{49}$$

Both of these matrices asymptotically approach (aside from a sign change)

$$S_{a,0} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (1-\rho_0^2)^{-1} & 0 \\ 0 & 0 & 1/2 \end{pmatrix}. \tag{50}$$

Estimation of ρ with EMM uses the expectation of (46) conditional on ρ. Some algebra reveals that this expectation is

$$H_{a,T}(\rho, \beta) = \beta_2^{-1} \begin{pmatrix} -\beta_0 \\ (\rho - \beta_1)\mathrm{Var}(y|\rho) \\ \frac{1}{2}\left(\beta_2^{-1} - 1\right) + \frac{1}{2}\left((\rho-\beta_1)^2 \mathrm{Var}(y_t|\rho) + \beta_0^2\right)\beta_2^{-1} \end{pmatrix}. \tag{51}$$

The T subscript on $H_{a,T}$ is included for notational consistency, but it is unnecessary because the auxiliary likelihood is a conditional likelihood. Because the auxiliary likelihood is correctly specified, the consistency of maximum likelihood implies that the asymptotic indirect inference binding function is

$$b(\rho) = \begin{pmatrix} 0 \\ \rho \\ 1 \end{pmatrix}. \tag{52}$$

The finite-sample binding function is approximated with $H = 4$. With our choice of auxiliary model, estimation of the auxiliary parameters does not require numerical approximation, and computation of $\tilde{b}_T^H(\rho)$ is therefore tractable.

3.2 The Criterion Functions and Confidence Bounds

As in the case of full maximum likelihood, a closer look at the EMM and asymptotic indirect inference criterion functions helps us understand the differences between their performance. The EMM moment vector is (51) evaluated at $\hat{\beta}_T$. The moments are weighted by the inverse of (49). The corresponding asymptotic indirect inference moment vector is $\hat{d}_{a,T}(\hat{\beta}_T - b(\rho))$:

$$d_{a,T}(\hat{\beta}_T - b(\rho)) = \hat{\beta}_{2,T}^{-1} \begin{pmatrix} -\hat{\beta}_{0,T} + \overline{y_{t-1}}(\rho - \hat{\beta}_{1,T}) \\ (\rho - \hat{\beta}_{1,T})\overline{y_{t-1}^2} - \hat{\beta}_{0,T}\overline{y_{t-1}} \\ \frac{1}{2}(\hat{\beta}_{2,T}^{-1} - 1) \end{pmatrix}. \tag{53}$$

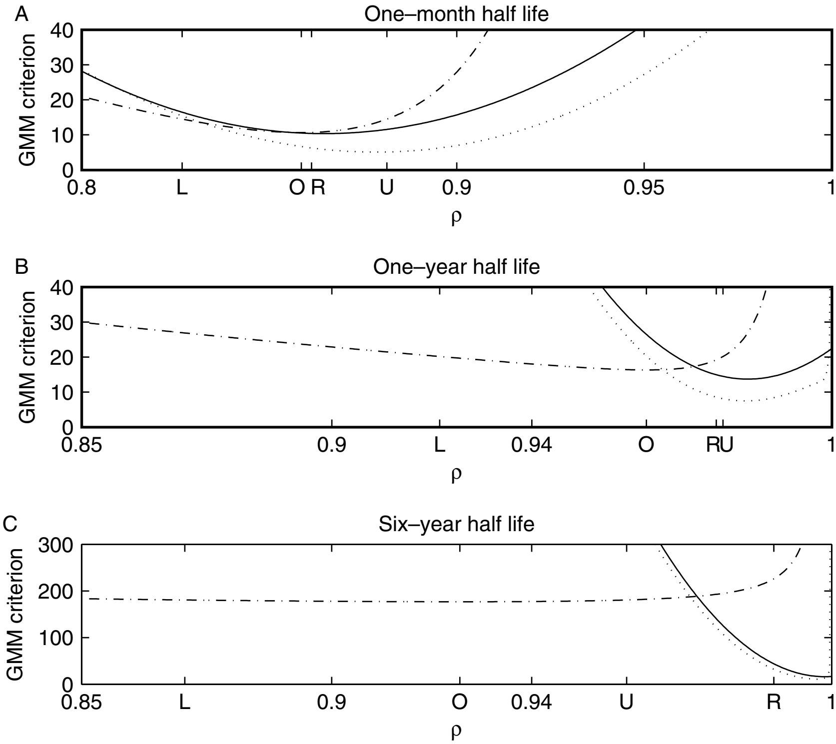

Figure 3 Criterion functions for estimation of an autocorrelation coefficient, using a conditional Gaussian auxiliary auxiliary likelihood.

