Econometric Theory, **12**, 1996, 657–681. Printed in the United States of America.

WHICH MOMENTS TO MATCH?

A. RONALD GALLANT
University of North Carolina
GEORGE TAUCHEN
Duke University

We describe an intuitive, simple, and systematic approach to generating moment conditions for generalized method of moments (GMM) estimation of the parameters of a structural model. The idea is to use the score of a density that has an analytic expression to define the GMM criterion. The auxiliary model that generates the score should closely approximate the distribution of the observed data but is not required to nest it. If the auxiliary model nests the structural model then the estimator is as efficient as maximum likelihood. The estimator is advantageous when expectations under a structural model can be computed by simulation, by quadrature, or by analytic expressions but the likelihood cannot be computed easily.

1. INTRODUCTION

We present a systematic approach to generating moment conditions for the generalized method of moments (GMM) estimator (Hansen, 1982) of the parameters of a structural model. The approach is an alternative to the common practice of selecting a few low-order moments on an ad hoc basis and then proceeding with GMM. The idea is simple: Use the expectation under the structural model of the score from an auxiliary model as the vector of moment conditions.

This score is the derivative of the log density of the auxiliary model with respect to the parameters of the auxiliary model. Thus, the moment conditions depend on both the parameters of the auxiliary model and the parameters of the structural model. The parameters of the auxiliary model are replaced by their quasimaximum likelihood estimates, which are computed by maximizing the pseudolikelihood of the auxiliary model.

The estimates of the structural parameters are computed by minimizing a GMM criterion function. As seen later, the optimal weighting matrix for forming the GMM criterion from the moment conditions depends only on the auxiliary model and is easily computed.

We thank Laura Baldwin, Ravi Bansal, John Coleman, and Anthony Smith for helpful discussions and two referees for very useful comments. This paper was supported by the National Science Foundation. Address correspondence to: A. Ronald Gallant, Department of Economics, University of North Carolina, CB #3305, 6F Gardner Hall, Chapel Hill, NC 27599-3305, USA; e-mail: ron_gallant@unc.edu.

 0266-4666/96 $9.00 + .10

We call the auxiliary model the score generator. The score generator need not encompass (nest) the structural model. If it does, then the estimator is as efficient as the maximum likelihood estimator. Hence, our approach ensures efficiency against a given parametric model. If the score generator closely approximates the actual distribution of the data, even though it does not encompass it, then the estimator is nearly fully efficient.

The estimation context that we have in mind is one where a structural model defines a data generation process for the data. The key feature of this data generation process is that it is relatively easy to compute the expectation of a nonlinear function given values for the structural parameters. An expectation may be computed by simulation, by numerical quadrature, or by analytic expressions, whichever is the most convenient.

Examples of this estimation context are the panel data models motivating the simulated method of moments approach of Pakes and Pollard (1989) and McFadden (1989). Another is the asset pricing model that motivates the dynamic method of moments estimator of Duffie and Singleton (1993). In Section 4, we present three such situations drawn from macroeconomics, finance, and empirical auction modeling. In these examples, the likelihood is difficult to compute, so maximum likelihood is infeasible. Simulation and moment matching thus naturally arise.

As indicated, there is no presumption that the score generator encompasses the structural model, although an order condition for identification requires a minimal level of complexity of the score generator. Under weak regularity conditions, our estimator is root-n consistent and asymptotically normal with an asymptotic distribution that depends on both the structural model and the score generator. If there exists a local, smooth mapping of the structural parameters into the parameters of the score generator, then the estimator has the same asymptotic distribution as the maximum likelihood estimator under the structural model.

The asymptotic theory of the estimator subsumes situations with strictly exogenous variables, where one conditions on particular values of the explanatory variables. It also subsumes situations with predetermined but not strictly exogenous variables, as is typical of stationary Markov data generation processes. The most general version allows for processes with time-dependent laws of motion and dependence extending into the indefinite past.

Section 2 presents the asymptotic justification of the estimator. Section 3 presents some candidate specifications for the score generator. Section 4 presents three proposed applications, each of which is a substantive empirical project.

2. THEORY

For a stochastic process described by a sequence of densities $\{p_1(x_1|\rho), \{p_t(y_t|x_t,\rho)\}_{t=1}^{\infty}\}$ for which expectations of nonlinear functions are easily computed by simulation, by quadrature, or by analytic expressions, we

derive a computationally convenient GMM estimator for ρ that uses the scores $(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta)$ from another sequence of densities $\{f_1(x_1|\theta), \{f_t(y_t|x_t,\theta)\}_{t=1}^{\infty}\}$ to generate the moment conditions. As an example, $\{p_1(x_1|\rho), \{p_t(y_t|x_t,\rho)\}_{t=1}^{\infty}\}$ might be a model that describes asset prices y_t in terms of exogenous variables x_t and structural parameters ρ and $\{f_1(x_1|\theta), \{f_t(y_t|x_t,\theta)\}_{t=1}^{\infty}\}$ might be the sequence of densities of a GARCH process (Bollerslev, 1986, 1987). The estimator is consistent and asymptotically normally distributed in general. It is fully efficient if the model $\{p_1(x_1|\rho), \{p_t(y_t|x_t,\rho)\}_{t=1}^{\infty}\}_{\rho\in R}$ is smoothly embedded (Definition 1) within the model $\{f_1(x_1|\theta), \{f_t(y_t|x_t,\theta)\}_{t=1}^{\infty}\}_{\theta\in\Theta}$. The estimator is attractive when the density $f_t(y_t|x_t,\theta)$ has a convenient analytic expression whereas the density $p_t(y_t|x_t,\rho)$ does not.

Throughout, the observed data $\{\tilde{y}_t,\tilde{x}_t\}_{t=1}^{n}$ are assumed to have been generated from the sequence of densities

$$\{p_1(x_1|\rho^o), \{p_t(y_t|x_t,\rho^o)\}_{t=1}^{n}\}, \tag{1}$$

which is to say that ρ^o denotes the true value of the parameter ρ in the model

$$\{p_1(x_1|\rho), \{p_t(y_t|x_t,\rho)\}_{t=1}^{\infty}\}_{\rho\in R}, \tag{2}$$

where R denotes the parameter space. The model

$$\{f_1(x_1|\theta), \{f_t(y_t|x_t,\theta)\}_{t=1}^{\infty}\}_{\theta\in\Theta} \tag{3}$$

is called the score generator. The variables y_t and x_t can be univariate or multivariate or have a dimension that depends on t. The functional form of the score generator may act to exclude elements of the vector y_t; that is, the score generator may define a stochastic process for only some elements of y_t.

When we say that a process is time invariant we mean that the densities of the process do not depend on t, in which case the t subscripts on the densities may be suppressed and the dimensions of y_t and x_t are fixed. Writing $\sqrt{n}X_n \approx N(0,V_n)$ means $\sqrt{n}(V_n^{1/2})^{-1}X_n \xrightarrow{\mathcal{L}} N(0,I)$, where $V_n = (V_n^{1/2})(V_n^{1/2})'$. Smoothly embedded is defined as follows.

DEFINITION 1. *The model* $\{p_1(x_1|\rho), \{p_t(y_t|x_t,\rho)\}_{t=1}^{\infty}\}_{\rho\in R}$ *is said to be smoothly embedded within the score generator* $\{f_1(x_1|\theta), \{f_t(y_t|x_t,\theta)\}_{t=1}^{\infty}\}_{\theta\in\Theta}$ *if for some open neighborhood* R^o *of* ρ^o *there is a twice continuously differentiable mapping* $g: R^o \to \Theta$ *such that*

$$p_t(y_t|x_t,\rho) = f_t[y_t|x_t,g(\rho)], \qquad t = 1,2,\ldots, \tag{4}$$

for every $\rho \in R^o$ *and* $p_1(x_1|\rho) = f_1[x_1|g(\rho)]$ *for every* $\rho \in R^o$.

We consider three cases.

Case 1. All densities are time invariant, the analysis is conditional on the observed sequence $\{\tilde{x}_t\}_{t=1}^{n}$, and the data $\{(\tilde{y}_t,\tilde{x}_t)\}_{t=1}^{n}$ are a sample from

$\prod_{t=1}^{n} p(y_t|\tilde{x}_t,\rho^o)$. An example for which these assumptions are appropriate is the nonlinear regression model on cross-sectional data with independently and identically distributed errors. For this case, using simulation to compute the GMM estimator proposed here requires N simulated sequences $\{\{\hat{y}_{t\tau}\}_{t=1}^{n}\}_{\tau=1}^{N}$ from the density $\prod_{t=1}^{n} p(y_t|\tilde{x}_t,\rho)$. We impose Assumptions 1–6 of Gallant (1987, Ch. 3) on both $p(y|x,\rho)$ and $f(y|x,\theta)$.

