Exact Maximum Likelihood Estimation of Observation-Driven Econometric Models

Francis X. Diebold

Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297

Til Schuermann

AT&T Bell Labs
600 Mountain Avenue
Room 7E-530
Murray Hill, NJ 07974

Revised February 1996

Abstract: The possibility of exact maximum likelihood estimation of many observation-driven models remains an open question. Often only approximate maximum likelihood estimation is attempted, because the unconditional density needed for exact estimation is not known in closed form. Using simulation and nonparametric density estimation techniques that facilitate empirical likelihood evaluation, we develop an exact maximum likelihood procedure. We provide an illustrative application to the estimation of ARCH models, in which we compare the sampling properties of the exact estimator to those of several competitors. We find that, especially in situations of small samples and high persistence, efficiency gains are obtained. We conclude with a discussion of directions for future research, including application of our methods to panel data models.

Acknowledgments: This is a revised and extended version of our earlier paper, "Exact Maximum Likelihood Estimation of ARCH Models." Helpful comments were provided by Fabio Canova, Rob Engle, John Geweke, Werner Ploberger, Doug Steigerwald, and seminar participants at Johns Hopkins University and the North American Winter Meetings of the Econometric Society. All errors remain ours alone. We gratefully acknowledge support from the National Science Foundation, the Sloan Foundation, the University of

Pennsylvania Research Foundation, and the Cornell National Supercomputer Facility.

1. Introduction

Cox (1981) makes the insightful distinction between observation-driven and parameter-driven models. A model is observation-driven if it is of the form

$$y_t = f(y^{(t-1)}, \varepsilon_t),$$

and parameter-driven if it is of the form

$$y_t = h(\phi_t, \nu_t)$$

$$\phi_t = g(\phi^{(t-1)}, \eta_t),$$

where superscripts denote past histories, and ε_t, ν_t and η_t are white noise. If, moreover, the relevant part of $y^{(t-1)}$ is of finite dimension, we will call an observation-driven model finite-ordered, and similarly if the relevant part of $\phi^{(t-1)}$ is of finite dimension, we will call a parameter-driven model finite-ordered.

Of course the distinction is only conceptual, as various state-space and filtering techniques enable movement from one representation to another, but the idea of cataloging models as observation- or parameter-driven facilitates interpretation and provides perspective. The key insight is that observation-driven models are often easy to estimate, because their dynamics are defined directly in terms of observables, but they are often hard to manipulate. In contrast, the nonlinear state-space form of parameter-driven models makes them easy to manipulate but hard to estimate.

A simple comparison of ARCH and stochastic volatility models will clarify the concepts.[1] Consider the first-order ARCH model,

$$y_t = \sigma_t \varepsilon_t$$

$$\varepsilon_t \overset{iid}{\sim} N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2,$$

so that

$$y_t | y_{t-1} \sim N(0, \alpha_0 + \alpha_1 y_{t-1}^2).$$

The model is finite-ordered and observation-driven and, as is well-known (e.g., Engle, 1982), it is easy to estimate by (approximate) maximum likelihood. Alternatively, consider the first-order stochastic volatility model,

$$y_t = \sigma_t \nu_t$$

$$\nu_t \overset{iid}{\sim} N(0,1)$$

$$\ln\sigma_t^2 = \delta_0 + \delta_1 \sigma_{t-1}^2 + \eta_t$$

$$\eta_t \overset{iid}{\sim} N(0,1),$$

so that

$$y_t | \sigma_{t-1} \sim N(0, \exp(\delta_0 + \delta_1 \sigma_{t-1}^2 + \eta_t)).$$

The model is finite-ordered but parameter-driven and, as is also well-known, it is very difficult to construct the likelihood because σ_t is unobserved.

---

[1] This example draws upon Shephard's (1995) insightful survey.

In this paper we study finite-ordered observation-driven models. This of course involves some loss of generality, as some interesting models (like the stochastic volatility model) are not observation-driven and/or finite-ordered, but finite-ordered observation-driven models are nevertheless tremendously important and popular. Autoregressive models and ARCH models, for example, satisfy the requisite criteria, as do many more complex models. Moreover, observation-driven counterparts of parameter-driven models often exist, such as Gray's (1995) version of Hamilton's (1989) Markov switching model.