These moments are also weighted by the inverse of (49). Each element of the vector (53) differs slightly from its counterpart in (51). The effects of these differences are illustrated in Figure 3.

The figure displays the criterion functions for EMM, asymptotic indirect inference, and finite-sample indirect inference for the 1000-observation samples previously examined in Figure 1. The points labeled ''O'' are the EMM point estimates and the points labeled ''L'' and ''U'' are the upper and lower 95th percentile bounds on these estimates. The points labeled ''R'' are the estimates of persistence from the auxiliary model ($\hat{\beta}_{2,T}$), which are also the OLS regression estimates. Efficient method of moments criterion functions are dotted-dashed lines, asymptotic indirect inference functions are solid lines, and finite-sample indirect inference functions (using four simulations to approximate the binding function) are dotted lines.

The most striking feature of this figure is that the EMM criterion functions in Panels B and C look nothing like the indirect inference criterion functions. Consider, for example, Panel C. For this highly persistent sample of data, the OLS estimate of ρ is 0.9884. The asymptotic indirect inference estimate is 0.9987, at which point the value of the corresponding criterion function is about 16. The finite-sample indirect inference estimate is 0.9971, with a criterion function value of about 11. By contrast, the EMM estimate is 0.9256, with a criterion function value that exceeds 175.

What accounts for the anomalous behavior of the EMM criterion function? One reason is identified in Section 2.3: the presence of the unconditional variance in the EMM moment condition, which is extremely sensitive to small variations in ρ in the neighborhood of $\rho = 1$. A second reason is that the sample mean has a large variance when the data are highly persistent. If the auxiliary model specifies the true mean, as it does in Section 2, this is irrelevant. But with a more general auxiliary model, the EMM criterion function is extremely sensitive to the sample mean.

This sensitivity cannot be seen merely by studying the EMM moment vector (51) because it is driven by the interaction of the moment vector and the weighting matrix. More precisely, the sample mean of the data is correlated with sample covariance between the first two EMM moments, with unfortunate effects. The details (which are unavoidably mind-numbing) follow.

First note that in any finite sample, the mean of y_t will differ from the true mean, which is zero in this setting. In addition, in most finite samples where the true process is stationary, the OLS regression estimate of mean reversion, $\hat{\beta}_{1,T}$, is less than one. As a consequence, the estimate of the constant term in the OLS regression, $\hat{\beta}_{0,T}$, is positively correlated with the sample mean of y_t.[9] Therefore the first moment of (51) is negatively correlated with the sample mean of y_t. Asymptotically, this moment has no effect on the estimate of ρ. The parameter does not appear in the equation for this first moment, and we can see in (50) that the asymptotic covariances between this moment and the other two moments are both zero.

However, the criterion function uses sample covariances, not asymptotic covariances. The sample covariance between the first two moments is, from (49):

$$\overline{e_t^2 y_{t-1}} = \widehat{\text{Cov}}(e_t^2, y_{t-1}) + \overline{e_t^2}\,\overline{y_{t-1}}. \tag{54}$$

Thus if the sample mean of y_t happens to be greater (less) than its true mean, the sample covariance will also tend to be greater (less) than its true mean. Hence the first moment tends to be positively correlated with the second moment whenever the first moment is negative, and negatively correlated whenever the first moment is positive. This pattern means that the EMM criterion function is typically minimized at a negative value of the second moment, which corresponds to an EMM estimate of ρ less than the OLS estimate $\hat{\beta}_{1,T}$. The effect holds regardless of the true value of ρ, which is why the EMM estimate of ρ is less than the OLS regression estimate in each panel of Figure 3. The effect is magnified for highly

[9] This is from $\hat{\beta}_{0,T} = \overline{y_t} - \hat{\beta}_{1,T}\overline{y_{t-1}}$; ignore the difference between the sample mean of y_t and the sample mean of y_{t-1}.

persistent processes because the variance of the sample mean is higher for higher values of ρ.

