INTERNATIONAL ECONOMIC REVIEW
Vol. 25, No. 3, October, 1984

EFFICIENT ESTIMATION OF THE GEOMETRIC DISTRIBUTED LAG MODEL: SOME MONTE CARLO RESULTS ON SMALL SAMPLE PROPERTIES*

BY F. C. PALM, E. VOGELVANG AND D. A. KODDE**

1. INTRODUCTION

In recent years, several asymptotically efficient two-step and iterative estimators for dynamic models with autocorrelated errors have been presented and results on their finite sample properties have been obtained. Among closely related Monte Carlo studies on small sample properties, we specifically note the comparison of the finite sample properties of several estimators for the regression model with autoregressive errors by Rao and Griliches [1969] and by Park and Mitchell [1980], and for the Koyck [1954] distributed lag model by Morrison [1970] and Dhrymes [1971]. Hatanaka [1974] presents an efficient two-step estimator for a single equation dynamic adjustment model with first order autoregressive errors and reports results of a simulation experiment. More detailed results on the small sample performance of several estimators for the dynamic adjustment model are reported in Maeshiro [1980] for a trended explanatory variable and in Fomby and Guilkey [1983] for a stationary exogenous variable. Hendry and Sbra [1977] investigated the small sample properties of instrumental variables estimators in a simultaneous equation framework with autoregressive errors. Harvey and McAvinchey [1981] compared the efficiency in small samples of various two-step and iterative estimation procedures for regression models with moving average errors.

In this paper, we report Monte Carlo results on instrumental variables, efficient two-step and iterative Gauss-Newton estimators of a Koyck distributed lag model with uncorrelated errors (model 1) and with first order autoregressive errors (model 2). The distributed lag model with a Koyck scheme, perhaps the most widely used distributed lag model, is simple in the sense that it involves a small number of parameters. The parameter of the lag distribution can often be interpreted in terms of economic behavior such as adaptive expectation formation or partial adjustment. Still, the problems generally inherent in the estimation of distributed lag models are also present here, so that Koyck's model is a natural candidate for a simulation study. We only consider models for stationary data. As recent developments in time series modeling show, stationary processes are

* Manuscript received November, 1980; revised October, 1983.
** The authors wish to thank H. J. Blommestein and J. T. C. Kool for their help in carrying out the computational work. They are indebted to D. F. Hendry and A. C. Harvey and three unknown referees for their helpful comments. An earlier version of this paper was presented at the Econometric Society European Meeting in Amsterdam, August 31-September 4, 1981.

very useful for the analysis of economic time series. We note that the Koyck distributed lag model is a simple example of the rational transfer function model frequently used in time series analysis.

In the last decade, dynamic specification analysis has received much attention in the econometric literature. As econometric specification analysis usually requires estimates of several alternative dynamic model specifications possibly arranged as a uniquely ordered sequence of restricted models, the demand for computationally convenient estimation methods with desirable small and large sample statistical properties has arisen. Usually one has to choose between consistent but inefficient or consistent and asymptotically efficient estimators, which may or may not be iterative. The choice is usually based on criteria such as computational costs involved, small sample properties and asymptotic efficiency.

To offer some guidance for empirical work, we focus on the small sample properties of one estimator in each of the three classes of estimators, i.e. Liviatan's consistent instrumental variables estimator, an efficient two-step and an iterative Gauss-Newton estimator. The latter has been called minimum chi-square estimator by Dhrymes [1971] (see also Dhrymes [1974]) who shows that it becomes indistinguishable from the exact ML estimator in larger samples. The estimates of the distributed lag parameter are restricted to be positive and smaller than one. Similarly, the estimates of the autoregressive coefficient in the disturbance process are restricted to lie inside the unit circle. As these restrictions are plausible in many applications, we investigate the relevance of large sample theory for the small sample properties of restricted estimators.

Two main questions have been considered:

(1) are asymptotically efficient estimators to be preferred to a consistent but an inefficient instrumental variables estimator?

(2) does it pay to iterate asymptotically efficient estimators until convergence is achieved?

In section 2, we present the models and the estimation procedures. A more detailed presentation of the estimation methods and their large sample properties can be found in Dhrymes, Klein and Steiglitz [1970], Dhrymes [1971] or Harvey [1981]. In section 3, we describe the experiments. Section 4 contains the results of the simulations which are summarized using response functions. Instead of generating a large number of runs for each experiment, we use the technique of control variates to increase the precision of the outcome of the simulations. In the last section, we present conclusions.

2. THE MODELS AND THE ESTIMATION PROCEDURES

We analyze the geometric distributed lag model

$$y_t = \alpha_0 + \alpha_1 \sum_{i=0}^{\infty} \lambda^i x_{t-i} + u_t,\ t = 1,\ldots, T, \tag{2.1}$$

where y_t is the endogenous variable, $0<\lambda<1$, the explanatory variable x_{t-i} is independent of the error term u_t for all t and i, and T is the sample size. We consider the cases where $u_t=\varepsilon_t$ (model 1) with ε_t being a normally distributed white noise with variance σ^2, and where u_t is generated by a first-order autoregressive process (model 2)

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{2.2}$$

with $|\rho|<1$, $\rho\neq\lambda$. First, we present the estimation methods for the more general model. Then, we briefly indicate how these methods specialize for model 1.

When u_t is generated by the process (2.2), the likelihood function, conditional on initial values, is given by

$$L(y, x, \alpha_0, \alpha_1, \lambda, \rho, \sigma^2) = (\sqrt{2\pi}\sigma)^{-T} \exp \frac{-1}{2\sigma^2} \sum_{t=1}^{T} \varepsilon_t^2. \tag{2.3}$$

The model can be written as

$$y_t - \rho y_{t-1} = \alpha_0(1-\rho) + \alpha_1(x_t^* - \rho x_{t-1}^*) + \varepsilon_t, \tag{2.4}$$

where the variable x_t^* is defined as

$$x_t^* = \sum_{i=0}^{\infty} \lambda^i x_{t-i} = \frac{1}{1-\lambda L} x_t \tag{2.5}$$

for a sequence of variables x_t with L being the lag-operator.