Observation-driven models are often easy to estimate. The likelihood may be evaluated by prediction-error factorization, because the model is stated in terms of conditional densities that depend only on a finite number of past observables. The initial marginal term is typically discarded, however, as it can be difficult to determine and is of no asymptotic consequence in stationary environments, thereby rendering such "maximum likelihood" estimates *approximate* rather than *exact*. Because of the potential for efficiency gains, particularly in small samples with high persistence, exact maximum likelihood estimation may be preferable.

We will develop an exact maximum likelihood procedure for finite-ordered observation-driven models, and we will illustrate its feasibility and examine its sampling properties in the context ARCH models. Our procedure makes key use of simulation and nonparametric density estimation techniques to facilitate evaluation of the exact likelihood, and it is applicable quite generally

to any finite-ordered observation-driven model specified in terms of conditional densities.

In Section 2, we briefly review the exact estimation of the AR(1) model, which has been studied extensively. In that case, exact estimation may be done using procedures more elegant and less numerically intensive than ours, but those procedures are of course tailored to the AR(1) model. By showing how our procedure works in the simple AR(1) case, we provide motivation and intuitive feel for it, and we generalize it to much richer models in Section 3. In Sections 4 and 5, we use our procedure to obtain the exact maximum likelihood estimator for an ARCH model, and we compare its sampling properties to those of three common approximations. We conclude in Section 6.

2. Exact Maximum Likelihood Estimation of Autoregressions, Revisited

To understand the methods that we will propose for the exact maximum likelihood estimation of finite-ordered observation-driven models, it will prove useful to sketch the construction of the exact likelihood for a simple Gaussian AR(1) process.

The covariance stationary first-order Gaussian autoregressive process is

$$y_t = \rho y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

where $|\rho|<1$, t = 1, ..., T. The likelihood may be factored into the product of T-1 conditional likelihoods and an initial marginal likelihood. Specifically,

$$L(\theta) = l_T(y_T|\Omega_{T-1};\ \theta)\ l_{T-1}(y_{T-1}|\Omega_{T-2};\ \theta)\ \ldots\ l_2(y_2|\Omega_1;\ \theta)\ l_1(y_1;\ \theta),$$

where $\theta = (\rho,\ \sigma^2)'$ and $\Omega_t = \{y_t,\ \ldots,\ y_1\}$. The initial likelihood term $l_1(y_1;\ \theta)$ is known in closed form; it is

$$l_1(y_1;\ \theta) = (2\pi)^{-1/2} \sqrt{\frac{1-\rho^2}{\sigma^2}}\ \exp\left[-\frac{1-\rho^2}{2\sigma^2}\, y_1^2\right].$$

The remaining likelihood terms are

$$l_t(y_t|\Omega_{t-1};\ \theta) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(y_t - \rho y_{t-1})^2\right],$$

t = 2, ..., T.

Beach and MacKinnon (1978) show that small-sample bias reduction and efficiency gains are achieved by maximizing the exact likelihood, which includes the initial likelihood term, as opposed to the approximate likelihood, in which the initial likelihood term is either dropped or treated in an ad hoc manner. Moreover, they find that as ρ increases, the relative efficiency of exact maximum likelihood increases.

Now let us consider an alternative way of performing exact maximum likelihood. The key insight is that the initial likelihood term, for any given parameter configuration, is simply the unconditional density of the first observation, evaluated at y_1, which can be estimated to any desired degree of

accuracy using well-known techniques of simulation and consistent nonparametric density estimation.