The structure of the asymptotic indirect inference moment vector (53) avoids this problem. The steps used to conclude that the EMM criterion function is typically minimized at a negative value of the second moment also apply to this moment vector. However, the second moment of (53) has a term that does not appear in the EMM moment (51). This term picks up the negative value of the second moment. In fact, the asymptotic indirect inference estimate of ρ is usually larger than the OLS estimate $\hat{\beta}_{1,T}$. In addition, the asymptotic indirect inference moment vector depends on an estimate of the sample variance of y_t instead of the population variance. As a result, the asymptotic indirect inference criterion function is quadratic. Depending on the sample, the asymptotic indirect inference estimate of ρ can equal one, or equivalently the first-order conditions are not satisfied at an estimate less than one. This is a consequence of using a conditional likelihood as the auxiliary model. We now present Monte Carlo evidence on the finite-sample performance of the estimators.

3.3 Finite Sample Properties

We first summarize features of the empirical density functions of estimates of ρ. Return to Table 2, which reports means, medians, and RMSEs of the parameter estimates. The OLS estimates are the estimates $\hat{\beta}_{1,T}$ from the conditional Gaussian auxiliary likelihood. Two features of the table are worth highlighting. First, for all combinations of ρ_0 and T, the point estimates produced with EMM exhibit greater bias and greater RMSE than estimates produced with either asymptotic indirect inference or OLS estimation. Second, there is no clear winner in a horse race between asymptotic and finite-sample indirect inference. Unsurprisingly, point estimates produced with the latter technique exhibit less bias. Also unsurprisingly, the noise introduced by using a small number of finite-sample simulations typically inflates the RMSE of finite-sample indirect inference relative to that of asymptotic indirect inference. However, for the highest-persistence value of ρ_0, the bias reduction overcomes the noise, and the RMSEs of finite-sample indirect inference are actually lower than those of asymptotic indirect inference.

Figure 4 displays empirical density functions for ML, EMM, and asymptotic indirect inference when the sample size is 1000 observations. Each panel contains three plots. (To limit clutter, densities for finite-sample indirect inference are not displayed. They are not markedly different from those for asymptotic indirect inference.) The solid lines are the densities produced with full ML estimation. Equivalently, they are the densities produced with EMM and indirect inference using the full likelihood as the auxiliary likelihood. The dotted lines are densities produced with EMM estimation using the conditional Gaussian auxiliary likelihood. The dashed lines are densities produced with asymptotic indirect inference using the same auxiliary model. The true value of ρ is denoted with a ''p'' in each panel

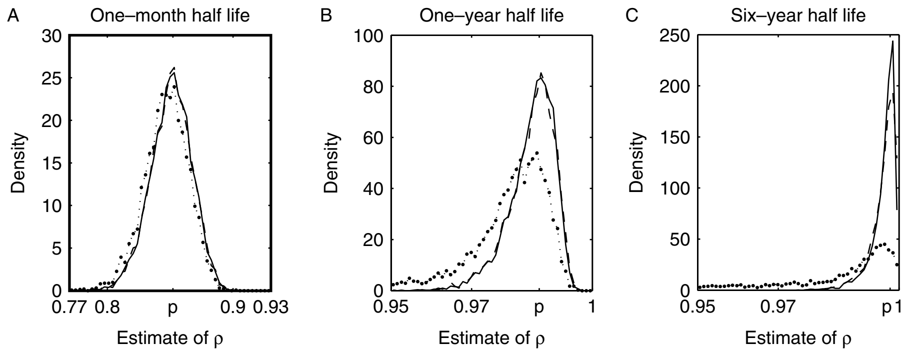

Figure 4 Empirical density functions of estimates of an autocorrelation coefficient.