The first-order conditions for a maximum of the log-likelihood function with respect to $\theta=(\alpha_0, \alpha_1, \lambda, \rho)'$ are given by

$$\frac{\partial \ln L}{\partial \theta} = -\sigma^{-2} \frac{\partial \varepsilon}{\partial \theta} \varepsilon = 0 \tag{2.6}$$

where $\frac{\partial \varepsilon}{\partial \theta}$ is the matrix of partial derivatives of the disturbance ε_t with respect to the elements in θ;

$$\frac{\partial \varepsilon}{\partial \theta} = - \begin{pmatrix} 1-\rho & \cdots & 1-\rho \\ x_1^* - \rho x_0^* & \cdots & x_T^* - \rho x_{T-1}^* \\ \alpha_1(x_0^{**} - \rho x_{-1}^{**}) & \cdots & \alpha_1(x_{T-1}^{**} - \rho x_{T-2}^{**}) \\ u_0 & \cdots & u_{T-1} \end{pmatrix} \tag{2.7}$$

with $x_t^{**} = \frac{1}{(1-\lambda L)^2} x_t$ and $\varepsilon=(\varepsilon_1, \varepsilon_2,\ldots, \varepsilon_T)'$ is the vector of disturbances.

The first order conditions in (2.6) are nonlinear in the parameter vector θ. We can solve them iteratively to obtain the conditional maximum likelihood (ML) estimator. However, Fisher [1925] (see also Dhrymes and Taylor [1976]) showed that under certain conditions the following two-step procedure

$$\hat{\hat{\theta}} = \hat{\theta} - \Gamma^{-1}(\hat{\theta}) \left.\frac{\partial \ln L}{\partial \theta}\right|_{\theta=\hat{\theta}} \tag{2.8}$$

has the same asymptotic properties as the ML estimator of θ. In particular, the result holds provided that

1) $\hat{\theta}$ is a consistent estimator of θ such that $\sqrt{T}(\hat{\theta}-\theta_0)$, with θ_0 being the true value of θ, has some limiting distribution, and

2) $\Gamma(\hat{\theta})$ is a non-singular matrix such that $T^{-1}\Gamma(\hat{\theta})$ converges in probability to the Hessian matrix at θ_0 of the log-likelihood function divided by T. We denote by "$\hat{\theta}$" and "$\hat{\hat{\theta}}$" the first and second step parameter estimates respectively.

One way to obtain an asymptotically efficient estimate by implementing (2.8) is to compute one step of the Gauss-Newton algorithm starting with a consistent estimate of θ_0. The formula for the Gauss-Newton algorithm is given by

$$\hat{\hat{\theta}} = \hat{\theta} - \left[\frac{\partial \varepsilon}{\partial \theta}\frac{\partial \varepsilon'}{\partial \theta}\right]^{-1} \frac{\partial \varepsilon}{\partial \theta}\varepsilon\Big|_{\theta=\hat{\theta}}. \tag{2.9}$$

The right side of (2.9) is evaluated at a consistent estimate $\hat{\theta}$. If we iterate the estimator (2.9) until convergence, we get the nonlinear least squares estimator or the conditional ML estimator. Whether the nonlinear least squares estimator is identical to the exact ML estimator depends on the treatment of the initial values for x_t and u_t. Notice also that the difference between the two-step and the initial consistent estimator, $\hat{\hat{\theta}}-\hat{\theta}$, in (2.9) can be computed by an ordinary least squares regression of the residuals $\hat{\varepsilon}$ on their partial derivatives with respect to θ, both evaluated at $\hat{\theta}$. These derivatives can be computed analytically as in (2.7) or numerically (for numerically computed derivatives, see e.g. Harvey and McAvinchey [1981]).

We now use the analytical formula for the derivatives and compute the two-step estimator in (2.9) as follows:

1. Consistent parameter estimates are obtained by Liviatan's instrumental variables method applied to the transformed model

$$y_t = \alpha_0(1-\lambda) + \lambda y_{t-1} + \alpha_1 x_t + v_t, \quad t = 2,\ldots, T \tag{2.10}$$

with $v_t = u_t - \lambda u_{t-1}$, using x_{t-1} as an instrument for y_{t-1}. The restriction $0<\lambda<1$ is imposed on the estimate $\hat{\lambda}$. If $\hat{\lambda}$ lies outside the interval [.05, .95], it is fixed at the corresponding boundary value and the parameters α_0 and α_1 are estimated in a regression of $y_t - \hat{\lambda} y_{t-1}$ on x_t. The boundary values for $\hat{\lambda}$ were chosen after some experimentation with the model when $\lambda = .9$. For a boundary value very close to one and $\lambda = .9$, the iterative estimator of λ often has a cyclical behavior. Next, we compute $\hat{u}_t = \hat{v}_t + \hat{\lambda}\hat{u}_{t-1}$, $t=3,\ldots, T$, $\hat{u}_2 = \hat{v}_2$, where $\hat{v}_t$ is an instrumental variables residual,

$$\hat{\rho} = \sum_{t=3}^{T} \hat{u}_t \hat{u}_{t-1} / \sum_{t=2}^{T} \hat{u}_t^2 \quad \text{and}$$

$$\hat{\sigma}^2 = \frac{1}{T-4} \sum_{t=3}^{T} \hat{\varepsilon}_t^2,$$

where

$$\hat{\varepsilon}_t = \hat{u}_t - \hat{\rho}\hat{u}_{t-1}.$$

2. The two-step estimator in (2.9) can be computed by OLS applied to the following equation (which is obtained by adding the same terms to both sides of equation (2.4))

$$\begin{aligned}(2.11)\qquad & [y_t - \rho y_{t-1} + \lambda\alpha_1(x_{t-1}^{**} - \rho x_{t-2}^{**}) + \rho u_{t-1}] \\ & = \alpha_0[1-\rho] + \alpha_1[x_t^* - \rho x_{t-1}^*] + \lambda[\alpha_1(x_{t-1}^{**} - \rho x_{t-2}^{**})] + \\ & \quad + \rho[u_{t-1}] + \varepsilon_t, \quad t = 3,\dots, T,\end{aligned}$$

after evaluating the regressand and the regressors between brackets at consistent first-step estimates. The variables involved in the regressand and in the regressors of (2.11) are computed as

$$\hat{x}_t^* = x_t + \hat{\lambda}\hat{x}_{t-1}^* \quad \text{and} \quad \hat{x}_t^{**} = \hat{x}_t^* + \hat{\lambda}\hat{x}_{t-1}^{**}$$

with $\hat{x}_0^*$ and $\hat{x}_0^{**}$ being set equal to the sample mean of x_t and x_t^* respectively, divided by $1-\hat{\lambda}$ (the process x_t is stationary).