Case 2. All densities are time invariant, the analysis is unconditional, and the data $\{(\tilde{y}_t,\tilde{x}_t)\}_{t=1}^{n}$ are a sample from $\prod_{t=1}^{n} p(y_t|x_t,\rho^o)p(x_1|\rho^o)$. An example for which these assumptions are appropriate is an autoregressive process where x_t is comprised of L lagged values of y_t. We impose Assumptions 1–6 of Gallant (1987, Ch. 7) on both $p(y|x,\rho)$ and $f(y|x,\theta)$. In addition, $\{(y_t,x_t)\}_{t=-\infty}^{\infty}$ is assumed to be stationary with joint density $p(y,x|\rho)$, marginal density $p(x|\rho)=\int p(y,x|\rho)\,dy$, and conditional density $p(y|x,\rho)=p(y,x|\rho)/p(x|\rho)$. Similarly for $f(y|x,\theta)$. For this case, using simulation to compute the GMM estimator requires only a single simulated sequence $\{(\hat{y}_\tau,\hat{x}_\tau)\}_{\tau=1}^{N}$ from the density $\prod_{\tau=1}^{N} p(y_\tau|x_\tau,\rho)p(x_1|\rho)$, generated as follows: Start at an arbitrary $\hat{x}_1=(\hat{y}_0,\ldots,\hat{y}_{-L+1})$, simulate $\hat{y}_1$ from $p(y_1|\hat{x}_1,\rho)$, put $\hat{x}_2=(\hat{y}_1,\ldots,\hat{y}_{-L+2})$, simulate $\hat{y}_2$ from $p(y_2|\hat{x}_2,\rho)$, and so on. So that $\hat{x}_\tau$ is plausibly a sample from $p(x|\rho)$, enough initial simulations are discarded for transients to die out and the next N simulations are retained as the sequence $\{(\hat{y}_\tau,\hat{x}_\tau)\}_{\tau=1}^{N}$.

Case 3. Densities are not time invariant, the analysis is unconditional, and the data $\{(\tilde{y}_t,\tilde{x}_t)\}_{t=1}^{n}$ are a sample from $\prod_{t=1}^{n} p_t(y_t|x_t,\rho^o)p_1(x_1|\rho^o)$. This framework does permit conditioning on the initial observation x_1 and conditioning on exogenous variables. Conditioning on x_1 is accomplished by letting $p_1(x_1|\rho)$ put its mass on a single point. Conditioning on exogenous variables w_t is accomplished through the dependence of $p_t(y_t|x_t,\rho)$ on t by putting $p_t(y_t|x_t,\rho)=p(y_t|x_t,w_t,\rho)$. An example for which these assumptions are appropriate is a nonlinear regression with fixed regressors w_t and lagged dependent variables x_t. For Case 3, using simulation to compute the GMM estimator may require N simulated sequences $\{\{\hat{y}_{t\tau},\hat{x}_{t\tau}\}_{t=1}^{n}\}_{\tau=1}^{N}$ from the density $\prod_{t=1}^{n} p_t(y_t|x_t,\rho)p_1(x_1|\rho)$. However, in the common case where the structural model is Case 2 and the score generator describes an asymptotically strictly stationary process, a single simulated sequence as in Case 2 suffices. We impose Assumptions 1–6 of Gallant (1987, Ch. 7) on both $p_t(y_t|x_t,\rho)$ and $f_t(y_t|x_t,\theta)$.

Our idea is to use the scores

$$(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta) \tag{5}$$

evaluated at the quasimaximum likelihood estimate

$$\tilde{\theta}_n=\underset{\theta\in\Theta}{\operatorname{argmax}}\ \frac{1}{n}\sum_{t=1}^{n}\ln f_t(\tilde{y}_t|\tilde{x}_t,\theta) \tag{6}$$

to generate GMM moment conditions. The GMM moment equations are

$$\text{Case 1:}\quad m_n(\rho,\tilde{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\int (\partial/\partial\theta)\ln f(y|\tilde{x}_t,\tilde{\theta}_n)p(y|\tilde{x}_t,\rho)\,dy, \tag{7}$$

$$\text{Case 2:}\quad m_n(\rho,\tilde{\theta}_n) = \iint (\partial/\partial\theta)\ln f(y|x,\tilde{\theta}_n)p(y|x,\rho)\,dy\,p(x|\rho)\,dx, \tag{8}$$

$$\text{Case 3:}\quad m_n(\rho,\tilde{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\int\cdots\int (\partial/\partial\theta)\ln f_t(y_t|x_t,\tilde{\theta}_n) \times \prod_{\tau=1}^{n} p_\tau(y_\tau|x_\tau,\rho)\,dy_\tau\,p_1(x_1|\rho)\,dx_1. \tag{9}$$

These are the moment conditions that define the estimator. In most applications, analytic expressions for the integrals will not be available and simulation or quadrature will be required to compute them.

If the integrals are computed by simulation, then the formulas used in practice are

$$\text{Case 1:}\quad m_n(\rho,\tilde{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{N}\sum_{\tau=1}^{N}(\partial/\partial\theta)\ln f(\hat{y}_{t\tau}|\hat{x}_t,\tilde{\theta}_n), \tag{10}$$

$$\text{Case 2:}\quad m_n(\rho,\tilde{\theta}_n) = \frac{1}{N}\sum_{\tau=1}^{N}(\partial/\partial\theta)\ln f(\hat{y}_\tau|\hat{x}_\tau,\tilde{\theta}_n), \tag{11}$$

$$\text{Case 3:}\quad m_n(\rho,\tilde{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{N}\sum_{\tau=1}^{N}(\partial/\partial\theta)\ln f_t(\hat{y}_{t\tau}|\hat{x}_{t\tau},\tilde{\theta}_n). \tag{12}$$

We assume that N is large enough that the Monte Carlo integral approximates the analytic integral to within a negligible error of the same sort as is made in computing any mathematical expression on a digital computer.

There are instances where the integral can be computed to a given accuracy at less cost by quadrature. Quadrature rules have the generic form

$$\text{Case 1:}\quad m_n(\rho,\tilde{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{N}\sum_{\tau=1}^{N}(\partial/\partial\theta)\ln f(\hat{y}_{t\tau}|\tilde{x}_t,\tilde{\theta}_n)W(\hat{y}_{t\tau}|\tilde{x}_t,\rho), \tag{13}$$

$$\text{Case 2:}\quad m_n(\rho,\tilde{\theta}_n) = \frac{1}{N}\sum_{\tau=1}^{N}(\partial/\partial\theta)\ln f(\hat{y}_\tau|\hat{x}_\tau,\tilde{\theta}_n)W(\hat{y}_\tau,\hat{x}_\tau|\rho), \tag{14}$$

where $W(\hat{y}_\tau,\hat{x}_\tau|\rho)$ and $(\hat{y}_\tau,\hat{x}_\tau)$ are the weights and the abscissae implied by the quadrature rule. Of course, N is dramatically smaller for quadrature rules than for Monte Carlo integration. Quadrature for Case 3 will be at too high a dimension to be practical in most applications.

As to statistical theory, Cases 1 and 2 are special cases of Case 3, so that throughout the rest of the discussion we can discuss Case 3 and specialize to Cases 1 and 2, as required.

The randomness in $m_n(\rho,\tilde{\theta}_n)$ is solely due to the random fluctuation of the maximum likelihood estimator $\tilde{\theta}_n$. Under the regularity conditions imposed earlier, there is a sequence $\{\theta_n^o\}$ such that $m_n(\rho^o,\theta_n^o)=0$, $\lim_{n\to\infty}(\tilde{\theta}_n-\theta_n^o)=0$ almost surely, and

$$\sqrt{n}(\tilde{\theta}_n-\theta_n^o)\approx N[0,(\mathcal{J}_n^o)^{-1}(\mathcal{I}_n^o)(\mathcal{J}_n^o)^{-1}], \tag{15}$$

where $\mathcal{J}_n^o=(\partial/\partial\theta')m_n(\rho^o,\theta_n^o)$ and

$$\mathcal{I}_n^o=\operatorname{Var}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\theta_n^o)\right] \tag{16}$$

(Gallant, 1987, Ch. 7, Theorem 6). Note that $\mathcal{I}_n^o$ and $\mathcal{J}_n^o$ are not random quantities because we have assumed that either quadrature has been employed to compute $m_n(\rho,\theta)$ or that N is as large as necessary to make the average essentially the same as the expected value. Using Taylor's theorem

$$\begin{aligned}\sqrt{n}m_n(\rho^o,\tilde{\theta}_n)&=\sqrt{n}m_n(\rho^o,\theta_n^o)+[\mathcal{J}_n^o+o_s(1)]\sqrt{n}(\tilde{\theta}_n-\theta_n^o)\\&=(\mathcal{J}_n^o)\sqrt{n}(\tilde{\theta}_n-\theta_n^o)+o_p(1).\end{aligned} \tag{17}$$

This implies that

$$\sqrt{n}m_n(\rho^o,\tilde{\theta}_n)\approx N(0,\mathcal{I}_n^o). \tag{18}$$

Thus, given an estimator $\tilde{\mathcal{I}}_n$ of $\mathcal{I}_n^o$ that is consistent in the sense that $\lim_{n\to\infty}(\tilde{\mathcal{I}}_n-\mathcal{I}_n^o)=0$ almost surely, the GMM estimator with an efficient weighting matrix is

$$\hat{\rho}_n=\underset{\rho\in R}{\operatorname{argmin}}\ m_n'(\rho,\tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1}m_n(\rho,\tilde{\theta}_n). \tag{19}$$

The computations necessary to estimate $\mathcal{I}_n^o$ depend on how well one thinks that the score generator approximates the true data generating process. If one is confident that the score generator is a good statistical approximation to the data generating process, then the estimator

$$\tilde{\mathcal{I}}_n=\frac{1}{n}\sum_{t=1}^{n}[(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\tilde{\theta}_n)][(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\tilde{\theta}_n)]' \tag{20}$$

can be used. This estimator can also be used with Gaussian QMLE scores if the conditional mean and variance functions are correctly specified (Bollerslev and Woolridge, 1992). A sufficient (but not necessary) condition is Assumption 2.

A weaker assumption that facilitates estimation of $\mathcal{I}_n^o$ is the following.

Assumption 1. There is a θ^o such that

$$\int\cdots\int(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)\prod_{\tau=1}^{n}p_\tau(y_\tau|x_\tau,\rho^o)\,dy_\tau\,p_1(x_1|\rho^o)\,dx_1=0 \tag{21}$$

for every $t\le n$.