There are three conclusions to draw from this figure. First, the asymptotic indirect inference densities are very close to the full ML densities. Second, the EMM density is similar to the others only when the data-generating process is not highly persistent. When the process has either a one-year or a six-year half life, the EMM density is significantly more diffuse than are the other two densities. In particular, the lower tails of the EMM densities are much fatter than the lower tails of the other densities. For specificity, consider the densities in Panel B (one-year half life). The interquartile range of the ML estimate is from 0.9819 to 0.9890. These values correspond to half lives of 0.73 years to 1.2 years. The interquartile range of the EMM estimate is from 0.9720 to 0.9857, corresponding to half lives of 0.47 years and 0.93 years. Not only is this range much larger, but it does not contain the true value of ρ.

Third, the efficiency of EMM estimation relative to the other estimation methods is lower when the data generating process exhibits greater persistence. (This conclusion is also evident in the RMSEs in Table 2.) Comparing Panels B and C, the ML and asymptotic indirect inference estimates are less diffuse when the process has a six-year half life than when it has a one-year half life. By contrast, the EMM estimates are more diffuse when the process has a six-year half life. The interquartile range of the EMM estimate in Panel B is from 0.9085 (outside of the plotted area) to 0.9924, corresponding to half lives of less than two months and 1.7 years.

We now turn to test statistics, including tests of over-identifying restrictions. Monte Carlo results for the EMM and indirect inference estimators are displayed in Table 4. The table reports median lengths of 95 percent confidence bounds for ρ. It also reports empirical rejection rates for tests of overall model adequacy and tests that $\rho = \rho_0$. Because three moments are used to identify a single parameter, the first category of statistics has an asymptotic $\chi^2(2)$ distribution and the second category has an asymptotic $\chi^2(1)$ distribution.[10]

[10] Indirect inference test statistics are not computed for samples in which the estimate of ρ is on the boundary of the stationary region. For the one-month and one-year half lives, none of the estimates

Table 4 Tests of an autocorrelation coefficient using a conditional Gaussian auxiliary model.

		Median length of 95% conf bound			Actual rejection rate at 95% critical value					
					Overident			$\rho = \rho_0$		
T	ρ_0	EMM	A-II	F-II	EMM	A-II	F-II	EMM	A-II	F-II
1000	0.8522	0.0685	0.0646	0.0726	0.066	0.061	0.057	0.086	0.050	0.056
2000	0.8522	0.0471	0.0457	0.512	0.057	0.054	0.055	0.075	0.057	0.060
10000	0.8522	0.0206	0.0205	0.0229	0.057	0.057	0.050	0.054	0.053	0.055
1000	0.9868	0.0564	0.0207	0.0209	0.126	0.079	0.097	0.344	0.056	0.144
2000	0.9868	0.0230	0.0144	0.0161	0.089	0.062	0.069	0.230	0.053	0.087
10000	0.9868	0.0070	0.0064	0.0071	0.061	0.057	0.051	0.102	0.050	0.057
1000	0.9978	0.1455	0.0078	0.0048	0.375	0.136	0.315	0.739	0.051	0.239
2000	0.9978	0.0892	0.0058	0.0042	0.234	0.110	0.203	0.590	0.049	0.176
10000	0.9978	0.0048	0.0026	0.0029	0.100	0.066	0.065	0.261	0.055	0.097

This table summarizes results from 5,000 Monte Carlo simulations. The true data-generating process is

$$y_t = \rho_0 y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, 1).$$

For each simulation, a sample of length T is generated. The parameter ρ is treated as unknown, and is estimated with EMM, asymptotic indirect inference (A-II), and finite-sample indirect inference (F-II) where the binding function is approximated with four simulations. The auxiliary likelihood is the conditional likelihood of a three-parameter AR(1) with Gaussian innovations. The columns labeled "Overident" report rejection rates of $\chi^2(2)$ tests of the overidentifying restrictions. The columns labeled "$\rho = \rho_0$" report empirical rejection rates for test statistics that have asymptotic $\chi^2(1)$ distributions under the null. They are equivalent to empirical rejection rates of the confidence bounds on ρ.