If the estimate $\hat{\hat{\lambda}}$ does not lie inside the interval [.05, .95], it is fixed at the corresponding boundary value and α_0, α_1 and ρ are estimated using (2.11). The runs, for which the restriction $|\hat{\hat{\rho}}|<1$ is not satisfied, are disregarded. The latter restriction has been satisfied in most cases, although we do not impose the asymptotic independence between the estimates of ρ and those of the remaining parameters by using a block-diagonal matrix Γ in the two-step and iterative estimation procedure. The iterative Gauss-Newton algorithm stops when the change in the estimates of α_1, λ and ρ is smaller than .001 or when the number of iterations is equal to 100. It also stops when the restriction on λ is violated for the second time or when the Hessian matrix becomes singular. The variance of ε_t is estimated as in step 1 using the residuals of step 2 or those of the iterative estimator respectively. For model 1 ($\rho=0$), the appropriate estimates are obtained by deleting the last row of (2.7) and setting $\rho=0$. Observations for $t=2, \dots, T$, are used to compute the twostep and iterative estimates for α_0, α_1, λ and σ^2. For both models we ignore the first observations.

We have not investigated whether this affects the conclusions about the finite sample properties, as has been found by e.g. Beach and MacKinnon [1978] and by Park and Mitchell [1980] for a linear regression model with autoregressive errors and by Maeshiro [1980] and Fomby and Guilkey [1983] for the dynamic adjustment model with autoregressive disturbances. With the exception of Fomby and Guilkey [1983], these authors consider models with trended explanatory variables and find, for $\rho>0$, that the autoregressive transformation of x_t (e.g. $x_t - \rho x_{t-1}$) reduces the variability of the independent variable and thereby induces a loss of efficiency. Notice, however, that for the Koyck distributed lag model, the treatment of the initial conditions for x_t can be expected to have a similar favorable effect on the efficiency of parameter estimates as the inclusion of the first observation in GLS- or ML-estimation.

Finally, it should be noted that there are many other ways to generate two-step estimators with the same asymptotic distribution as the ML estimator. Any matrix Γ satisfying the requirement for (2.8) characterizes a two-step estimator which is asymptotically equivalent to the ML estimator. For example the estimators proposed by Hannan [1965] and by Steiglitz and McBride [1965] have this property. The small sample properties of these estimators and of Liviatan's instrumental variables estimator for model 1 have been investigated by Morrison [1970].

3. THE DESIGN OF THE EXPERIMENTS

The complete model used to generate the data is defined by the following

$$y_t = \alpha_0 + \alpha_1 \sum_{i=0}^{\infty} \lambda^i x_{t-i} + u_t, \quad 0 < \lambda < 1 \tag{3.1a}$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad \rho \neq \lambda, \quad |\rho| < 1 \tag{3.1b}$$

$$\varepsilon_t \sim IN(0, \sigma^2)\forall t, \tag{3.1c}$$

$$x_t = \gamma x_{t-1} + \eta_t, \quad 0 < \gamma < 1, \quad \text{and} \tag{3.1d}$$

$$\eta_t \sim IN(0, 10)\forall t. \tag{3.1e}$$

ε_t and $\eta_{t'}$ are independent for all t and t'.

The following parameter values are considered

$$\alpha_0 = 50, \quad \alpha_1 = .9$$
$$\lambda \in \{.3, .6, .9\}$$
$$\rho \in \{-.85, -.5, 0, .5, .85\}$$
$$\gamma \in \{0, .7, .95\}$$
$$\sigma^2 \in \{5, 10\}.$$

These values cover the range of plausible values for the parameters λ and ρ and for the theoretical R^2. The sample size T is equal to 30 and 60. The process for x_t is stationary. For $\gamma = .95$, the spectrum for x_t has approximately the "typical shape of the spectrum of an economic variable". The random numbers are generated from a uniform distribution using the FORTRAN subroutine RANF available on the CDC CYBER 170–750 computer. They are transformed into ε_t and η_t according to (3.1c) and (3.1e) using the probability integral theorem. The random variables u_t and x_t are generated according to (3.1b) and (3.1d) respectively, with $u_1 = \varepsilon_1 \sqrt{\frac{1}{1-\rho^2}}$ and $x_1 = \eta_1 \sqrt{\frac{1}{1-\gamma^2}}$.

Then, for a given set of parameters α_0, α_1 and λ, 60 independent samples of size $40+T$ are generated for the variable y_t using the model (3.1a), with $x_t=0$ for $t\leq 0$. To guarantee the independence of y_t from the initial values of x_t, only the last T(i.e., 30 and 60 respectively) observations are used in the simulation study. As an alternative, we could have generated y_0 using its marginal density function implied by model (3.1) and the y_t's $t=1,\ldots, T$ using equation (2.4).

4. THE RESULTS OF THE SIMULATIONS

For each of the 60 independent runs of an experiment, we estimate the parameters using Liviatan's instrumental variables (IV) method, the two-step (2S) and the iterative Gauss-Newton (IGN) procedure as described in section 2. We compute and analyze the simulation mean and standard errors (SE) for these estimators. We investigate the relationships between simulation mean and SE's and the characteristics of the experiments. We model these relationships as response function equations and estimate them by OLS. Our analysis focuses on the appropriateness of large sample theory for finite sample situations. The existence of finite sample moments of the estimators has not been investigated. Possibly, the use of restricted estimators guarantees the existence of their finite sample moments.