We proceed as follows. At any numerical iteration (the j^{th}, say) en route to finding a maximum of the likelihood, a current "best guess" of the parameter vector exists; call it $\theta^{(j)}$. Therefore, we can simulate a very long realization of the process with parameter $\theta^{(j)}$ and estimate its unconditional density at y_1; denote it $\hat{l}_1(y_1;\ \theta^{(j)})$. The estimated density at y_1 is the first observation's contribution to the likelihood for the particular parameter configuration $\theta^{(j)}$. Then we construct the Gaussian likelihood

$$L(\theta^{(j)}) \approx \hat{l}_1(y_1;\ \theta^{(j)}) \prod_{t=2}^{T} \sigma^{-1} \exp\left[-\frac{1}{2\sigma^2}\ (y_t - \rho y_{t-1})^2\right],$$

and we maximize it with respect to θ using standard numerical techniques. The approximation error goes to zero -- that is, $\hat{l}_1(y_1;\ \theta^{(j)}) \rightarrow l_1(y_1;\ \theta^{(j)})$, so we obtain the exact likelihood function -- as the size of the simulated sample whose density we consistently estimate goes to infinity.

Obviously, it would be wasteful to adopt the simulation-based approach outlined here for exact estimation of the first-order autoregressive model, because the unconditional density of y_1 is known in closed form. In other important models, however, the unconditional density is *not* known in closed form, and in such cases our procedure provides a solution. Thus, we turn now to a general statement of our procedure, and then to a detailed illustration.

3. Observation-Driven Models with Arbitrary Conditional Density

The observation-driven form

$$y_t = f(y^{(t-1)}, \varepsilon_t)$$

usually makes it a simple matter to find the conditional density

$$y_t | y^{(t-1)} \sim D(y^{(t-1)}; \theta),$$

where the form of the conditional density D depends on f(.) and the density of ε_t. Many observation-driven models are in fact specified directly in terms of the conditional density D, which is typically assumed to be a member of a convenient parametric family. The likelihood is then just the product of the usual conditional densities and the initial joint marginal D^* (which is p-dimensional, say),

$$L(y_T,\ldots,y_1; \theta^{(j)}) = D^*(y_1,\ldots,y_p; \theta^{(j)}) \prod_{t=(p+1)}^{T} D(y^{(t-1)}; \theta^{(j)}).$$

The difficulty of constructing the exact likelihood function stems from the fact that the unconditional density D^* is typically not known in closed form, even when a large amount of structure (e.g., normality) is placed on the conditional density D. In a fashion that precisely parallels the above AR(1) discussion, however, we can consistently estimate D^* from a long simulation of the model, resulting in

$$L(y_T,\ldots,y_1; \theta^{(j)}) \approx \hat{D}^*(y_1,\ldots,y_p; \theta^{(j)}) \prod_{t=(p+1)}^{T} D(y^{(t-1)}; \theta^{(j)}).$$

As in the AR(1) case, the approximation error is under the control of the investigator, regardless of the sample size T, and it can be made arbitrarily small by simulating a long enough realization.

A partial list of observation-driven models for which exact maximum-likelihood estimation may be undertaken using the techniques proposed here includes Engle's (1982) ARCH model, models of higher-order conditional dynamics (e.g., time-varying conditional skewness or kurtosis), Poisson models with time-varying intensity, Hansen's (1994) autoregressive conditional density model, Cox's (1981) dynamic logit model, and Engle and Russell's (1995) conditional duration model. Moreover, the conditional density needn't be Gaussian, and the framework is not limited to pure time series models. It applies, for example, to regressions with disturbances that follow observation-driven processes.

4. Exact Maximum Likelihood Estimation of ARCH Models

Volatility clustering and leptokurtosis are routinely found in economic and financial time series, but they elude conventional time series modeling techniques. Engle's (1982) ARCH model and its generalizations are consistent with volatility clustering by construction and with unconditional leptokurtosis by implication; hence their popularity. ARCH models are now widely used in the analysis of economic time series and are implemented in popular computer packages like Eviews and PC-GIVE. Applications include modeling exchange

rate, interest rate and stock return volatility, modeling time-varying risk premia, asset pricing (including options), dynamic hedging, event studies, and many others.[2]

Engle's (1982) ARCH process is a classic and simple example of a model amenable to exact estimation with the techniques developed here. The known conditional probability structure of ARCH models facilitates approximate maximum likelihood estimation by prediction-error factorization of the likelihood. *Exact* maximum likelihood estimation has not been attempted, however, because the unconditional density l_p is not known in closed form. The prevailing view (namely, that exact maximum likelihood estimation is effectively impossible) is well summarized by Nelson and Cao (1992), who assert that[3]

> "...in practice (for example in estimation) it is necessary to compute [the conditional variance] recursively ... assuming arbitrary fixed values for $\{\varepsilon_0^2, \ ..., \ \varepsilon_{-p+1}^2\}$."
>
> (p. 232)

In short, the issue of exact maximum likelihood estimation is, without exception among the hundreds of published studies using ARCH techniques, skirted by conditioning upon ad hoc assumptions about l_p. Although the

---

[2] See Diebold and Lopez (1995).