There are three main conclusions to draw from this table. First, asymptotic indirect inference, in combination with the conditional Gaussian likelihood, works well in finite samples even when the data are highly persistent. The median confidence bounds for indirect inference are typically slightly larger than those reported in Table 3 for ML. The larger confidence intervals are warranted: the empirical rejection rates of the hypothesis $\rho = \rho_0$ correspond almost exactly to the asymptotic rejection rates, in contrast to the modest over-rejections reported for ML in Table 3. Tests of the over-identifying restrictions are somewhat less well-behaved, but the differences between empirical and asymptotic rejection rates are small except for the six-year half life. The largest difference occurs with 1,000 observations: 14% of the statistics exceed the asymptotic 95% critical value.

is on the boundary. For the six-year half life and 1000 observations, slightly less than two percent of the estimates are on the boundary. Doubling the number of observations reduces the fraction on the boundary to about one quarter of one percent. None of the estimates are on the boundary for the six-year half life and 10,000 observations.

Second, EMM, in combination with the conditional Gaussian likelihood, works poorly when the data are highly persistent. The confidence bounds are large and the empirical rejection rates of truth are high. The poor performance of EMM is evident even in very large samples. With a six-year half life and 10,000 observations (almost 200 years of weekly data), the median EMM confidence interval is almost twice the length of the median asymptotic indirect inference confidence interval. Yet over one-quarter of the EMM test statistics for $\rho = \rho_0$ exceed the asymptotic 95% critical value. Naturally, shorter samples correspond to poorer finite-sample behavior. With 1,000 observations, almost three-fourths of the test statistics exceed the same critical value. Tests of the EMM overidentifying restrictions are somewhat better behaved, but also strongly over-reject the null. With a one-year half life the over-rejections are smaller, but remain significant. For example, with a one-year half life and 2,000 observations, almost one-quarter of the statistics for $\rho = \rho_0$ exceed the asymptotic 95% critical value.

Third, the empirical rejection rates for finite-sample indirect inference are generally higher than those for asymptotic indirect inference. This result conforms with the auxiliary likelihood that we observed in Section 2.4.2. Although finite-sample indirect inference produces less biased point estimates, it does not necessarily produce more accurate test statistics.

Comparing the results here with the results in Section 2 shows that implementing EMM with this conditional Gaussian auxiliary likelihood instead of the true full likelihood results in both less efficient parameter estimation and larger discrepancies between finite-sample and asymptotic critical values. The conditional Gaussian auxiliary likelihood differs from the true likelihood in two (obviously related) respects—it uses less information about the true model and it is more highly parameterized. A natural question is whether the consequences of these two aspects of the auxiliary likelihood can be disentangled.

3.4 Does Imposing Structural Model Restrictions Improve Efficiency?

Imagine that instead of estimating (2), we estimate a more general Gaussian AR(1):

$$y_t - \alpha = \rho(y_{t-1} - \alpha) + \epsilon_t, \quad \epsilon_t \sim N(0, v). \tag{55}$$

In other words, instead of using our knowledge of the true model to fix the unconditional mean to zero and the variance of shocks to one, we treat these parameters as unknown.

Full ML estimates of ρ are less efficient when the parameters α and v are treated as unknown than when they are fixed to their true values.[11] As can be seen in the RMSEs reported in Table 2, the same result holds for the auxiliary model used by EMM and asymptotic indirect inference. In other words, when estimating the single parameter ρ with either EMM or indirect inference, using the full likelihood as the auxiliary model produces more accurate point estimates than

[11] Although perhaps obvious, we verified this result with Monte Carlo simulations; the evidence is available on request.

using a conditional Gaussian likelihood. (The reduction in accuracy with indirect inference is fairly small, but as we saw in Figure 4, it is dramatic with EMM.)

Surprisingly, this intuition does not carry over to the *structural* model used by EMM estimation, at least for the process examined in this paper. In other words, holding the auxiliary model constant, estimates of ρ using the structural model that imposes the true values of α and v are less efficient than estimates of ρ using the structural model that treats α and v as unknown.