Case 2 will always satisfy Assumption 1 because of stationarity and time invariance. Thus, it is an assumption that only affects Cases 1 and 3.

For Case 1, the preceding estimator,

$$\tilde{\mathfrak{I}}_n = \frac{1}{n}\sum_{t=1}^{n}[(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\tilde{\theta}_n)][(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\tilde{\theta}_n)]', \tag{22}$$

retains its consistency under the weaker Assumption 1.

For Cases 2 and 3, the following estimator is consistent under Assumption 1:

$$\tilde{\mathfrak{I}}_n = \sum_{\tau=-[n^{1/5}]}^{[n^{1/5}]} w\left(\frac{\tau}{[n^{1/5}]}\right)\tilde{S}_{n\tau}, \tag{23}$$

where

$$w(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3 & \text{if } 0 \le x \le \frac{1}{2}, \\ 2(1-|x|)^3 & \text{if } \frac{1}{2} \le x \le 1, \end{cases} \tag{24}$$

and

$$\tilde{S}_{n\tau} = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{t=1+\tau}^{n}[(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\tilde{\theta}_n)][(\partial/\partial\theta)\ln f_{t-\tau}(\tilde{y}_{t-\tau}|\tilde{x}_{t-\tau},\tilde{\theta}_n)]' \\ \qquad\qquad \text{if } \tau \ge 0, \\ (\tilde{S}_{n,-\tau})' \quad \text{if } \tau < 0 \end{cases} \tag{25}$$

(Gallant, 1987, Ch. 7, Theorem 5). See Andrews (1991) for alternative suggestions as to appropriate weights and rates one might use instead of $w(x)$ and $n^{1/5}$. The Parzen weights suggested above guarantee the positive definiteness of $\tilde{\mathfrak{I}}_n$, which is essential. Weights that do not guarantee positive definiteness cannot be used.

If one is unwilling to accept Assumption 1, then the estimator $\tilde{\mathfrak{I}}_n$ is modified as follows. First, compute the initial estimator

$$\rho_n^{\#} = \underset{\rho\in R}{\operatorname{argmin}}\, m_n'(\rho,\tilde{\theta}_n)m_n(\rho,\tilde{\theta}_n). \tag{26}$$

Compute

$$\tilde{\mu}_t = \int\cdots\int (\partial/\partial\theta)\ln f_t(y_t|x_t,\tilde{\theta}_n)\prod_{\tau=1}^{n} p_\tau(y_\tau|x_\tau,\rho_n^{\#})\,dy_\tau\, p_1(x_1|\rho_n^{\#})\,dx_1 \tag{27}$$

using the integration methods already described. For Case 1, use the estimator

$$\tilde{\mathfrak{I}}_n = \frac{1}{n}\sum_{t=1}^{n}[(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\tilde{\theta}_n) - \tilde{\mu}_t][(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\tilde{\theta}_n) - \tilde{\mu}_t]'. \tag{28}$$

For Case 3, use the formula

$$\tilde{\mathcal{I}}_n = \sum_{\tau=-[n^{1/5}]}^{[n^{1/5}]} w\left(\frac{\tau}{[n^{1/5}]}\right)\tilde{S}_{n\tau} \tag{29}$$

with $\tilde{S}_{n\tau}$ above modified to read

$$\tilde{S}_{n\tau} = \frac{1}{n}\sum_{t=1+\tau}^{n} [(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\tilde{\theta}_n) - \tilde{\mu}_t] \times [(\partial/\partial\theta)\ln f_{t-\tau}(\tilde{y}_{t-\tau}|\tilde{x}_{t-\tau},\tilde{\theta}_n) - \tilde{\mu}_{t-\tau}]' \tag{30}$$

for $\tau \geq 0$. It is unlikely that this generality will be necessary in practice because the use of this formula means that one thinks that the score generator is a poor statistical approximation to the data generating process, which is unlikely to be true for the following reasons. The score generator is conceptually a reduced form model, not a structural model. Thus, it is ordinarily easy to modify it by adding a few parameters so that it fits the data well. The situation where one thinks the score generator is a poor approximation might arise in hypothesis testing, but even then the null hypothesis will usually imply either Assumption 1 or Assumption 2 and the generality is, again, unnecessary.

Theorem 1 gives the asymptotic distribution of $\hat{\rho}_n$.

THEOREM 1. *For Case* 1, *let Assumptions* 8–11 *of Gallant* (1987, *Ch.* 3) *hold. For Cases* 2 *and* 3, *let Assumptions* 8–11 *of Gallant* (1987, *Ch.* 7) *hold. Then,*

$$\lim_{n\to\infty} \hat{\rho}_n = \rho^o \quad a.s., \tag{31}$$

$$\sqrt{n}(\hat{\rho}_n - \rho^o) \approx N\{0, [(M_n^o)'(\mathcal{I}_n^o)^{-1}(M_n^o)]^{-1}\}, \tag{32}$$

$$\lim_{n\to\infty} (\hat{M}_n - M_n^o) = 0 \quad a.s., \tag{33}$$

where $\hat{M}_n = M_n(\hat{\rho}_n,\tilde{\theta}_n)$, $M_n^o = M_n(\rho^o,\theta_n^o)$, *and* $M_n(\rho,\theta) = (\partial/\partial\rho')m_n(\rho,\theta)$.

Proof. Apply Theorems 7 and 9 of Gallant (1987, Ch. 3) for Case 1 and Theorems 8 and 10 of Gallant (1987, Ch. 7) for Cases 2 and 3. Make these associations: $\lambda = \rho$, $\lambda_n^o = \rho^o$, $m_n(\lambda) = m_n(\rho,\tilde{\theta}_n)$, $m_n^o(\lambda) = m_n(\rho,\theta_n^o)$, and $S_n^o = \text{Var}[\sqrt{n}m_n(\lambda_n^o)] = \mathcal{I}_n^o$. ■

The identification condition

$$m_n(\rho,\theta_n^o) = 0 \Rightarrow \rho = \rho^o \quad \text{for all } n \text{ larger than some } n^o \tag{34}$$

is among the regularity conditions of Theorem 1. The situation is analogous to verification of the order and rank conditions of simultaneous equations models. The order condition is that the dimension of θ must exceed the dimension of ρ. However, due to nonlinearity, analytic verification of the analog of the rank condition, which is that the equations $m_n(\rho,\theta_n^o) = 0$ do not have multiple solutions for $\rho \in R$, is difficult. See Gallant (1977) for dis-

cussion and examples. It is usually adequate to rely on the optimization program used to compute $\operatorname{argmin}_{\rho\in R} m_n'(\rho,\tilde{\theta}_n)(\tilde{\mathcal{J}}_n)^{-1}m_n(\rho,\tilde{\theta}_n)$ to indicate the flat spots on the surface $m_n'(\rho,\tilde{\theta}_n)(\tilde{\mathcal{J}}_n)^{-1}m_n(\rho,\tilde{\theta}_n)$ that suggest identification failure. For example, the parameters of the mixing process of a stochastic volatility model (see Section 4.2) require third- and fourth-order moment information for identification. Using the score of a Gaussian vector autoregression will not provide this information. We have actually done this inadvertently by setting some tuning parameters erroneously in a computation and learned of the error by the behavior of the optimizer with respect to the parameters of the mixing process.

Direct use of Theorem 1 for setting confidence intervals on the elements of $\hat{\rho}_n$ or testing hypotheses with the Wald test requires computation of $M_n(\rho,\theta)$. This is probably easiest to do by saving the trial values of ρ and $m(\rho,\tilde{\theta}_n)$ generated over the course of the optimization that computes $\operatorname{argmin}_{\rho\in R} m_n'(\rho,\tilde{\theta}_n)(\tilde{\mathcal{J}}_n)^{-1}m_n(\rho,\tilde{\theta}_n)$, fitting the local quadratic regressions $m_i = b_{oi} + b_i'(\rho - \hat{\rho}_n) + (\rho - \hat{\rho}_n)'B_i(\rho - \hat{\rho}_n)$ for $i = 1,2,\ldots,\dim(\theta)$ to the elements of $m_n(\rho,\tilde{\theta}_n)$ at points near $\hat{\rho}_n$ and taking $\hat{M}_n$ to be the matrix with rows b_i'. Computation of $M_n(\rho,\theta)$ can be avoided by testing hypotheses using the criterion difference test statistic (Gallant, 1987, Ch. 7, Theorem 15) and setting confidence intervals by inverting it. Under Assumption 1, the condition $HVH' = H\mathcal{J}^{-1}H'$ of Gallant (1987, Ch. 7, Theorem 15) will be satisfied.

It is important to note that we have not, as yet, made use of an assumption that the score generator $\{f_1(x_1|\theta),\{f_t(y_t|x_t,\theta)\}_{t=1}^{\infty}\}_{\theta\in\Theta}$ contains the true model. That is, we have not yet imposed the following assumption.

Assumption 2. There is a θ^o such that $p_t(y_t|x_t,\rho^o) = f_t(y_t|x_t,\theta^o)$ for $t = 1,2,\ldots$ and $p_1(x_1|\rho^0) = f_1(x_1|\theta^o)$.