[3] Their notation has been changed to match ours.

treatment of l_p is asymptotically inconsequential, it may be important in small samples, particularly when conditional variance persistence is high. With this in mind, we construct the exact likelihood function of an ARCH process using the procedure outlined earlier.

Consider the sample path $\{\varepsilon_t\}_{t=1}^{T}$ governed by the pth-order ARCH process,

$$\varepsilon_t = \sigma_t \eta_t$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_p \varepsilon_{t-p}^2$$

$$\eta_t \overset{iid}{\sim} N(0,1),$$

where $\sum_{i=1}^{p} \alpha_i < 1$, $\omega > 0$, $\alpha_i \geq 0$, $\forall\ i = 1, \ldots, p$.[4] Let $\theta = (\omega, \alpha_1, \ldots, \alpha_p)'$. The exact likelihood for a sample of size T is the product of the T-p conditional point likelihoods corresponding to observations (p+1) through T, and the unconditional joint likelihood for observations 1 through p. That is,

$$L(\varepsilon_T, \ldots, \varepsilon_1; \theta) = l_T(\varepsilon_T | \Omega_{T-1}; \theta)\ l_{T-1}(\varepsilon_{T-1} | \Omega_{T-2}; \theta) \ldots$$

$$\ldots\ l_{p+1}(\varepsilon_{p+1} | \Omega_p; \theta)\ l_p(\varepsilon_p, \ldots, \varepsilon_1; \theta)\ .$$

We simulate a very long realization of the process with parameter $\theta^{(j)}$ and consistently estimate the height of the unconditional density of the first p observations, evaluated at $\{\varepsilon_1, \ldots, \varepsilon_p\}$; denote it $\hat{l}_p(\varepsilon_1, \ldots, \varepsilon_p; \theta^{(j)})$. We substitute this estimated p-dimensional unconditional density into the likelihood

[4] We adopt the conditional normality assumption only because it is the most common. Alternative distributions, such as the Student's t advocated by Bollerslev (1987), could be used with no change in our procedure.

where the true unconditional density appears, yielding the full conditionally Gaussian likelihood,

$$L(\varepsilon_T,\ldots,\varepsilon_1;\ \theta^{(j)}) \ \approx\ \hat{l}_p(\varepsilon_1,\ldots,\varepsilon_p;\ \theta^{(j)}) \prod_{t=(p+1)}^{T} \left[\sigma_t^{-1}(\theta^{(j)}) \quad \exp\left(\frac{-1}{2\sigma_t^2(\theta^{(j)})}\ \varepsilon_t^2 \right) \right],$$

which we maximize using standard numerical techniques.

5. Comparative Finite-Sample Properties of Exact and Approximate Maximum Likelihood Estimators of ARCH Models

For purposes of illustration, we study a conditionally Gaussian ARCH(1) process with unit unconditional variance,

$$\varepsilon_t | \varepsilon_{t-1} \ \sim\ N(0,\ \sigma_t^2)$$
$$\sigma_t^2 \ =\ (1-\alpha) + \alpha\varepsilon_{t-1}^2 \ .$$

The stark simplicity of this data generating process is intentional. Although the model is restrictive, all the points that we want to make can be made within its simple context, and the simplicity of the model (in particular, the one-dimensional parameter space) renders it amenable to Monte Carlo analysis. Moreover, the ARCH(1) *is* sometimes used in practice; the popular PC-GIVE software, for example, permits only ARCH(1) estimation. It should be kept in mind that our procedure is readily applied in higher-dimensional situations, even though the associated increased computational burden makes Monte Carlo analysis infeasible.