The evidence for this conclusion is hidden in Table 2. The information in the table allows us to compare the accuracy of two particular estimates of ρ. One estimate uses EMM combined with the conditional Gaussian auxiliary likelihood to estimate the single-parameter model (2). The other estimate uses EMM combined with the same conditional Gaussian auxiliary likelihood to estimate the three-parameter model (55). This estimate is identical to the OLS estimate of ρ. The reason for this equivalence is that the conditional Gaussian likelihood is the true conditional likelihood for (55). Therefore the EMM point estimate of the structural model equals the point estimate of the auxiliary model.[12] As noted in Section 3.1, the auxiliary-model point estimate is the OLS point estimate. Thus the columns labeled ''OLS'' in Table 2 can be interpreted as results for the estimate of ρ from EMM estimation of (55) using the conditional Gaussian auxiliary likelihood.

We see in the table that the OLS estimates are less biased than those from EMM estimation of (2) using the conditional Gaussian auxiliary likelihood. They are also less diffuse, in a RMSE sense. Put differently, when the same auxiliary model is used to estimate both the restrictive model (2) and the broader model (55) with EMM, the estimates of ρ using the broader model are more accurate.

What accounts for this apparently counterintuitive behavior? Although more information is available to pin down ρ when the restrictive model (2) is estimated than when (55) is estimated, EMM estimation does not necessarily use the information appropriately. Section 3.2 placed part of the blame for the poor EMM estimates of ρ on the correlation between the sample mean of the data and the weighting matrix. Hence the information in the difference between the true mean and the sample mean tends to be used incorrectly in EMM estimation of (2). Estimation of the broader model (55) throws away this information because it treats the true mean as an unknown parameter. For the sample sizes and values of ρ_0 examined here, throwing away the information is better than using the information improperly.

[12] This statement is a little loose. The estimate of the structural model will impose stationarity, but the estimate of the conditional Gaussian model will not. Inspection of the EMM moment condition for the structural model (55) combined with its conditional likelihood used as the auxiliary likelihood reveals that the point estimates of the auxiliary model equal the point estimates of the structural model when the point estimates for the auxiliary likelihood are in the stationary region. This condition was satisfied for all but 84 of the 5000 Monte Carlo simulations that specified 1000 observations and a six-year half life. It was satisfied for all but 4 of the simulations that specified 2000 observations and a six-year half life, and was always satisfied for all other combinations of T and ρ_0.

A more provocative way to express this conclusion is that with EMM estimation of highly persistent processes, it is more important to impose model-based restrictions on the auxiliary model than on the structural model. This runs counter to the standard advice to use a data-driven auxiliary model. We hasten to add that we do not know whether this result generalizes to other settings.

3.5 Some Caveats

The results of this section, combined with those in Section 2, do not place EMM estimation in a favorable light. Thus it is important to recognize that, in many ways, the AR(1) setting we study rules out many of the advantages of the EMM procedure that others have noted. Perhaps the most obvious difficulty in generalizing our results is that there is no need for EMM estimation in our setting because of the tractability of the likelihood function. Dynamic simulation methods are commonly used in settings where structural models are nonlinear, when shocks are conditionally heteroskedastic and conditionally nonnormal, and where likelihood functions are unknown or intractable. What we really want to know is how estimation techniques perform when confronted with the combination of highly persistent data and these more realistic dynamic properties. Unfortunately, the analytics are beyond us.

A clear limitation of our analysis is that the strong asymmetry of the EMM criterion function, which is largely responsible for the poor performance of EMM identified here, may well be a desirable property in more complicated settings. When the econometrician does not know how to impose stability on the parameter space, EMM ensures that neither the parameter estimates nor the confidence bounds include the explosive region. This enforced stability is not shared by indirect inference. Thus the asymmetry of the EMM criterion function might well produce more accurate point estimates and confidence regions than indirect inference.

From the perspective of dynamic term structure estimation (the motivation behind our study), this limitation is not particularly important, because for these models the stationary region of the parameter space is typically known. But even in this case, criterion function asymmetry may be useful in damping the effects of nonlinearities and fat tails near the boundary of stationarity. An investigation of this question is beyond the scope of the current work.