Because Assumption 2 implies that the score generator is a correctly specified model, it implies Assumption 1 and the following standard results from the theory of maximum likelihood estimation:

$$\int (\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)p_t(y_t|x_t,\rho^o)\,dy_t = 0 \tag{35}$$

for $t = 1,2,\ldots,$

$$\int (\partial/\partial\theta)\ln f_1(x_1|\theta^o)p_1(x_1|\rho^o)\,dx_1 = 0, \tag{36}$$

and

$$\int\cdots\int [(\partial/\partial\theta)\ln f_s(y_s|x_s,\theta^o)]\,[(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)]' \times \prod_{\tau=1}^{n} p_\tau(y_\tau|x_\tau,\rho^o)\,dy_\tau\, p_1(x_1|\rho^o)\,dx_1 = 0 \tag{37}$$

when $t \neq s$. These results allow use of the estimator

$$\tilde{\mathfrak{I}}_n = \frac{1}{n}\sum_{t=1}^{n}[(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\tilde{\theta}_n)][(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\tilde{\theta}_n)]' \tag{38}$$

in Cases 1–3. Moreover,

$$\begin{aligned}
&\int\cdots\int[(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)][(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)]' \\
&\quad\times\prod_{\tau=1}^{n}p_\tau(y_\tau|x_\tau,\rho^o)\,dy_\tau\,p_1(x_1|\rho^o)\,dx_1 \\
&=-\int\cdots\int(\partial^2/\partial\theta\,\partial\theta')\ln f_t(y_t|x_t,\theta^o) \\
&\quad\times\prod_{\tau=1}^{n}p_\tau(y_\tau|x_\tau,\rho^o)\,dy_\tau\,p_1(x_1|\rho^o)\,dx_1
\end{aligned} \tag{39}$$

for $t = 1,2,\ldots$, so that

$$\sqrt{n}(\tilde{\theta}_n - \theta^o) \approx N[0,(\mathfrak{I}_n^o)^{-1}]. \tag{40}$$

Now let us examine the consequences of the smoothly embedded assumption (see Definition 1, earlier).

Assumption 3. The model $\{p_1(x_1|\rho),\{p_t(y_t|x_t,\rho)\}_{t=1}^{\infty}\}_{\rho\in R}$ is smoothly embedded within the score generator $\{f_1(x_1|\theta),\{f_t(y_t|x_t,\theta)\}_{t=1}^{\infty}\}_{\theta\in\Theta}$.

Assumption 3 implies Assumption 2. Moreover, the consistency of $\hat{\rho}_n$ implies that $\hat{\rho}_n$ is tail equivalent (Gallant, 1987, p. 187) to a GMM estimator obtained by optimizing over the closure of R^o instead of over R. Therefore, without loss of generality, we may assume that the twice continuously differentiable function g given by Definition 1 is defined over R. Let $G(\rho) = (\partial/\partial\rho')g(\rho)$, $G^o = G(\rho^o)$, and $\hat{G} = G(\hat{\rho})$.

A consequence of Assumption 3 is that the minimum chi-square estimator

$$\hat{\rho}_{mcs} = \underset{\rho\in R}{\operatorname{argmin}}\,[\tilde{\theta}_n - g(\rho)]'(\tilde{\mathfrak{I}}_n)[\tilde{\theta}_n - g(\rho)] \tag{41}$$

is as efficient as the maximum likelihood estimator for $\{p_1(x_1|\rho),\{p_t(y_t|x_t,\rho)\}_{t=1}^{\infty}\}$. To see this, first note that

$$\sqrt{n}(\hat{\rho}_{mcs} - \rho^o) \approx N\{0,[(G_n^o)'(\mathfrak{I}_n^o)(G_n^o)]^{-1}\}. \tag{42}$$

Now, if $\hat{\rho}_{mle}$ denotes the maximum likelihood estimator for $\{f_1[x_1|g(\rho)],\{f_t[y_t|x_t,g(\rho)]\}_{t=1}^{\infty}\}$, then

$$\sqrt{n}(\hat{\rho}_{mle} - \rho^o) \approx N\{0,[(G_n^o)'(\mathfrak{I}_n^o)(G_n^o)]^{-1}\} \tag{43}$$

(Gallant, 1987, Ch. 7, Sect. 4). Because

$$\{f_1[x_1|g(\rho)],\{f_t[y_t|x_t,g(\rho)]\}_{t=1}^{\infty}\} = \{p_1(x_1|\rho),\{p_t(y_t|x_t,\rho)\}_{t=1}^{\infty}\}, \tag{44}$$

$\hat{\rho}_{mle}$ is also the maximum likelihood estimator for the process $\{p_1(x_1|\rho), \{p_t(y_t|x_t,\rho)\}_{t=1}^{\infty}\}$.

If g cannot be computed, the minimum chi-square estimator is not practical. However, $\mathcal{E}(\tilde{\theta}_n|\rho)$ can be computed by simulation and the preceding remarks suggest that minimum chi-square with $\mathcal{E}(\tilde{\theta}_n|\rho)$ replacing $g(\rho)$ would be a practical, fully efficient estimator. See Gouriéroux, Monfort, and Renault (1993) and Smith (1993) for examples. The difficulty with this approach is that the simulated minimum chi-square estimator is computationally inefficient relative to the GMM estimator proposed here because at each of the N Monte Carlo repetitions in the expression $\mathcal{E}(\tilde{\theta}_n|\rho) = (1/N)\sum_{\tau=1}^{N}\tilde{\theta}_{n\tau}$ an optimization to compute $\tilde{\theta}_{n\tau}$ is required. The GMM estimator requires only the one optimization to compute $\tilde{\theta}_n$ and avoids the N extra optimizations required to compute $\mathcal{E}(\tilde{\theta}_n|\rho)$. Moreover, one would actually have to invoke Assumption 2 or estimate $\mathcal{I}_n^o$ to follow this approach. See Gouriéroux et al. (1993) for additional remarks on the relationships among various approaches.

We conclude this section by showing that $\hat{\rho}_n$ has the same asymptotic distribution as $\hat{\rho}_{mle}$.

THEOREM 2. *Assumption* 3 *implies*

$$\sqrt{n}(\hat{\rho}_n - \rho^o) \approx N\{0, [(G_n^o)'(\mathcal{I}_n^o)(G_n^o)]^{-1}\}. \tag{45}$$

Proof. From the first-order conditions

$$\begin{aligned} 0 &= (\partial/\partial\rho)\,[m_n'(\hat{\rho}_n,\tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1}m_n(\hat{\rho}_n,\tilde{\theta}_n)] \\ &= 2[(\partial/\partial\rho)m_n'(\hat{\rho}_n,\tilde{\theta}_n)]\,(\tilde{\mathcal{I}}_n)^{-1}m_n(\hat{\rho}_n,\tilde{\theta}_n), \end{aligned} \tag{46}$$

we have, after a Taylor's expansion of $m_n(\hat{\rho}_n,\tilde{\theta}_n)$,

$$[(\hat{M}_n)'(\tilde{\mathcal{I}}_n)^{-1}(\overline{M}_n)]\sqrt{n}(\hat{\rho}_n - \rho^o) = -[(\hat{M}_n)'(\tilde{\mathcal{I}}_n)^{-1}(\partial/\partial\theta)\overline{m}_n]\sqrt{n}(\tilde{\theta}_n - \theta^o), \tag{47}$$

where the overbars indicate that the rows of $M_n(\rho,\theta) = (\partial/\partial\rho')m_n(\rho,\theta)$ and $(\partial/\partial\theta')m_n(\rho,\theta)$ have been evaluated at points on the line segment joining $(\hat{\rho}_n,\tilde{\theta}_n)$ to (ρ^o,θ^o). Recall that $\hat{M}_n$ and M_n^o indicate evaluation of $M_n(\rho,\theta)$ at $(\hat{\rho}_n,\tilde{\theta}_n)$ and (ρ^o,θ^o), respectively. Now $\lim_{n\to\infty}(\overline{M}_n - M_n^o) = 0$, $\lim_{n\to\infty}(\hat{M}_n - M_n^o) = 0$, $\lim_{n\to\infty}[-(\partial/\partial\theta)\overline{m}_n - \mathcal{I}_n^o] = 0$, and $\lim_{n\to\infty}(\tilde{\mathcal{I}}_n - \mathcal{I}_n^o) = 0$ a.s. Furthermore,

$$\sqrt{n}(\tilde{\theta}_n - \theta^o) \approx N[0, (\mathcal{I}_n^o)^{-1}]. \tag{48}$$

Therefore, the preceding equation can be rewritten as

$$[(\hat{M}_n)'(\tilde{\mathcal{I}}_n)^{-1}(\overline{M}_n)]\sqrt{n}(\hat{\rho}_n - \rho^o) = (M_n^o)'\sqrt{n}(\tilde{\theta}_n - \theta^o) + o_p(1), \tag{49}$$

which implies that

$$\sqrt{n}(\hat{\rho}_n - \rho^o) \approx N\{0, [(M_n^o)'(\mathcal{I}_n^o)^{-1}(M_n^o)]^{-1}\}. \tag{50}$$

We complete the proof by showing that $M_n^o = \mathcal{I}_n^o G^o$.