The Monte Carlo experiments were done in vectorized FORTRAN 77 at the Cornell National Supercomputer Facility. We report the results of nine experiments, corresponding to $\alpha = .9, .95, .99$, and $T = 10, 25, 50$, each with 1000 Monte Carlo replications performed. The nonparametric estimation of the initial likelihood term is done by the kernel method, using a standard normal kernel, fit to a simulated series of length 1000. The bandwidth is set to $\hat{\gamma}(1000)^{-\frac{1}{5}}$, where $\hat{\gamma}(\alpha^{(j)}) = (\sum_{i=1}^{1000} x_i^2(\alpha^{(j)})/1000)^{1/2}$ and $x_i(\alpha^{(j)})$, i = 1, ..., 1000, is the simulated sample.[5] The same random numbers are used to construct the simulated sample at each evaluation of the likelihood and across Monte Carlo replications.

Because the effect of initial conditions is central to this small-sample exercise, we take care to let the process run for some time before sampling. Specifically, each Monte Carlo sample is taken as the last T elements of a vector of length 500+T, thus eliminating any effects that the starting value (0) might have.

The calculation of the likelihood for observations 2 through T is the same for the exact and approximate methods; the methods differ only in the calculation of the initial likelihood. Our exact method, specialized to the case at hand, yields

---

[5] Silverman (1986) advocates the use of such a bandwidth selection procedure, and it satisfies the conditions required for consistency of the density estimator. More sophisticated "optimal" bandwidth selection procedures may of course be employed if desired.

$$\hat{l}_1(\varepsilon_1;\ \alpha) \ = \ \frac{1}{1000\ \hat{\gamma}(\alpha^{(j)})\ (1000)^{-1/5}} \sum_{i=1}^{1000} K\left(\frac{\varepsilon_1 \ - \ x_i(\alpha^{(j)})}{\hat{\gamma}(\alpha^{(j)})\ (1000)^{-1/5}} \right),$$

where $K(\cdot)$ is the N(0,1) density function, and x_i, i = 1, ..., 1000 is a simulated ARCH(1) process with parameter α.

Three approximations to the initial likelihood are considered:

(A1) We simply set $l_1(\varepsilon_1;\ \alpha) \ = \ 1$. This is of course a perfectly well defined likelihood, but it does not make full use of all information contained in the sample.

(A2) The functional form of $l_1(\varepsilon_1;\ \alpha)$ is assumed (incorrectly) to be Gaussian, as with all of the conditional densities, and the unconditional variance (1) is substituted for the unavailable ε_0^2, which yields $l_1(\varepsilon_1;\ \alpha) \ = \ \exp(-\varepsilon_1^2/2)$.

(A3) The functional form of $l_1(\varepsilon_1;\ \alpha)$ is assumed (incorrectly) to be Gaussian, and the unconditional mean (0) is substituted for the unavailable ε_0, which yields

$$l_1(\varepsilon_1;\ \alpha) \ = \ (1-\alpha)^{-1/2}\ \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_1^2}{1-\alpha} \right) \right).$$

To be certain that global maxima are found, we maximize the exact and approximate likelihoods using a grid search over the relevant parameter space (in this case, the unit interval). The grid mesh is of width .01, and it is reduced to .002 when the distance from either boundary is less than or equal to .05, and when the distance from the true parameter value is less than or equal to .05.

The exact and approximate estimators' biases, variances and mean-squared errors are displayed in Table 1. Efficiency of all methods increases with T and α. The exact method, however, consistently outperforms all approximate methods, especially for small T and large α. The mean-squared error reductions afforded by the exact estimator typically come both from variance and bias reductions. In Figure 1, we graphically highlight the results for small samples (T = 5, 10, 15, 20, 25) with high persistence (α = .99); the efficiency gains from exact maximum likelihood are immediately visually apparent.