4.1 *Control variates.* To reduce the variance of the simulation results, we use the technique of control variates (CV) (see Mikhail [1972, 1975]). (For a more detailed description of this variance reduction technique, see Hendry and Srba [1977] and the references therein). Suppose that we want to simulate the finite sample mean (assumed to exist) of an estimator $\hat{\theta}$ of the parameter θ. We can compute the sample mean of the outcome $\hat{\theta}_j$ of m independent runs

$$\bar{\hat{\theta}} = \frac{1}{m}\sum_{j=1}^{m}\hat{\theta}_j. \tag{4.1}$$

Consider now an alternative estimator $\bar{\theta}^\circ$ with known mean $E(\bar{\theta}^\circ)$. Then, the quantity $\tilde{\theta}=\bar{\hat{\theta}}-\bar{\theta}^\circ+E(\bar{\theta}^\circ)$ will have the same expectation as $\bar{\hat{\theta}}$. Its variance

$$\operatorname{var}(\tilde{\theta}) = \operatorname{var}(\bar{\hat{\theta}}) + \operatorname{var}(\bar{\theta}^\circ) - 2\operatorname{cov}(\bar{\hat{\theta}}, \bar{\theta}^\circ) \tag{4.2}$$

will be smaller than the variance of $\bar{\hat{\theta}}$, provided

$$2\operatorname{cov}(\bar{\hat{\theta}}, \bar{\theta}^\circ) > \operatorname{var}(\bar{\theta}^\circ). \tag{4.3}$$

The technique of CV's consists in choosing an estimator $\bar{\theta}^\circ$ (called CV) with known mean, satisfying (4.3) and using $\tilde{\theta}$ instead of $\bar{\hat{\theta}}$ to estimate the (unknown) finite sample expectation of $\hat{\theta}$. From (4.2), it is obvious that the precision of $\tilde{\theta}$ is very high when the CV $\bar{\theta}^\circ$ has approximately the same variance as $\bar{\hat{\theta}}$ and is almost perfectly correlated (positively) with $\bar{\hat{\theta}}$. Therefore, we derive the CV $\bar{\theta}^\circ$ from the asymptotic distribution of $\hat{\theta}$. We choose $\bar{\theta}^\circ$ such that its finite sample first and second moments are equal to the corresponding large sample moments of $\bar{\theta}$. As these moments only depend on the parameters of the model, which are known in a simulation study, they can be computed and used to obtain $\tilde{\theta}$.

The IV estimator of $\theta=(\beta', \rho)'$, where $\beta'=(\alpha_0, \alpha_1, \lambda)$, is

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'y, \ \hat{\rho}_{IV} = (\hat{u}'_{-1}\hat{u}_{-1})^{-1}\hat{u}'_{-1}\hat{u}, \tag{4.4}$$

with Z, X and $\hat{u}_{-1}=(\hat{u}_1, \hat{u}_2,\ldots, \hat{u}_{T-1})'$ being the matrices of instruments, regressors, and the vector of lagged residuals, respectively, of the first step estimation of (2.10). The CV's are given by

$$\beta^{\circ}_{IV} = E^{-1}(Z'X)Z'v + \beta, \tag{4.5a}$$

where $v=(v_2, v_3,\ldots, v_T)'$, and

$$\rho^{\circ}_{IV} = E^{-1}(u'_{-1}u_{-1})\, u'_{-1}\varepsilon + \rho = \frac{1-\rho^2}{(T-1)\sigma^2}(u'_{-1}\varepsilon)+\rho. \tag{4.5b}$$

The control variate β°_{IV} has as expectation β. The variance of $\sqrt{T}\beta^{\circ}_{IV}$ is equal to the asymptotic variance of $\sqrt{T}\hat{\beta}_{IV}$,

$$\Omega_{IV} = TE^{-1}(Z'X)E(Z'VZ)E^{-1}(X'Z), \tag{4.6}$$

where V is the covariance matrix of v. The variable v_t is generated by an ARMA (1, 1)-model $v_t=\frac{1-\lambda L}{1-\rho L}\varepsilon_t$, so that it is straightforward to express the elements of V as functions of λ, ρ and σ^2 (see Box and Jenkins [1970, p. 76]).

The control variate ρ°_{IV} is centered at ρ. The variance of $\sqrt{T-1}\rho^{\circ}_{IV}$ equals the asymptotic variance of $\sqrt{T}\hat{\rho}_{IV}$,

$$\operatorname{Var}(\sqrt{T}\hat{\rho}_{IV}) = 1 - \rho^2. \tag{4.7}$$

Notice that $\hat{\rho}_{IV}$ and $\hat{\beta}_{IV}$ are independent in large samples. The CV's given in (4.5) are expected to be almost perfectly correlated with the IV estimates in large samples. As the two-step and the iterative estimators have the same asymptotic distribution, we use the same CV's

$$\theta^{\circ}_{2S} = \theta^{\circ}_{IGN} = E^{-1}(P'P)P'\varepsilon + \theta, \tag{4.8}$$

where $P'=-\frac{\partial\varepsilon}{\partial\theta}$ defined in (2.7) but for $t=3,\ldots T$.

The mean of the control variates, $E(\theta^{\circ}_{2S})$ is equal to the true parameter values. The finite sample variance of $\sqrt{T}\theta^{\circ}_{2S}$ is the same as the large sample variance of the $2S$-estimator,

$$\operatorname{Var}(\sqrt{T}\hat{\theta}_{2S}) = \sigma^2_{\varepsilon}TE^{-1}(P'P). \tag{4.9}$$

The matrix $E(P'P)$ will be given in the appendix. The CV's for model 1 are easily obtained from (4.5) and (4.8) by setting $\rho=0$ and deleting the last column of P.

4.2. *Results for Some Selected Experiments.* In table 1, we report detailed results for 12 experiments. The values of the parameters and the sample size in these experiments are close to those often encountered in empirical econometric work. The results of table 1 have not been used to estimate the response functions in section 4.3. They have been kept back to investigate the predictive performance of the response functions in section 4.4.