$$
\begin{aligned}
M_n^o &= (\partial/\partial\rho')\,\frac{1}{n}\sum_{t=1}^{n}\int\cdots\int(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)\\
&\quad\times\prod_{\tau=1}^{n}p_\tau(y_\tau|x_\tau,\rho)\,dy_\tau\,p_1(x_1|\rho)\,dx_1\big|_{\rho=\rho^o}\\
&=(\partial/\partial\rho')\,\frac{1}{n}\sum_{t=1}^{n}\int\cdots\int(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)\\
&\quad\times\prod_{\tau=1}^{n}f_\tau[y_\tau|x_\tau,g(\rho)]\,dy_\tau\,p_1(x_1|\rho)\,dx_1\big|_{\rho=\rho^o}\\
&=\frac{1}{n}\sum_{t=1}^{n}\int\cdots\int(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)\\
&\quad\times\sum_{s=1}^{n}(\partial/\partial\rho')f_s[y_s|x_s,g(\rho)]_{\rho=\rho^o}\\
&\quad\times\prod_{\tau\neq s}^{n}f_\tau(y_\tau|x_\tau,\theta^o)\,dy_\tau\,dy_s\,p_1(x_1|\rho^o)\,dx_1\\
&\quad+\frac{1}{n}\sum_{t=1}^{n}\int\cdots\int(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)\\
&\quad\times\prod_{\tau=1}^{n}f_\tau(y_\tau|x_\tau,\theta^o)\,dy_\tau\,(\partial/\partial\rho')p_1(x_1|\rho)\big|_{\rho=\rho^o}dx_1\\
&=\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\int\cdots\int(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)\\
&\quad\times[(\partial/\partial\theta')f_s(y_s|x_s,\theta)\big|_{\theta=g(\rho)}G(\rho)]_{\rho=\rho^o}\\
&\quad\times\prod_{\tau\neq s}^{n}f_\tau(y_\tau|x_\tau,\theta^o)\,dy_\tau\,dy_s\,p_1(x_1|\rho^o)\,dx_1\\
&\quad+\frac{1}{n}\sum_{t=1}^{n}\int 0(\partial/\partial\rho')p_1(x_1|\rho)\big|_{\rho=\rho^o}dx_1\\
&=\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\int\cdots\int[(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)]\\
&\quad\times[(\partial/\partial\theta)\ln f_s(y_s|x_s,\theta^o)]'f_s(y_s|x_s,\theta^o)G^o\\
&\quad\times\prod_{\tau\neq s}^{n}f_\tau(y_\tau|x_\tau,\theta^o)\,dy_\tau\,dy_s\,p_1(x_1|\rho^o)\,dx_1\\
&=\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\int\cdots\int[(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)][(\partial/\partial\theta)\ln f_s(y_s|x_s,\theta^o)]'
\end{aligned}
$$

$$\times \prod_{\tau=1}^{n} f_\tau(y_\tau|x_\tau,\theta^o)\,dy_\tau\, p_1(x_1|\rho^o)\,dx_1\, G^o$$

$$= \frac{1}{n}\sum_{t=1}^{n} \int \cdots \int [(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)][(\partial/\partial\theta)\ln f_t(y_t|x_t,\theta^o)]'$$

$$\times \prod_{\tau=1}^{n} f_\tau(y_\tau|x_\tau,\theta^o)\,dy_\tau\, p_1(x_1|\rho^o)\,dx_1\, G^o$$

$$= \mathcal{I}_n^o G^o. \tag{51}$$

■

3. GENERAL PURPOSE SCORE GENERATORS

As pointed out in Section 2, if a model $\{f_1(x_1|\theta),\{f_t(y_t|x_t,\theta)\}_{t=1}^{\infty}\}_{\theta\in\Theta}$ is known to accurately describe the distribution of the data $\{\tilde{y}_t\}_{t=1}^{n}$, then that model should be the score generator that defines $m_n(\rho,\tilde{\theta}_n)$ and $\hat{\rho}_n$. If not, we can suggest two general purpose score generators.

The first is the SNP score, which can be expected to closely approximate any nonlinear Markovian process. An example of its use in connection with the estimator $\hat{\rho}_n$ proposed here is that by Bansal, Gallant, Hussey, and Tauchen (1995), who fit a general equilibrium, two-country, monetary model using high-frequency financial market data. The second is the neural net score, which can be expected to closely approximate any cross-sectional nonlinear regression or any dynamic nonlinear autoregression, including deterministic chaos. An example of its use in connection with $\hat{\rho}_n$ is that by Ellner, Gallant, and Theiler (1995), who use data widely believed to exhibit chaotic dynamics to calibrate the parameters of the SEIR model, which is a model of epidemics often used in health economics. The cited applications contain descriptions of the SNP and neural net scores, respectively.

In terms of convenience, what one would like is for

$$\int (\partial/\partial\theta)\ln f_t(y_t|x_t,\theta_n^o)p_t(y_t|x_t,\rho^o)\,dy_t \tag{52}$$

to be small enough that $\tilde{\mu}_t$ can be put to zero with little effect upon the accuracy of the computation of $\tilde{S}_{n\tau}$ and small enough that

$$\tilde{S}_{n\tau}, \qquad \tau \neq 0, \tag{53}$$

can be put to zero with little effect on the accuracy of the computation of $\tilde{\mathcal{I}}_n$. The estimator of $\mathcal{I}_n^o$ would then assume its simplest form:

$$\tilde{\mathcal{I}}_n = \frac{1}{n}\sum_{t=1}^{n} [(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\tilde{\theta}_n)][(\partial/\partial\theta)\ln f_t(\tilde{y}_t|\tilde{x}_t,\tilde{\theta}_n)]'. \tag{54}$$

Both SNP and neural nets are series expansions that have the property that (52) can be made arbitrarily small by using enough terms in the expansion

(Gallant and Nychka, 1987; Gallant and White, 1992). Hence, $\tilde{\mu}_t$ and (53) can be made arbitrarily small by using enough terms. The appropriate number of terms relative to the sample size are suggested by the results of Fenton and Gallant (1996) and McCaffrey and Gallant (1994). However, there is as yet no general theory giving the rate at which terms can be added so as to retain $\sqrt{n}$-asymptotic normality so one must guard against taking too many terms and then claiming that standard asymptotics apply.

4. APPLICATIONS

We discuss three classes of applications of the estimator developed in the previous sections. In the setup for each application, it is relatively simple to generate simulated realizations from the structural model while computation of the likelihood is infeasible. Hence, simulation and moment matching are appropriate estimation strategies.

4.1. Consumption and Asset Returns in a Production Economy

Consider the following version of the Brock–Mirman one-sector setup. The representative agent's problem is

$$\max \mathcal{E}_t\left[\frac{1}{1-\gamma}\sum_{i=0}^{\infty}\beta^i c_{t+i}^{1-\gamma} v_{2,t+i}\right] \tag{55}$$

subject to

$$c_t + k_{t+1} - k_t \le A k_t^{\alpha} v_{1t}, \tag{56}$$

where c_t is consumption at time t and k_t the capital stock at the beginning of period t (i.e., inherited from period $t-1$); v_{1t} and v_{2t} are strictly positive shocks to technology and preferences; $\mathcal{E}_t(\cdot)$ is shorthand for the conditional information given all variables in the model dated time t and earlier; and the parameters satisfy $0 < \beta < 1$, $\gamma \ge 0$, $A > 0$, and $0 < \alpha \le 1$. The agent's choice variables at time t are c_t and k_{t+1}. The stochastic process $v_t = (v_{1t}, v_{2t})'$ is strictly stationary and Markovian of order r, with conditional density $\phi(v_{t+1} \mid v_t^*, \delta)$, where $v_t^* = (v_t', \ldots, v_{t-r}')'$ and δ is a parameter vector.

The Euler equation for this problem is

$$c_t^{-\gamma} = \mathcal{E}_t(\beta c_{t+1}^{-\gamma} \alpha A k_{t+1}^{\alpha-1}). \tag{57}$$

The solution of the optimization problem is

$$k_{t+1} = \psi_k(k_t, v_t^*), \tag{58}$$

$$c_t = \psi_c(k_t, v_t^*), \tag{59}$$

where ψ_c and ψ_k are the policy functions.

There is no known closed form solution for the policy functions, though the policy functions can be well approximated using one of the newly developed methods for solving nonlinear rational expectations models. The 1990 symposium in the *Journal of Business and Economic Statistics* (Tauchen, 1990) surveys many of the extant methods. For this model, and the proposed application, the method of Coleman (1990), which uses quadrature for numerical integration and computes the policy function over an extremely fine grid, is probably the most accurate and numerically efficient.

Using Coleman's method to evaluate the policy functions, one can then easily simulate from this model. Given an initial value k^o for the capital stock, and a simulated realization $\{\hat{v}_\tau\}$ generated from $\phi(v|v*,\delta)$, one generates simulated $\{\hat{k}_\tau, \hat{c}_\tau\}$ by recursively feeding the $\hat{v}_\tau$ and $\hat{k}_\tau$ through the policy function for capital. Good practice is to allow the iterations to run for a long while in order to let the effects of the transients wear off. A simulated realization of length N, $\{\hat{k}_\tau, \hat{v}_\tau\}_{\tau=1}^N$, would be the last N values of the iterations.

Strategies to implement empirically the corresponding competitive equilibrium of this model differ depending on which variables are used to confront the model to data, that is, which variables enter the score generator. For example, with good data on both consumption and capital, the researcher could use $(c_t, k_t)'$. However, if capital is poorly measured but output well measured, then it would be better to use $(c_t, q_t)'$, where q_t is total output, which in the simulation would be computed as $\hat{q}_\tau = A\hat{k}_\tau^\alpha \hat{v}_{1\tau}$. Neither of these strategies, though, makes use of price data.