Our results are consistent with existing literature. Beach and MacKinnon (1978), in particular, report efficiency gains from exact maximum likelihood estimation in autoregressive processes. But ARCH processes *are* autoregressions in squares; that is, if ε_t is an ARCH(p) process,

$$\varepsilon_t \mid \varepsilon_{t-1}, \ ..., \ \varepsilon_{t-p} \ \sim \ N(0, \sigma_t^2)$$

$$\sigma_t^2 \ = \ \omega \ + \ \alpha(L)\varepsilon_t^2,$$

where $\alpha(L) \ = \ \sum_{i=1}^{p} \alpha_i L^i$, $\omega>0$, $\alpha_i \geq 0 \ \forall i$, and $\alpha(1)<1$, then ε_t^2 has the covariance-stationary autoregressive representation

$$\varepsilon_t^2 \ = \ \omega \ + \ \alpha(L)\varepsilon_t^2 \ + \ \nu_t,$$

where $\nu_t \ = \ \varepsilon_t^2 \ - \ \sigma_t^2$ is the difference between the squared innovation and the conditional variance at time t.[6]

---

[6] See Diebold and Lopez (1995) for additional discussion.

6. Summary and Directions for Future Research

We have proposed an exact estimator for finite-ordered observation-driven models. The exact estimator is more efficient than commonly-used approximate estimators. Our methods are computationally intense but nevertheless entirely feasible, even accounting for the "curse of dimensionality" associated with higher-dimensional situations, due to the fact that the simulation sample size may be made very large.

Our "exact" estimator, like its approximate competitors, is in fact an approximation, but with the crucial difference that the size of the approximation error is under the control of the investigator. In real applications, a very large simulation sample size can be used in order to guarantee that the approximation error is negligible. Similarly, more sophisticated methods of bandwidth selection and likelihood maximization may be used.

In closing, let us sketch a potentially fruitful direction for future research -- application of our likelihood evaluation technique to panel data, the time series dimension of which is often notoriously small. Consider, for example, a simple dynamic model for panel data:[7]

$$y_{it} = \rho y_{i,t-1} + x_{it}'\beta + \mu_i + \varepsilon_{it},$$

$i = 1, ..., N$ and $t = 1, ..., T$, where

[7] For background on such models, see Bhargava and Sargan (1983), Sevestre and Trognon (1992), and Nerlove (1996).

$$E(\varepsilon_{it}\varepsilon_{js}) = \sigma_\varepsilon^2,\ j = i, t = s$$
$$= 0 \ otherwise.$$

In a fixed effects model, μ_i is an individual-specific parameter, whereas in a random- effects model μ_i is a zero-mean random variable with

$$E(\mu_i\mu_j) = \sigma_\mu^2 \ for \ i = j$$
$$= 0 \ otherwise.$$

Assume that the

densities of ε_{it} and μ_i are Gaussian, and that the regressors x_{it} are iid over space and time.[8]

Because of the independence across i, the complete likelihood is simply the product of the likelihoods for the N individuals. In an obvious notation, the i^{th} likelihood for either the fixed or random coefficient model is

$$L(Y_i, X_i, \mu_i; \theta) = l_T(y_{iT}, x_{iT}, \mu_i | y_{i,T-1}; \theta)\ l_{T-1}(y_{i,T-1}, x_{i,T-1}, \mu_i | y_{i,T-2}; \theta) \ ...$$
$$... \ l_2(y_{i2}, x_{i2}, \mu_i | y_{i1}; \theta)\ l_1(y_{i1} x_{i1}, \mu_i; \theta).$$

This is our familiar likelihood factorization. As before, the only complication is evaluation of the unconditional likelihood of the initial observation, and as before we simply simulate a long realization of y_{it} from which we can estimate the unconditional density at y_{i1}. At iteration j the model is

[8] A challenging extension will be to allow for serial dependence and spatial heterogeneity in the regressors.

$$y_{it} = \rho^{(j)} y_{i,t-1} + x_{it}'\beta^{(j)} + \mu_i + \varepsilon_{it}.$$

If we estimate a fixed effects model, we also have $\mu_i^{(j)}$ at iteration j; under random effects we instead have $\sigma_\mu^{2\,(j)}$. Thus the only barrier to performing the required simulation is the set of regressors, x_{it}. But the regressors are uncorrelated over space and time; thus we can sample with replacement from the observed set of NT regressors. The likelihood evaluation algorithm is as follows. At iteration j:

(1) Initialize y_{i0}.