In columns 3, 8 and 14 of table 1, the simulation mean (M) for the IV, 2S and IGN estimators respectively of a parameter θ_i is given

$$\bar{\theta}_i = \frac{1}{m} \sum_{j=1}^{m} \hat{\theta}_{ij} \tag{4.10}$$

where $m=60$ minus the number of times convergence as described in section 2 is not achieved. In columns 4, 9 and 15, the simulation standard errors (SSE) for the estimators are computed as

$$SSE_i = \left[\frac{1}{m-1} \sum_{j=1}^{m} (\hat{\theta}_{ij} - \bar{\theta}_i)^2\right]^{\frac{1}{2}}. \tag{4.11}$$

In columns 5 and 10, the mean of the control variates for the IV and 2S-estimator respectively (MCV) is given by

$$\bar{\theta}_i^\circ = \frac{1}{m} \sum_{j=1}^{m} \theta_{ij}^\circ. \tag{4.12}$$

In columns 6 and 11, the standard deviation of the control variates (SDCV) is computed as

$$SDCV_i = \left[\frac{1}{m-1} \sum_{j=1}^{m} (\theta_{ij}^\circ - \bar{\theta}_i^\circ)^2\right]^{\frac{1}{2}}. \tag{4.13}$$

In columns 12 and 16, the square root of the mean of the variances of the estimators obtained from the conventional formula for the estimated standard errors (ESE) is computed as

$$ESE_i = \left[\frac{1}{m-1} \sum_{j=1}^{m} DE_{ij}\right]^{\frac{1}{2}}, \tag{4.14}$$

where DE_{ij} is the i-th diagonal element of $\hat{\sigma}_j^2[\hat{P}_j'\hat{P}_j]^{-1}$ for run j, with $\hat{\sigma}_j^2$ and $\hat{P}_j$ being evaluated at the 2S and iterated estimates, respectively. For the IV estimator, the appropriate formula for the estimated variance of $\hat{\beta}_{IV}$ is given in (4.6), with the moments replaced by their sample equivalents. As the formula is almost never used in empirical work, we have not computed ESE's for the IV estimator.

In columns 7 and 13, the asymptotic standard errors (ASE) are equal to the square root of the i-th diagonal element of the covariance matrices in (4.6) and (4.9) divided by T. A CV estimate of the finite sample bias of the IV estimator [2S, IGN] can be obtained by subtracting column 5 [10, 10] from column 3 [8, 14]. Similarly, a CV estimate of the variance of the IV estimator [2S, IGN] can be obtained by subtracting the square of an element in column 6 [11, 11] from that of the corresponding element in column 4 [9, 15] and adding that of the

TABLE 1 SIMULATION RESULTS
(α_0=50, α_1=.9,

		Instrumental Variables Estimator					
		M	SSE	MCV	SDCV	ASE	M
$T = 40$	α_0	50.005	2.301	49.193	4.166	5.875	49.836
$\lambda = .7$	α_1	.8677	.1745	.8644	.1802	.2386	.9125
$\rho = 0$	λ	.71539	.10649	.71616	.08378	.10736	.70710
$m = 55$	σ^2	15.093	8.407				12.499
$T = 40$	α_0	49.501	9.428	50.258	2.272	4.766	47.851
$\lambda = .9$	α_1	.9341	.1473	.9239	.1760	.2898	1.0038
$\rho = 0$	λ	.88990	.05165	.89488	.04545	.07568	.89544
$m = 50$	σ^2	23.663	14.210				42.285
$T = 60$	α_0	50.791	5.698	51.060	3.786	4.776	48.463
$\lambda = .7$	α_1	.9478	.1712	.9459	.1693	.1940	.9225
$\rho = 0$	λ	.67728	.09156	.67863	.07529	.08729	.69806
$m = 58$	σ^2	12.533	3.493				11.871
$T = 60$	α_0	48.685	8.924	50.641	2.530	3.875	50.094
$\lambda = .9$	α_1	.9625	.1587	.9524	.1978	.2356	.9571
$\rho = 0$	λ	.87526	.06006	.88706	.05089	.06153	.90785
$m = 58$	σ^2	38.678	29.175				49.140
$T = 40$	α_0	50.009	1.941	49.497	3.493	4.257	49.688
$\lambda = .7$	α_1	.8740	.1524	.8738	.1586	.1773	.9288
$\rho = .4$	λ	.70909	.08603	.70968	.07038	.07978	.68998
$m = 55$	ρ	.39459	.17761	.37479	.14873	.14491	.30222
	σ^2	10.235	2.747				10.227
$T = 40$	α_0	48.753	8.556	50.095	1.915	3.172	51.975
$\lambda = .9$	α_1	.9236	.1199	.9121	.1510	.2078	.9730
$\rho = .4$	λ	.89272	.04195	.89818	.03789	.05425	.89375
$m = 42$	ρ	.53732	.16842	.37954	.14569	.14491	.33946
	σ^2	11.479	2.351				13.122
$T = 60$	α_0	50.640	4.651	50.783	3.281	3.460	45.122
$\lambda = .7$	α_1	.9309	.1575	.9339	.1476	.1441	.9285
$\rho = .4$	λ	.68304	.07949	.68417	.06504	.06486	.68677
$m = 59$	ρ	.38905	.12211	.38143	.11490	.11832	.32857
	σ^2	10.542	2.199				10.435
$T = 60$	α_0	49.865	7.059	50.135	1.719	2.575	52.183
$\lambda = .9$	α_1	.9343	.1481	.9137	.1357	.1689	.9554
$\rho = .4$	λ	.89314	.04117	.89737	.03471	.04411	.89228
$m = 50$	ρ	.58553	.20494	.35321	.12091	.11832	.40398
	σ^2	12.356	3.862				13.272
$T = 40$	α_0	50.878	11.332	49.458	3.319	4.321	46.887
$\lambda = .7$	α_1	.8967	.1376	.8957	.1212	.1646	.9200
$\rho = .8$	λ	.70761	.08587	.70933	.06326	.07406	.69019
$m = 60$	ρ	.66911	.11181	.80530	.08284	.09487	.64577
	σ^2	9.913	2.339				10.111
$T = 40$	α_0	48.846	13.879	49.640	1.781	2.428	59.561
$\lambda = .9$	α_1	.9008	.1112	.8801	.1182	.1646	.9429
$\rho = .8$	λ	.90206	.05130	.90772	.03525	.04297	.88346
$m = 45$	ρ	.75950	.12804	.80156	.08662	.09487	.68367
	σ^2	10.278	2.526				10.302
$T = 60$	α_0	50.090	2.534	50.116	3.144	3.523	49.972
$\lambda = .7$	α_1	.9082	.1642	.9024	.1550	.1338	.8966
$\rho = .8$	λ	.70087	.03017	.69706	.06436	.06021	.70321
$m = 59$	ρ	.70770	.09420	.79902	.08609	.07746	.70246
	σ^2	9.535	1.596				9.661
$T = 60$	α_0	50.489	9.008	50.181	1.565	1.962	52.907
$\lambda = .9$	α_1	.9223	.1171	.9223	.1216	.1338	.9404
$\rho = .8$	λ	.89425	.03770	.89645	.03009	.03494	.89786
$m = 54$	ρ	.75118	.12353	.80652	.08705	.07746	.69698
	σ^2	10.115	1.963				10.894

As the ASE's for the IV and the 2S-estimators are divided by the number of observations than those for the IV estimator.