A strategy that incorporates price information is to use c_t along with the returns on a pure discount risk-free bond, r_{bt}, and a stock, r_{st}. Asset returns are determined via asset pricing calculations, carried out as follows. (It turns out to be a bit easier to think of the equations defining returns between t and $t+1$.) The bond return, $r_{b,t+1}$, is the solution to

$$c_t^{-\gamma} = \mathcal{E}_t(\beta c_{t+1}^{-\gamma})(1 + r_{b,t+1}), \tag{60}$$

and $r_{b,t+1}$ is known to agents at time t. For the stock return, the dividend process is $d_{st} = Ak_t^\alpha v_{1t} - r_{b,t}k_t$, and the stock price process $\{p_{st}\}$ is the solution to the expectational equation

$$p_{st}c_t^{-\gamma} = \mathcal{E}_t[\beta c_{t+1}^{-\gamma}(p_{s,t+1} + d_{s,t+1})]. \tag{61}$$

The stock return between t and $t+1$ is $r_{s,t+1} = (p_{s,t+1} + d_{s,t+1})/p_{st}$. Solving for the asset returns entails additional computation that could potentially be as numerically intensive as approximating the policy functions.

This formulation presumes that, in the competitive equilibrium, the firm uses 100% debt financing to rent from a household the capital stock k_{t+1} for one period at interest rate $r_{b,t+1}$. (Both $r_{b,t+1}$ and k_{t+1} are determined and known at time t.) The firm distributes to the household as the dividend $d_{s,t+1} = Ak_{t+1}^\alpha v_{1,t+1} - r_{b,t+1}k_{t+1}$, which is the firm's cash flow in period $t+1$,

that is, the proceeds after paying off the bondholder. Other conceptualizations are possible, and, in particular, the stock price and returns process could be different if the firm retains earnings or uses different forms of debt financing.

One typically does not observe a risk-free real bond return. Common practice in empirical asset pricing is to use the consumption series along with either the real ex-post return on the stock (deflated using a price index) or the excess of the stock return over the bond return, $r_{et} = r_{st} - r_{bt}$, from which inflation cancels out. This practice presumes that the observed data come from a monetary economy with exactly the same real side as above and a nominal side characterized by a binding cash-in-advance constraint, which implies unitary monetary velocity.

We show how to implement the estimator on data consisting of consumption and the excess stock return. This is done for illustrative purposes. The proposal offers an alternative to the standard SMM strategy of selecting out a set of low-order moments, as in Gennotte and Marsh (1993), for estimation of an asset pricing model. In actual practice, one would want to employ more sophisticated versions of the model with time nonseparabilities in consumption and production and also include additional latent taste and/or technology shocks when additional asset returns are observed. Common practice in stochastic modeling is to include sufficient shocks or measurement errors to preclude the predicted distribution of the data from being concentrated on a lower dimensional manifold, which is normally counterfactual, and the model being dismissed out of hand immediately.

Put $y_t = (r_{et}, c_t)'$. Let $\rho = (\gamma, A, \alpha, \delta')'$ denote the vector of structural parameters. The numerical solution of the model provides a means to simulate data given a value of ρ.

Experience with financial data suggests that a reasonable choice for the score generator is the sequence of densities defined by an ARCH (Engle, 1982) or GARCH (Engle and Bollerslev, 1986) process. For ease of exposition, we show the ARCH case. Consider the multivariate ARCH model

$$y_t = b_0 + \sum_{j=1}^{L_1} B_j y_{t-j} + u_t, \tag{62}$$

$$\operatorname{vech}(\Sigma_t) = c_0 + \sum_{j=1}^{L_2} C_j \operatorname{vech}(u_{t-j} \otimes u_{t-j}), \tag{63}$$

where $u_t \sim N(0, \Sigma_t)$. Let $\operatorname{pdf}(y_t | x_t, \psi)$ denote the implied conditional density of $\{y_t\}$ under the ARCH model, where $x_t = (y'_{t-L}, \ldots, y'_{t-1})'$, $L = L_1 + L_2$, and

$$\psi = (b'_0, \operatorname{vec}([B_1 B_2 \ldots B_{L_1}])', c'_0, \operatorname{vec}([C_1 C_2 \ldots C_{L_2}])')'. \tag{64}$$

Common practice in ARCH modeling is to impose a priori restrictions, such as diagonality or factor restrictions, so as to constrain $\psi = \psi(\theta)$ to depend

on a lower dimensional parameter, θ. Let $f(y_t|x_t,\theta) = \text{pdf}(y_t|x_t,\psi[\theta])$ denote the ARCH conditional density under the restrictions, which we take as the score generator.

Given the observed data set $\{\tilde{y}_t\}_{t=1}^n$, the first step in the estimation is to apply quasimaximum likelihood to the ARCH model

$$\tilde{\theta}_n = \underset{\theta\in\Theta}{\text{argmax}} \frac{1}{n}\sum_{t=1}^{n} \ln f_t(\tilde{y}_t|\tilde{x}_t,\theta). \tag{65}$$

The second step is to estimate ρ by

$$\hat{\rho} = \underset{\rho\in R}{\text{argmin}}\, m_n'(\rho,\tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1}m_n(\rho,\tilde{\theta}_n), \tag{66}$$

where $m_n(\rho,\tilde{\theta}_n) = (1/N)\sum_{\tau=1}^{N}(\partial/\partial\theta)\ln f_t(\hat{y}_\tau|\hat{x}_\tau,\tilde{\theta}_n)$, $\{\hat{y}_\tau\}$ is a simulated realization given ρ from the model, and $\hat{x}_\tau = (\hat{y}_{t-L}',\ldots,\hat{y}_{t-1}')'$.

The relevant asymptotic distribution theory is that of Case 2 in Section 2. The order condition for identification is that the length of θ be at least as long as the length of ρ. The analog of the rank condition is given in the discussion following Theorem 1. It is exceedingly difficult to determine analytically whether the ARCH scores suffice to identify the asset pricing model, which, as noted earlier, is typical of nonlinear statistics. In practice, near flat spots in the sample objective function would be a strong indicator of failure of identification. In such a case, further expansion of the score generator such as relaxing conditional normality or using a non-Markov (GARCH) model could bring in additional score components to achieve identification. The mechanics of implementing a GARCH-type score generator are similar to that just described for ARCH, though the notation is more cumbersome. In either case, use of this estimator provides a means to bring to bear on the task of selecting moments the knowledge that ARCH-GARCH models fit returns data well.

Another possible score generator would be to use the SNP model of Gallant and Tauchen (1989, 1992); this strategy is employed by Bansal et al. (1995) for estimation of a model of weekly currency market data. Use of an SNP model would give the exercise a more nonparametric slant, as the choice of dimension of the score generator model would be data-determined. Either choice would ensure efficiency against a class of models known to capture much of the first and second moment dynamics of asset prices and other macro aggregates.

4.2. Stochastic Volatility

Consider the stochastic volatility model

$$y_t - \mu_y = c(y_{t-1} - \mu_y) + \exp(w_t)r_y z_t,$$
$$w_t = aw_{t-1} + r_w\tilde{z}_t.$$

The first equation is the mean equation with parameters μ_y, c, and r_y; the second is the volatility equation with parameters a and r_w. $\{y_t\}$ is an observed financial returns process and $\{w_t\}$ is an unobserved volatility process. In the basic specification, z_t and $\tilde{z}_t$ are mutually independent iid $N(0,1)$ shocks. The model can be generalized in an obvious way to accommodate longer lag lengths in either equation. Versions of this model have been examined by Clark (1973), Melino and Turnbull (1990), Harvey, Ruiz, and Shephard (1993), Jacquier, Polson, and Rossi (1994), and many others. The appeal of the model is that it provides a simple specification for speculative price movements that accounts, in qualitative terms, for broad general features of data from financial markets such as leptokurtosis and persistent volatility. The complicating factor for estimation is that the likelihood function is not readily available in closed form, which motivates consideration of other approaches.

Gallant, Hsieh, and Tauchen (1994) employ the estimator of this paper to estimate the stochastic volatility model on a long time series comprised of 16,127 daily observations $\{\tilde{y}_t\}_{t=1}^{16,127}$ on adjusted movements in the Standard and Poor's Composite Index, 1928–1987. The score generator is an SNP model, as described in Section 3. The specification search for appropriate auxiliary models for $\{\tilde{y}_t\}_{t=1}^{16,127}$ leads to two scores: a "Nonparametric ARCH Score," when errors are constrained to be homogeneous, and a "Nonlinear Nonparametric Score," when errors are allowed to be conditionally heterogeneous. The Nonparametric ARCH Score contains indicators for both deviations from conditional normality and ARCH. Together, these scores suffice to identify the stochastic volatility model; indeed, the stochastic volatility model places overidentifying restrictions across these scores. The Nonlinear Nonparametric ARCH Score contains additional indicators for conditional heterogeneity, most importantly, the leverage type effect of Nelson (1991), which is a form of dynamic asymmetry. These additional indicators identify dynamic asymmetries like those suggested by Harvey and Shephard (1993), which the Nonparametric ARCH Score does not identify. When fitted to either of these two scores, the standard stochastic volatility model fails to approximate the distribution of the data adequately; it is overwhelmingly rejected on the chi-square goodness-of-fit tests. After altering the distribution of z_t to accommodate thickness in both tails along with left skewness and generalizing the volatility equation to include long memory (Harvey, 1993), the stochastic volatility model can match the moments defined by the simpler Nonparametric ARCH Score, but not those defined by the Nonlinear Nonparametric Score. Introducing cross-correlation between z_{t-1} and $\tilde{z}_t$ as in Harvey and Shephard (1993) improves the fit to the Nonlinear Nonparametric Score substantially, but still the stochastic volatility model cannot fit that score. Overall, Gallant et al. (1994) find the estimation provides a computationally tractable means to assess the relative plausibility of a wide class of alternative specifications of the stochastic volatility model. They show how

to use the score vector of a rejected model to elucidate useful diagnostic information.