(2) Draw $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^{2\,(j)})$, t = 1, ..., R.

(3) Draw x_{it} by sampling with replacement from $X_{[NxT]}$, t = 1, ..., R.

(4) If random effects, draw $\mu_i \sim N(0, \sigma_\mu^{2\,(j)})$; else if fixed effects, let $\mu_i = \mu_i^{(j)}$.

(5) Generate $y_{it} = \rho^{(j)} y_{i,t-1} + x_{it}'\beta^{(j)} + \mu_i + \varepsilon_{it}$, $t = 1, ..., R$.

(6) Estimate the unconditional density (likelihood) of y_{it} at the initial observation y_{i1}.

(7) Form the complete likelihood for individual i.

(8) Repeat for i = 1,..., N and form the complete likelihood.

The separability of the likelihood across *i* makes for simple likelihood evaluation. In particular, the dimension of the required density estimation for each i is only as large as the order of serial dependence, just as in the univariate

time series case. Thus, for the prototype model at hand, evaluation of the likelihood requires only one-dimensional density estimates.

References

Beach, C.M. and MacKinnon, J.G. (1978), "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors," *Econometrica*, 46, 51-58.

Bhargava, A. and Sargan, J.D. (1983), Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods," Econometrica, 51, 1635-1659.

Bollerslev, T. (1987), "A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69, 542-547.

Cox, D.R. (1981), "Statistical Analysis of Time Series: Some Recent Developments," *Scandinavian Journal of Statistics*, 8, 93-115.

Diebold, F.X. and Lopez, J. (1995), "Modeling Volatility Dynamics," in Kevin Hoover (ed.), *Macroeconometrics: Developments, Tensions and Prospects*. Boston: Kluwer Academic Publishers, 427-472, 1995.

Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

Engle, R.F. and Russell, J. (1995), "Forecasting Transaction Rates: The Autoregressive Conditional Duration Model," Manuscript, Department of Economics, University of California, San Diego.

Gray, S.F. (1995), "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," Manuscript, Fuqua School of Business, Duke University.

Hamilton, James D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.

Hansen, B.E. (1994), "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705-730.

Nelson, D.B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347-370.

Nelson, D.B. and Cao, C.Q. (1992), "Inequality Constraints in the Univariate GARCH Model," *Journal of Business and Economic Statistics*, 10, 229-235.

Nerlove, M., (1996), *Econometrics for Applied Economists*. San Diego: Academic Press, forthcoming.

Sevestre, P. and Trognon, A. (1992), "Linear Dynamic Models," in L. Matyas and P. Sevestre (eds.), *Econometrics of Panel Data: Handbook of Theory and Applications*. Amsterdam: Kluwer.

Shephard, N. (1995), "Statistical Aspects of ARCH and Stochastic Volatility," Nuffield College Oxford, Economics Discussion Paper No. 94.

Silverman, B.W. (1986), *Density Estimation*. New York: Chapman and Hall.