FOR SOME SELECTED EXPERIMENTS
$\gamma = .85$, $\sigma^2 = 10$).

	Two-Step Estimator				Iterative Estimator		
SSE	MCV	SDCV	ESE	ASE	M	SSE	ESE
1.667	49.988	.466	1.178	.506	49.768	1.330	1.030
.1127	.8954	.0642	.0818	.0638	.9591	.0975	.0846
.03887	.70284	.02107	.03847	.02320	.69180	.03171	.04219
3.532					12.224	3.423	
15.410	50.046	.536	5.390	.506	48.630	11.760	4.046
.2499	.9004	.0150	.0776	.0231	1.0461	.2531	.0794
.03488	.90031	.00228	.01866	.00308	.89112	.03213	.02040
37.704					38.462	35.239	
12.166	49.974	.401	3.836	.412	50.067	.767	.658
.0839	.9017	.0569	.0696	.0519	.9379	.0738	.0647
.03444	.69940	.02132	.02889	.01886	.69224	.02811	.02738
2.706					11.694	2.739	
15.054	49.968	.411	2.652	.412	49.892	10.343	2.488
.2140	.9020	.0150	.0707	.0188	1.0065	.2023	.0580
.02548	.89970	.00195	.01588	.00250	.89748	.01852	.01097
43.768					45.533	42.699	
2.100	50.063	5.026	1.820	.855	49.760	1.326	1.441
.1009	.8900	.0968	.1165	.0964	.9310	.0953	.1128
.04345	.70377	.03195	.05500	.03566	.69624	.04102	.05442
.17945	.31556	.14331	.16414	.14868	.28477	.17918	.16424
2.503					10.227	2.553	
16.514	50.698	4.557	8.234	.855	49.762	8.457	4.134
.1495	.8998	.0257	.0948	.0380	.9950	.1694	.0805
.03333	.90058	.00368	.02696	.00510	.89290	.02463	.02100
.22740	.33580	.17956	.18071	.14868	.26117	.28615	.21287
6.774					16.869	13.600	
37.826	49.700	4.021	10.877	.692	50.053	.860	.952
.1124	.9074	.0881	.0949	.0780	.9260	.1018	.0878
.04741	.69718	.03358	.04168	.02886	.69342	.04259	.03841
.13250	.33536	.11114	.12760	.12035	.31667	.13155	.12800
1.912					10.406	1.952	
13.864	50.102	4.132	7.220	.692	51.176	5.755	2.633
.1216	.9027	.0233	.0868	.0308	.9828	.1284	.0604
.03250	.89946	.00339	.02176	.00413	.89626	.01579	.01465
.19758	.30886	.15912	.14004	.12035	.25708	.28675	.22258
5.776					18.755	13.751	
22.327	51.193	62.127	15.445	2.565	50.345	5.761	3.508
.1456	.9024	.1051	.1445	.1389	.9208	.1492	.1447
.08171	.70762	.05459	.07092	.06039	.69828	.08049	.07407
.14474	.74238	.07238	.13269	.09733	.64013	.15377	.13409
2.349					10.244	2.466	
28.179	57.492	64.131	18.306	2.565	51.654	7.675	7.415
.1192	.9109	.0701	.1126	.0850	.9685	.1547	.1055
.03460	.89943	.01304	.03507	.01251	.88649	.03796	.03270
.15296	.75831	.09891	.12902	.09733	.62370	.23959	.19342
2.714					12.339	9.040	
2.858	45.468	53.718	2.091	2.076	49.965	2.979	2.145
.1456	.8880	.1265	.1128	.1124	.8936	.1419	.1115
.08221	.70351	.05253	.05156	.04888	.70504	.08634	.05295
.10691	.75565	.05999	.09779	.07878	.70107	.10934	.09797
1.574					9.665	1.555	
10.989	49.300	64.649	8.231	2.076	51.507	5.784	7.174
.1282	.9061	.0720	.0910	.0688	.9507	.1201	.0838
.02417	.89979	.00948	.02372	.01013	.89889	.02073	.02177
.15675	.75140	.07681	.10618	.07878	.65389	.20697	.14982
2.641					12.434	5.835	

used in the estimation, T–1 and T–2 resp., the ASE's for the 2S-estimator for ρ are greater asymptotic standard errors in column 7 [13, 13]. Although a CV estimate of the variance is sometimes greater than the simulation variance, it is expected to be a more efficient estimate of the unknown variance.

Notice also that for most experiments, the simulation standard errors are closer to the asymptotic standard errors than the estimated standard errors. The variance of α_0 is high and usually differs substantially from its asymptotic value. In those cases, the results for σ^2 are not very satisfactory either. Whether this is an indication of the non-existence of finite sample moments of the estimators or of multicollinearity due to the autoregressive transformation has not been investigated. The bias of the 2S estimator of α_0, for $\rho \neq 0$ and $T=40$, is much greater than that of the IV or IGN estimator. Although we do not report additional results for the parameter α_0, we should mention that the bias and the standard errors are sometimes large. In general, the results for the parameters α_1, λ and ρ are satisfactory. The bias and the SE's of the 2S and IGN estimators for these parameters are very similar. The results in the table do not indicate a dominance of IGN on the 2S estimator. For the 2S and IGN estimator in model 1, the SSE's are usually smaller than the ESE's. For model 2, both are fairly good (especially when $T=60$), except for the parameter λ, for which the SSE is closer to the ASE than the ESE. The results in table 1 give an overall picture of the finite sample properties of the three estimators considered. Still, the results should not be carried over straightforwardly to other experiments.