There are other ongoing applications of the estimator in the context of stochastic volatility. Engle (1994) employs it to estimate a continuous time stochastic volatility model, with the score generator being a GARCH model fitted to the discrete time data. Ghysels and Jasiak (1994) use it to estimate a continuous time model of stock returns and volume subject to time deformation like that of Clark (1973) and Tauchen and Pitts (1983). Their score generator is an SNP model very similar to that of Gallant, Rossi, and Tauchen (1992) fitted to the discrete time returns and volume data.

4.3. Empirical Modeling of Auction Data

Auctions are commonly used to sell assets. Game theoretic models of auctions provide a detailed theory of the mapping from the disparate values that bidders place on the asset to the final outcome (the winner and the sales price). The predictions of this theory depend strongly on the assumptions regarding the characteristics of the auction and the bidders. Generally, the specific rules of the auction along with the information structure, the attitudes of the bidders toward risk, and the bidders' strategic behavior all matter a great deal in determining the final outcome (Milgrom, 1986).

Empirical implementation of game theoretic models of auctions lags well behind the theory. The extreme nonlinearities and numerical complexity of auction models presents substantial obstacles to direct implementation. Two recent papers, by Paarsch (1991) and Laffont, Ossard, and Vuong (1991) make substantial progress, however. In both papers, the task is to estimate the parameters of the distribution of values across bidders. Paarsch develops a framework based on standard maximum likelihood. His approach can handle a variety of informational environments but is restricted to a relatively narrow set of parametric models for the valuation distribution—essentially the Pareto and Weibull. Laffont et al. use a simulation approach, and they can thereby handle a much broader class of valuation distributions. However, their approach imposes only the predictions of the theory regarding first moments and ignores higher order structure, which can cause problems of inefficiency and identification.

The method set forth in Section 2 imposes all restrictions and generates an efficient estimate of the valuation distribution. In what follows, we illustrate how one would implement the method for some of the simpler models of auctions. A full empirical study would go much further and, in particular, would relax our strong assumptions and consider other environments known to be theoretically important.

We first provide a short overview of some of the simplest auction models and then proceed to the econometrics.

Two auction models under independent private valuations. An item, such as a tract of land or stand of timber, is to be sold at auction. The item will be sold so long as a selling price at least as large as a reservation price $r_0 > 0$ is realized; otherwise, it is left unsold.

There are two commonly used auction designs. In an *oral ascending auction*, the selling value of the item starts at r_0 and then increases. Bidders drop out as the selling value rises until one bidder remains, who pays the selling value at which the last of the other bidders dropped out. In a *sealed bid first price auction*, all bids are collected simultaneously. The object is sold to the highest bidder, who pays his bid so long as it exceeds the reservation price.

The independent private value paradigm is a set of assumptions regarding the characteristics of bidders; the paradigm is applied to either type of auction. In this paradigm, each of B bidders is assumed to have a private valuation, v_i, $i = 1,2,\ldots,B$, for the item to be sold. Each bidder knows his or her own private valuation but does not know the valuation of other bidders. The bidders act as if the B valuations are i.i.d. drawings from a common valuation distribution $H(v|q,\rho)$, with density $h(v|q,\rho)$, where q is a vector of covariates defining characteristics of the item to be sold and ρ is a parameter vector. Each bidder knows q,ρ, the functional form of $H(v|q,\rho)$, and the reservation price, r_0. Also, each bidder is assumed to be risk-neutral and the equilibrium concept is the symmetric Bayesian Nash equilibrium.

For the oral ascending auction, the winning bid, y, is

$$y = \max[v_{(B-1:B)}, r_0] I(v_{(B:B)} \geq r_0) \quad \text{if } B \geq 2, \tag{67}$$

$$y = r_0 I(v_1 \geq r_0) \quad \text{if } B = 1, \tag{68}$$

where $v_{(1:B)} \leq \cdots \leq v_{(B:B)}$ are the order statistics of $v_1,\ldots,v_B$, and $I(\cdot)$ is the zero-one indicator function. On the event $v_{(B:B)} < r_0$, the winning bid is defined as zero and the item is unsold.

Let $p_{oa}(y|r_0,B,q,\rho)$, or simply $p_{oa}(y|x,\rho)$ with $x = (r_0,B,q)$, denote the conditional probability density of the winning bid. Below, we write either $p_{oa}(y|r_0,B,q,\rho)$ or $p_{oa}(y|x,\rho)$, depending on whether or not we wish to emphasize dependence on each of the different components of x. In general, $p_{oa}(y|x,\rho)$ is an ordinary density on the region $y > r_0$, so long as $h(v|q,\rho)$ is smooth, whereas $p_{oa}(y|x,\rho)$ has atoms at $y = 0$ and $y = r_0$. In certain circumstances—for example, $h(v|q,\rho)$ is Pareto or Weibull as in Paarsch (1991)—$p_{oa}(y|x,\rho)$ has a manageable closed-form expression. In other circumstances—for example, $h(v|q,\rho)$ is lognormal as in Laffont et al. (1991)—$p_{oa}(y|x,\rho)$ admits no tractable expression. However, so long as it is easy to simulate from $h(v|q,\rho)$, then it is easy to simulate from $p_{oa}(y|x,\rho)$.

For the sealed bid first price auction, the winning bid is

$$y = E\{\max(v_{(B-1:B)}, r_0) | v_{(B:B)}\} I(v_{(B:B)} \geq r_0) \quad \text{if } B \geq 2, \tag{69}$$

$$y = r_0 I(v_1 \geq r_0) \quad \text{if } B = 1. \tag{70}$$

Thus, when there are two or more bidders and $v_{(B:B)} \geq r_0$, then the winning bid follows the distribution of the conditional expectation of $\max(v_{(B-1:B)}, r_0)$ given $v_{(B:B)}$. Let $p_{sp}(y|r_0, B, q, \rho)$, or $p_{sp}(y|x, \rho)$, denote the implied conditional density of the winning bid in the sealed bid case.

Generally, $p_{sp}(y|x, \rho)$ is less manageable in practice than is $p_{oa}(y|x, \rho)$. Generation of a simulated draw from $p_{sp}(y|x, \rho)$ entails either numerical integration of the cumulative distribution of the valuation distribution or a double-nested set of simulations.

A new estimation strategy for auction models. Suppose an econometrician observes $\{\tilde{y}_t, \tilde{x}_t\}_{t=1}^n$, where $\tilde{y}_t$ is the winning bid and $\tilde{x}_t = (\tilde{r}_{0t}, \tilde{B}_t, \tilde{q}_t)$ contains the reservation price, the number of bidders, and covariates for each of n auctions. In what follows, we take the auctions to be oral ascending auctions and point out, where appropriate, how things differ for sealed bid auctions. The econometrician assumes that the same valuation density, $h(v|q, \rho^o)$, describes the bidder valuations for each auction. The analysis is conditional (the x's are strictly exogenous); the econometrician assumes that $\tilde{y}_t$ and $\tilde{y}_s$ are statistically independent for $t \neq s$, conditional on the sequence $\{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n\}$. The task is to estimate the true underlying parameter vector, ρ^o.

One estimation strategy is straight maximum likelihood. Under special distributional assumptions on the valuation density such as Weibull or Pareto, the conditional density of the winning bid $p_{oa}(y|r_0, B, q, \rho)$ has a manageable closed form. Conventional maximum likelihood estimation can then be undertaken. This is the strategy of Paarsch (1991).

Laffont et al. (1991) developed a simulated nonlinear least squares (SNLLS) estimator that can handle a broader class of parent densities for the valuation distribution. Their approach is to apply nonlinear least squares:

$$\tilde{\rho} = \underset{\rho \in R}{\operatorname{argmin}} \left(\sum_{t=1}^{n} [\tilde{y}_t - \mu_{oa}(\tilde{r}_{0t}, \tilde{B}_t, \tilde{q}_t, \rho)]^2 \right), \tag{71}$$

where $\mu_{oa}(\tilde{r}_{0t}, \tilde{B}_t, \tilde{q}_t, \rho) = \int y p_{oa}(y|\tilde{r}_{0t}, \tilde{B}_t, \tilde{q}_t, \rho)\, dy$. In practice, $\mu_{oa}(\tilde{r}_{0t}, \tilde{B}_t, \tilde{q}_t, \rho)$ is approximated via Monte Carlo integration:

$$\mu_{oa}(\tilde{r}_{0t}, \tilde{B}_t, \tilde{q}_t, \rho) \approx \frac{1}{N} \sum_{\tau=1}^{N} \max(v_{\tau,(\tilde{B}_t-1:\tilde{B}_t)}, \tilde{r}_{0t}), \tag{72}$$

where $v_{\tau,(\tilde{B}_t-1:\tilde{B}_t)}$ is the second highest order statistic of the τth independent simulated realization of $(v_{\tau 1}, \ldots, v_{\tau \tilde{B}_t})$ i.i.d. from $h(v|\tilde{q}_t, \rho)$. In their motivating examples and empirical applications, v is conditionally lognormal with a mean that depends on $\tilde{q}_t$, and ρ contains the parameters of this conditional lognormal distribution. The SNLLS estimator is nonlinear least squares with a heteroskedasticity-robust estimate of the asymptotic variance of $\tilde{\rho}$ that accounts for conditional heteroskedasticity of

$$\epsilon_t = \tilde{y}_t - \mu_{oa}(\tilde{r}_{0t}, \tilde{B}_t, \tilde{q}_t, \rho^o). \tag{73}$$

Laffont et al. (1991) noted that revenue equivalence implies the same formulation of the conditional mean function applies for a sealed bid auction. Revenue equivalence implies

$$\begin{aligned}\mu_{oa}(r_0, B, q, \rho) &= \int y p_{oa}(y|r_0, B, q, \rho)\, dy \\ &= \int y p_{sb}(y|r_0, B, q, \rho)\, dy \\ &= \mu_{sb}(r_0, B, q, \rho) \end{aligned} \tag{74}$$

for all r_0, B, q, and ρ. Hence, one can evaluate the conditional mean function at the data, that is, compute $\mu_{sb}(\tilde{r}_{0t}, \tilde{B}_t, \tilde{q}_t, \rho)$, by simulating and averaging exactly as one does under oral ascending rules. The result can be a significant reduction in computational demands.