Table 1
Exact and Approximate Maximum Likelihood Estimation

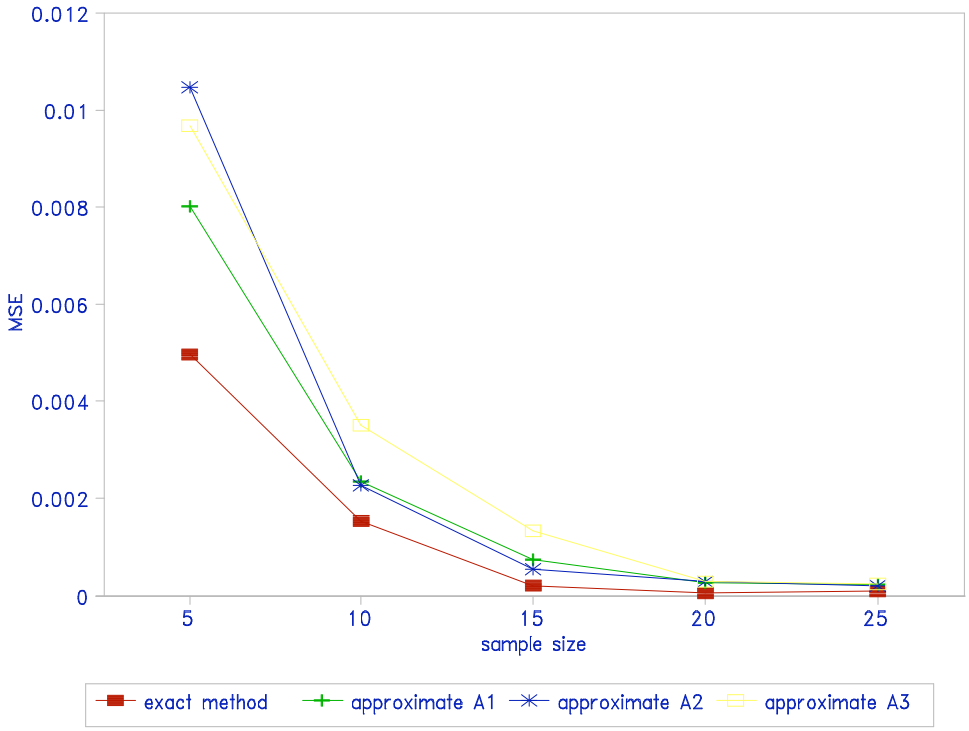
		Exact	A1	A2	A3
	α=.9				
	Bias	0.01822	0.04967	0.04628	0.05147
	Var	0.01640	0.02327	0.02218	0.02593
	MSE	0.01674	0.02573	0.02432	0.02858
	α=.95				
	Bias	0.00843	0.03028	0.02662	0.03241
T=10	Var	0.00570	0.01010	0.00807	0.01229
	MSE	0.00577	0.01102	0.00878	0.01334
	α=.99				
	Bias	0.00258	0.00923	0.00831	0.01119
	Var	0.00152	0.00226	0.00219	0.00339
	MSE	0.00153	0.00234	0.00225	0.00351

		Exact	A1	A2	A3
	α=.9				
	Bias	0.00699	0.02372	0.02228	0.01959
	Var	0.00398	0.00685	0.00685	0.00682
	MSE	0.00403	0.00741	0.00735	0.00720
	α=.95				
	Bias	0.00339	0.01425	0.01328	0.01199
T=25	Var	0.00121	0.00268	0.00273	0.00294
	MSE	0.00122	0.00288	0.00290	0.00307
	α=.99				
	Bias	0.0019	0.00432	0.00308	0.00293
	Var	9.72E-5	2.15E-4	1.88E-4	2.49E-4
	MSE	9.72E-5	2.26E-4	1.98E-4	2.57E-4

		Exact	A1	A2	A3
	α=.9				
	Bias	0.00122	0.00940	0.00801	0.00914
	Var	0.00129	0.00226	0.00165	0.00252
	MSE	0.00129	0.00235	0.00171	0.00261
	α=.95				
	Bias	0.00024	0.00532	0.00426	0.00493
T=50	Var	0.00032	0.00074	0.00046	0.00074
	MSE	0.00032	0.00077	0.00047	0.00076
	α=.99				
	Bias	0.00003	0.00122	0.00101	0.00114
	Var	1.23E-5	3.52E-5	1.96E-5	3.38E-5
	MSE	1.23E-5	3.67E-5	2.06E-5	3.51E-5

Notes to Table 1: The data are generated as an ARCH(1) process; α is the ARCH parameter and T is the sample size. Three estimators are compared: exact maximum likelihood ("Exact"), and three approximations ("A1," "A2," and "A3"). We report the bias, variance and mean-squared error for each estimator ("Bias," "Var," and "MSE"). See the text for details.

Figure 1
Mean-Squared Error Comparison, $\alpha = .99$



Notes to Figure 1: The data are generated as an ARCH(1) process; α is the ARCH parameter and T is the sample size. We show the mean-squared error (MSE) of three estimators of α as a function of T: exact maximum likelihood ("Exact method"), and three approximations ("A1," "A2," and "A3"). See the text for details.