The efficiency gain for the bias when using CV estimates can be measured in terms of the ratio of the simulation variance over the variance of the CV estimate. A variance ratio of two indicates that the gain in efficiency from using CV's is equal to that of doubling the number of runs. In general, the efficiency gain for 2S approximately equals that for IGN. For $\gamma=.95$, $\sigma^2=10$, $\rho \neq .85$ and $\lambda \neq .9$, the efficiency gain varies between 1.5 and 10 for all parameters and for all three estimators. The variance reduction is sufficiently large to justify using CV's instead of generating additional runs. When $\rho=.85$ or $\lambda=.9$, the efficiency gain for 2S and IGN sometimes becomes smaller than 1.5.

4.3. *Response Functions for Bias and Standard Errors.* In table 2, we report the estimated response functions (RF) for the bias and the standard errors for α_1 estimated by means of the three estimation methods described in section 2. The results for the parameters λ, ρ and σ^2 can be obtained on request from the authors.

The response functions given in table 2 summarize the properties of the estimators for the experiments described in section 3. The RF's are estimated using 36 experiments for model 1 and 144 experiments for model 2. In each experiment, the 60 independent samples for ε_t and η_t are reused, thereby limiting the computational costs and reducing the inter-experiment variability, which is especially useful for detecting invariances between experiments (see e.g. Hammersley and Handscomb [1964] and Hendry [1980]). Under ergodicity, the specification and the estimates for the RF's should not be seriously affected by the dependence between experiments. To avoid misspecification, very flexible forms were chosen for the RF's. Contrary to the common use of analytical RF's in simulation analysis, we use dummy variables to model the effects on the simulation results of varying parameter values and/or sample size across experiments. A dummy variables specification has several advantages compared with an analytical relationship. First, in addition to summarizing the main results of the simulations in a convenient way it allows for a common but very flexible specification for the finite sample moments of the different parameter estimates. This makes comparisons across methods, parameters and models quite easy. For instance, if the same matrix of explanatory variables is used in two different regressions, the standard errors for the regression coefficients will be the same up to a factor of proportionality. Second, it is possible to specify an analytical form for the RF by choosing a particular method to interpolate for the parameter values and/or sample size not considered in the experiments. The specifications are not restrictive in the sense that they do not imply assumptions on those regions of the parameter space for which no experiment has been carried out. Finally, a very flexible specification may be needed to characterize the relationship between small sample moments and the parameters of the model that can be very complex as has been shown in small sample theory.

When modeling the outcome of the experiments in a RF, we rely as much as possible on asymptotic properties of the estimators. As a dependent variable in the RF's for the bias, we first use the standardized variable

$$B_i = \sqrt{m}\,\frac{(\bar{\hat{\theta}}_i - \theta_i)}{ASE_i} \tag{4.15}$$

for the simulation bias, and

$$BCV_i = \sqrt{m}\,\frac{(\bar{\hat{\theta}}_i - \bar{\theta}_i^0)}{ASE_i} \tag{4.16}$$

for the CV bias, where m is equal to the number of runs for which the convergence criteria given in section 2 have been satisfied.

Usually, m is close to 60, but for values of λ close to one, m might be reduced to 40. The lowest value for m is 19 and occurs when $\lambda = .9$, $\gamma = .95$, $\rho = -.85$ and $T = 30$. Notice that the RF's for the IV and 2S methods are estimated from the results of the runs for which the IGN procedure has converged. In this way, the precision of the outcome of a given experiment is the same for the three estimation methods. In most cases where the IGN estimate does not converge, it exhibits a cyclical pattern, i.e. overestimation of a parameter alternates with underestimation. Moreover, the IV estimate is fairly precise while the 2S procedure is rather inaccurate. Finally, we notice that the Hessian matrix very rarely becomes singular. According to asymptotic distribution theory, the variable B_i in (4.15) is approximately $N(0, 1)$. The following RF has been estimated

$$y_i = \sqrt{m}\sqrt{\frac{30}{T}}D_i'\delta + u_i, \tag{4.17}$$

where the dependent variable y_i is B_i or BCV_i, D_i is a vector of dummy variables, that characterize the experiment (the dummy variable takes the value one when the parameters have the value indicated in the first row of the table 2, otherwise the dummy variable equals zero) and δ is a vector of parameters. The dummy variables included in the RF's have been chosen after a detailed analysis of the outcome of the experiments as a function of the parameters and the sample size. Their choice has also been determined by the analytical RF specifications fitted to the same data by Palm et al. [1980]. The variable $\sqrt{m}$ has been included to account for possible heteroscedasticity due to a varying number of runs. The presence of $T^{-\frac{1}{2}}$ assures that the bias is centered at zero in large samples. The factor 30 rescales the dummy variables.

When analyzing response functions for B_i and BCV_i, we came to the conclusion that the transformation of the bias by the inverse of ASE_i did not lead to homoscedasticity. The main reason why this transformation did not reduce the heteroscedasticity lies in the fact that we analyze the finite sample bias of restricted estimators. The restrictions reduce the variation of the results for the experiments for which the parameters are close to the boundaries. Our empirical finding is a first indication that the asymptotic standard errors are usually larger than the estimated finite sample standard errors of the restricted estimators. Interestingly, part of the transformation motivated by large sample theory, i.e. the premultiplication by $T^{\frac{1}{2}}$, appeared to be in accordance with the outcome of the simulations. This led us to estimating RF's of the form (4.17) for the variables

$$b_i = \sqrt{m}\sqrt{T}(\bar{\bar{\theta}}_i - \theta_i) \quad \text{and} \quad bcv_i = \sqrt{m}\sqrt{T}(\bar{\bar{\theta}}_i - \bar{\theta}_i^0) \tag{4.18}$$

respectively.

The specification of these RF's is in accordance with the consistency and possibly with the asymptotic normality of the restricted estimators. Not surprisingly, it is not in agreement with the large sample variance for the unrestricted estimators. Also, the control variate estimate of the bias is expected to have a large sample variance that is smaller than the ASE_i, so that the transformation by ASE_i^{-1} is not appropriate for this reason either. The point estimates for the parameters δ in (4.17) were similar for b_i and bcv_i. The control variate estimate of the bias, bcv_i, however, had a substantially smaller standard error of regression, indicating the gain of precision obtained by using control variates. To limit the amount of information, we decided to report only the estimated RF's for bcv_i in table 2.