4 CONCLUSIONS

EMM and indirect inference are both asymptotically equivalent to ML when the auxiliary model nests the true likelihood function. This paper confirms that their finite-sample properties are also similar to those of ML when estimating an AR(1) process that is not close to a unit root. However, when the persistence of the data is similar to the observed persistence of interest rates, the finite-sample properties of EMM estimates differ substantially both from their asymptotic properties and from the finite-sample properties of indirect inference and ML.

The most obvious implication of our results is that any estimation of a structural model with highly persistent data should be accompanied by a Monte Carlo analysis of its finite-sample properties. Without this, it is impossible to draw reliable inferences from reported test statistics. We can also draw three further preliminary implications. Although we cannot be sure that these necessarily hold unaltered in more complex settings than considered here, their significance warrants further investigation. First, researchers using EMM to estimate such models might want to consider using model-based auxiliary models instead of data-driven models. Unfortunately, researchers lose much of their ability to test for misspecification when using a model-based auxiliary model. But the results here indicate that the efficiency and size properties of EMM estimates are improved substantially by imposing model-based restrictions on the mean of the data. Second, the superior performance of asymptotic indirect inference documented here suggests that researchers should consider using asymptotic indirect inference as an alternative to EMM estimation of such models. Third, our results also cast doubts on a common perception that finite-sample indirect inference (averaging over independent replications) is superior to asymptotic indirect inference (working with a single long simulation). Finite-sample indirect inference does, indeed, possess bias-reduction properties, but this advantage, at least in the setting we consider, is outweighed by its high computational cost and relatively poor finite-sample test statistics.

Received October 2, 2006; revised March 12, 2007; accepted July 24, 2007.

REFERENCES

Andersen, Torben G., Hyung-Jin Chung, and Bent E. Sørensen. (1999). ''Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study.'' *Journal of Econometrics 91*, 61–87.

Andrews, Donald W. K. (1993). ''Exactly Median Unbiased Estimation of First Order Autoregressive/Unit Root Models.'' *Econometrica 61*, 139–165.

Andrews, Donald W. K., and Hong-Yuan Chen. (1994). ''Approximately Median-Unbiased Estimation of Autoregressive Models with Application to U.S. Macroeconomic and Financial Time Series.'' *Journal of Business and Economic Statistics 12*, 187–204.

Ball, Clifford A., and Walter N. Torous. (1996). ''Unit Roots and the Estimation of Interest Rate Dynamics.'' *Journal of Empirical Finance 3*, 215–238.

Bansal, Ravi, A., Ronald Gallant, Robert Hussey, et al. (1993). *Computational Aspects of Nonparametric Simulation Estimation* In David A. Belsley (ed.), *Computational Techniques for Econometrics and Economic Analysis*. Boston: Kluwer.

Bansal, Ravi, A. Ronald Gallant, Robert Hussey, et al. (1995). ''Nonparametric Estimation of Structural Models for High-Frequency Currency Market Data.'' *Journal of Econometrics 66*, 251–287.

Chernozhukov, Victor, and Han Hong. (2003). ''An MCMC Approach to Classical Estimation.'' *Journal of Econometrics 115*, 293–346.

Chumacero, Rómulo A. (1997). ''Finite Sample Properties of the Efficient Method of Moments.'' *Studies in Nonlinear Dynamics and Econometrics 2*, 35–51.

Chumacero, Rómulo A. (2001). ''Estimating ARMA Models Efficiently.'' *Studies in Nonlinear Dynamics and Econometrics 5*, 103–114.

Clarida, Richard, Jordi Galí, and Mark Gertler. (2000). ''Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory.'' *Quarterly Journal of Economics 115*, 147–180.

Dickey, David A., and Wayne A. Fuller. (1979). ''Distribution of the Estimators for Autoregressive Time Series with a Unit Root.'' *Journal of the American Statistical Society 74*, 427–431.

Duffee, Gregory R., and Richard H. Stanton. (2004). Estimation of Dynamic Term Structure Models. Working Paper, U.C. Berkeley.