The SNLLS approach works off of the conditional first moment implications alone, though, and auction models place additional structure on the data. An auction model has second moment implications as well as first moment implications. In fact, it actually dictates the functional form of the conditional heteroskedasticity in the nonlinear regression equation, which suggests additional moment conditions. There are practical consequences from not incorporating additional restrictions beyond first moment information. Laffont et al. (1991) and Baldwin (1992) find it difficult to estimate the variance of the underlying parent lognormal using SNLLS. Bringing in second moment estimation can be expected to alleviate this difficulty. In general, there are further implications beyond first and second moments as well; imposition of all implications of the model can be expected to sharpen even further the estimates of the parameter ρ.

Ideally, one wants to do this by doing maximum likelihood using either $p_{oa}(y|r_0, B, q, \rho)$ or $p_{sb}(y|r_0, B, q, \rho)$ as appropriate to define the likelihood. The difficulty is that both densities are intractable, except in the special circumstances assumed by Paarsch (1991).

The approach outlined in Section 2 can come close to the maximum likelihood ideal. Our analysis pertains to the just-described situation where the likelihood is smooth but intractable; it does not cover cases where the likelihood is nondifferentiable in parameters. The consistency of the estimator $\hat{\rho}_n$ is not affected by nondifferentiability but asymptotic normality may be. See Hansen, Heaton, and Luttmer (1995, Appendix C) for a discussion of differentiability considerations with respect to GMM estimators.

The approach would be applied to the auction data as follows. $\tilde{\theta}_n$ is obtained as

$$\tilde{\theta}_n = \underset{\theta}{\operatorname{argmax}} \left(\frac{1}{n} \sum_{t=1}^{n} \ln f[\tilde{y}_t | \tilde{x}_t, \theta] \right), \tag{75}$$

where $f(y|x,\theta)$ is a score generator that gives a good approximation to the conditional distribution of y given the exogenous variables. One choice for $f(y|x,\theta)$ is a truncated Hermite expansion, or SNP model of Gallant and Tauchen (1992), which has been found in practice to be sufficiently flexible to approximate well a wide class of densities. Note that $f(y|x,\theta)$ does not have to smoothly embed $p_{oa}(y|x,\rho)$, although, if it does, then the estimator is equally efficient as maximum likelihood.

Our estimator is GMM using the score function of the θ estimation to define the moment conditions:

$$\tilde{\rho} = \underset{\rho \in R}{\operatorname{argmin}}\, m_n'(\rho,\tilde{\theta}_n)(\tilde{\mathfrak{I}}_n)^{-1} m_n(\rho,\tilde{\theta}_n), \tag{76}$$

where

$$m_n(\rho,\tilde{\theta}_n) = (1/n)\sum_{t=1}^{n}(1/N)\sum_{\tau=1}^{N}(\partial/\partial\theta)\ln f_t(\hat{y}_{\tau t}|\tilde{x}_t,\tilde{\theta}_n) \tag{77}$$

and where, for each t, $\{\hat{y}_{\tau t}\}_{\tau=1}^{N}$ is a simulated realization of length N from either $p_{oa}(y|\tilde{r}_{0t},\tilde{B}_t,\tilde{q}_t,\rho)$ or $p_{sb}(y|\tilde{r}_{0t},\tilde{B}_t,\tilde{q}_t,\rho)$, depending on whether the data are from an oral ascending or sealed bid auction. Sampling from $p_{oa}(y|\tilde{r}_{0t},\tilde{B}_t,\tilde{q}_t,\rho)$ is relatively easy while sampling from $p_{sb}(y|\tilde{r}_{0t},\tilde{B}_t,\tilde{q}_t,\rho)$ is more difficult. (The revenue equivalence property only simplifies the sampling for the conditional first moment.)

The appropriate asymptotic theory for this estimator is Case 1, as the entire analysis is conditional on the realization of the strictly exogenous process $\{\tilde{x}_t\}$. To the extent $f(y|x,\theta)$ provides a good approximation of the distribution of y given the exogenous variables, then this estimator will have efficiency close to that of maximum likelihood.

REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 307–346.

Baldwin, L. (1992) *Essays on Auctions and Procurement.* Ph.D. Dissertation, Duke University.

Bansal, R., A.R. Gallant, R. Hussey, & G. Tauchen (1995) Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics* 66, 251–287.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Bollerslev, T. (1987) A conditionally heteroskedastic time series model of speculative prices and rates of return. *Review of Economics and Statistics* 64, 542–547.

Bollerslev, T. & J.M. Wooldridge (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11, 143–172.

Clark, P.K. (1973) A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41, 135–56.

Coleman, J. (1990) Solving the stochastic growth model by policy-function iteration. *Journal of Business and Economic Statistics* 8, 27–30.

Duffie, D. & K.J. Singleton (1993) Simulated moments estimation of Markov models of asset prices. *Econometrica* 61, 929–952.

Ellner, S., A.R. Gallant, & J. Theiler (1995) Detecting nonlinearity and chaos in epidemic data. In D. Mollison (ed.), *Epidemic Models: Their Structure and Relation to Data*, pp. 54–78. Cambridge: Cambridge University Press.

Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.

Engle, R.F. (1994) Indirect Inference on Volatility Diffusions and Stochastic Volatility Models. Manuscript, University of California at San Diego.

Engle, R.F. & T. Bollerslev (1986) Modeling the persistence of conditional variance. *Econometric Reviews* 5, 1–50.

Fenton, V. & A.R. Gallant (1996) Convergence rates of SNP density estimators. *Econometrica* 64, 719–727.

Gallant, A.R. (1977) Three stage least squares estimation for a system of simultaneous, nonlinear, implicit equations. *Journal of Econometrics* 5, 71–88.

Gallant, A.R. (1987) *Nonlinear Statistical Models.* New York: Wiley.

Gallant, A.R., D.A. Hsieh, & G. Tauchen (1994) Estimation of stochastic volatility models with diagnostics. Manuscript, Duke University.

Gallant, A.R. & D.W. Nychka (1987) Semi-nonparametric maximum likelihood estimation. *Econometrica* 55, 363–390.

Gallant, A.R., P.E. Rossi, & G. Tauchen (1992) Stock prices and volume. *Review of Financial Studies* 5, 199–242.

Gallant, A.R. & G. Tauchen (1989) Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica* 57, 1091–1120.

Gallant, A.R. & G. Tauchen (1992) A nonparametric approach to nonlinear time series analysis: Estimation and simulation. In E. Parzen, D. Brillinger, M. Rosenblatt, M. Taqqu, J. Geweke, & P. Caines (eds.), *New Dimensions in Time Series Analysis*, pp. 71–92. New York: Springer-Verlag.

Gallant, A.R. & H. White (1992) On learning the derivatives of an unknown mapping with multilayer feedforward networks. *Neural Networks* 5, 129–138.

Gennotte, G. & T.A. Marsh (1993) Variations in economic uncertainty and risk premiums on capital assets. *European Economic Review* 37, 1021–1041.

Ghysels, E. & J.J. Jasiak (1994) Stochastic volatility and time deformation: An application to trading volume and leverage effects. Manuscript, University of Montreal.

Gouriéroux, C., A. Monfort, & E. Renault (1993) Indirect inference. *Journal of Applied Econometrics* 8, S85–S118.

Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

Hansen, L.P., J. Heaton, & E.J.G. Luttmer (1995) Econometric evaluation of asset pricing models. *The Review of Financial Studies* 8, 237–274.

Harvey, A.C. (1993) Long Memory in Stochastic Volatility. Manuscript, London School of Economics.

Harvey, A.C., E. Ruiz, & N. Shephard (1993) Multivariate stochastic variance models. *Review of Economic Studies* 61, 247–264.

Harvey, A.C. & N. Shephard (1993) Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns. Manuscript, London School of Economics.

Jacquier, E., N.G. Polson, & P.E. Rossi (1994) Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 12, 371–388.

Laffont, J.-J., H. Ossard, & Q. Vuong (1991) The Econometrics of First-Price Auctions. Document de Travail 7, Institut d'Economie Industrielle, Toulouse.

McCaffrey, D.F. & A.R. Gallant (1994) Convergence rates for single hidden layer feedforward networks. *Neural Networks* 7, 147–158.

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

Melino, A. & S.M. Turnbull (1990) Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45, 239–266.

Milgrom, P. (1986) Auction theory. In T. Bewley (ed.), *Advances in Economic Theory*, pp. 1–32. Cambridge: Cambridge University Press.

Nelson, D. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347–370.

Paarsch, H.J. (1991) Empirical Models of Auctions and an Application to British Columbian Timber Sales. Discussion paper 91-19, University of British Columbia.

Pakes, A. & D. Pollard (1989) Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027–1058.

Smith, A.A. (1993) Estimating nonlinear time series models using vector autoregressions: Two approaches. *Journal of Applied Econometrics* 8, 63–84.

Tauchen, G. (1990) Associate editor's introduction. *Journal of Business and Economic Statistics* 8, 1-1.

Tauchen, G. & M. Pitts (1983) The price variability–volume relationship on speculative markets. *Econometrica* 51, 485–505.