Efron, Bradley. (1979). ''Bootstrap Methods: Another Look at the Jackknife.'' *Annals of Statistics 7*, 1–26.

Gallant, A. Ronald, and George E. Tauchen. (1992). ''A Nonparametric Approach to Nonlinear Time Series Analysis: Estimation and Simulation.'' In David Brillinger, Peter Caines, John Geweke, Emanuel Parzen, Murray Rosenblatt, and Murad S. Taqqu (eds.), *New Directions in Time Series Analysis* Part II. New York: Springer-Verlag.

Gallant, A. Ronald, and George E. Tauchen. (1996). ''Which Moments to Match?.'' *Econometric Theory 12*, 657–681.

Gallant, A. Ronald, and George E. Tauchen. (1998). ''EMM: A Program for Efficient Method of Moments Estimation, Version 1.4.'' User's guide, University of North Carolina, Chapel Hill.

Gallant, A. Ronald, and George Tauchen. (2007). ''EMM: A Program for Efficient Method of Moments,'' http://www.econ.duke.edu/~get/emm.htm.

Gouriéroux, Christian, and Alain Monfort. (1996). *Simulation-Based Econometric Methods*. New York: Oxford University Press.

Gouriéroux, Christian, Alain Monfort, and Eric Renault. (1993). ''Indirect Inference.'' *Journal of Applied Econometrics 8*, S85–S118.

Gouriéroux, Christian, Eric Renault, and Nizar Touzi. (2000). ''Calibration by Simulation for Small Sample Bias Correction.'' In Roberto Mariano, Til Schuerman, and Melvyn J. Weeks (eds.), *Simulation-Based Inference in Econometrics*. Cambridge: Cambridge University Press.

Hansen, Lars Peter, John C. Heaton, and Amir Yaron. (1996). ''Finite-Sample Properties of Some Alternative GMM Estimators.'' *Journal of Business and Economic Statistics 14*, 262–280.

Kendall, M. G. (1954). ''Note on Bias in the Estimation of Autocorrelation.'' *Biometrika 41*, 403–404.

Marriott, F. H. C., and J. A. Pope. (1954). ''Bias in the Estimation of Autocorrelations.'' *Biometrika 41*, 390–402.

Michaelides, Alexander, and Serena Ng. (2000). ''Estimating the Rational Expectations Model of Speculative Storage: A Monte Carlo Comparison of Three Simulation Estimators.'' *Journal of Econometrics 96*, 231–266.

Monfardini, Chiara. (1998). ''Estimating Stochastic Volatility Models through Indirect Inference.'' *Econometrics Journal 1*, C113–C128.

Orcutt, G. H. (1948). ''A Study of the Autoregressive Nature of the Time Series used for Tinbergen's Model of the Economic System of the United States, 1919–1932.'' *Journal of the Royal Statistical Society Series B 10*, 1–45.

Quenouille, M. H. (1949). ''Approximate Tests of Correlation in Time Series.'' *Journal of the Royal Statistical Society Series B 11*, 68–84.

Rudebusch, G. D. (1992). ''Trends and Random Walks in Macroeconomic Time Series: A Re-examination.'' *International Economic Review 33*, 661–680.
Sims, Christopher A., and Harald Uhlig. (1991). ''Understanding Unit Rooters: A Helicopter Tour.'' *Econometrica 59*, 1591–1599.
Smith, Anthony A. (1990). ''Three Essays on the Solution and Estimation of Dynamic Macroeconomic Models.'' Ph.D. thesis, Duke University.
Smith, Anthony A. (1993). ''Estimating Nonlinear Time Series Models using Simulated Vector Autoregressions.'' *Journal of Applied Econometrics 8*, S63–S84.
Tauchen, George E. (1998). ''The Objective Function of Simulation Estimators Near the Boundary of the Unstable Region of the Parameter Space.'' *Review of Economics and Statistics 80*, 389–398.
Zhou, Hao. (2001). ''Finite Sample Properties of EMM, GMM, QMLE, and MLE for a Square-Root Interest Rate Diffusion Model.'' *Journal of Computational Finance 5*, 89–122.