An alternative procedure to increase the precision of the RF's is to model the bias for the instrumental variables and two-step estimators respectively in deviation from the bias for the iterative Gauss-Newton estimator, e.g. $\sqrt{m}\sqrt{T}(\bar{\bar{\theta}}_{IIV} - \bar{\bar{\theta}}_{IIGN})$. As the three estimators are expected to be strongly positively correlated, the difference between the bias of two estimators should have — besides a zero asymptotic mean — a smaller variance than any of the estimators taken separately. In addition, the RF's directly show how the difference between the small sample bias of two estimators varies across the experiments. Results for the differences are also reported in table 2. The gain of efficiency due to differencing is of the same order of magnitude as that obtained by control variates. Notice that because all three estimators for ρ have the same asymptotic variance $1-\rho^2$, this variable cancels from the RF when differencing.

To obtain the RF's for the estimated standard errors (ESE) and the estimated residual variance, a log-linear relationship has been estimated

$$\sqrt{m-1}y_i = \sqrt{m-1}\left[\delta_0 x_i + \left(\frac{30}{T}\right)D_i'\delta\right] + u_i, \tag{4.19}$$

where y_i denotes $\ln ESE_i$ and $\ln \hat{\sigma}_i^2$, and x_i is the corresponding limiting value. By using a log-linear relationship, we hope to achieve homoscedasticity (see Rao [1952]). In addition, Campos [1981] establishes that for large m and T, the simulation standard error (SSE) is related log-linearly to the ASE_i, with $\delta_0=1$, and to an additional term of order T^{-1} and a standardized normal disturbance term u_i. For the instrumental variables estimator we fitted the specification (4.19) to the SSE (its ESE is almost never used in practice). For the 2S and the IGN estimator, we report RF's for the ESE's that are relevant for empirical work. The use of control variate estimates for the SE's computed as $\text{SECV}=[\text{SSE}^2-\text{SDCV}^2+\text{ASE}^2]^{\frac{1}{2}}$ for the IV estimator and as $\text{SECV}=[\text{ESE}^2-\text{SDCV}^2+\text{ASE}^2]^{\frac{1}{2}}$ for the 2S and IGN estimator did not yield a substantial gain of precision.

The parameter δ_0 in (4.19) is significantly different from (and smaller than) one in many specifications. These findings which can be explained by the fact that the ASE is not an appropriate yardstick in small samples for the precision of the restricted estimators led to the following specification

$$\sqrt{m-1}(y_i-x_i) = \sqrt{m-1}\left[\delta_0\left(\frac{30}{T}\right)^{1/3} x_i + \left(\frac{30}{T}\right)D_i'\delta\right] + u_i, \tag{4.20}$$

in which the bias of the logarithm of the estimated second moments is explained by the limiting value x_i and a set of dummy variables D_i. The factor $\left(\frac{30}{T}\right)^{1/3}$ has been included to allow for a decreasing impact of x_i on the bias as sample size increases. The exponent of 1/3 has been chosen on a priori grounds to allow the impact of x_i on y_i-x_i to diminish slowly as sample size grows.

We like to emphasize that to analyze the path and the speed of convergence of the estimation bias as T increases, experiments for many more (and larger) values of T have to be carried out. This would lead to natural extensions of our study. The relationship (4.20) should be interpreted as a specification which is reasonably in agreement with the results for $T=30$ and $T=60$ and which is in accordance with consistency and asymptotic normality of the estimators. Estimates for the parameters in (4.20) are reported in table 2.

Because values of λ near the unit circle deserve special attention, we split the experiments in two groups according to whether $\lambda=.9$ or $\lambda\neq.9$, and we estimated the RF's for the two subsamples. The estimate of δ_0 appeared to be sensitive to this partition of the experiments; the estimates of the remaining coefficients are not very sensitive. Consequently, we also estimated a RF specification with two explanatory variables for the ASE, one equal to $\left(\frac{30}{T}\right)^{1/3} x_i$ when $\lambda=.9$ and zero otherwise, the second equal to zero when $\lambda=.9$ and equal to $\left(\frac{30}{T}\right)^{1/3} x_i$ otherwise. Below the point estimates for the parameters of the RF's in table 2 we report (absolute) t-values.

Finally, the RF's in table 2 have been used to predict the outcome of the independent experiments given in table 1. We report the value of

$$Q_i(l) = \frac{\sum_{j=1}^{l} (O_{ij} - P_{ij})^2}{S_i^2}, \tag{4.21}$$

where l is the number of independent experiments to be predicted ($l=4$ for model 1, $l=8$ for model 2); O_{ij} is the standardized (premultiplied by $\sqrt{T}\sqrt{m}$ for the bias, by $\sqrt{m-1}$ for the ESE) outcome of experiment j for parameter i; P_{ij} is the prediction from the response function, and S_i^2 is the residual variance of the RF. Under the assumption that the RF is correctly specified and known, $Q_i(l)$ is approximately χ^2-distributed with l degrees of freedom. As the RF's are approximations for the relationship between the outcome and characteristics of an experiment, $Q_i(l)$ should be thought of as a measure of the prediction performance of RF's rather than be taken as a formal test of hypotheses.

4.4. *Simulation evidence.* We shall briefly draw some conclusions from the results in table 2. This should not dispense the reader from having a close look at the results. We also summarize the main findings for the parameters λ, ρ and σ^2. As a rule, we use the same specification for first and second moments and for all four parameters. We could have deleted some explanatory variables at the expense of introducing some specificity. For instance, for α_1, the dummy variables for $\gamma=.7$, and for $\gamma=.7$ and $\lambda=.6$ are insignificantly different from zero in all the specifications and could have been deleted. A similar remark can be made for ρ and σ^2.

The results in table 2 can be used as follows. For instance, for a model with true parameter values $\lambda=.6$, $\rho=0$, $\gamma=.7$ and $\sigma^2=5$, the small sample bias of the IV estimator for α_1 is approximately equal to $\delta_1+\delta_6+\delta_8+\delta_{10}=.095$ (see the first row of table 2). The point estimates of the parameters in the RF's are usually small, except for α_1, λ and σ^2, where the coefficients for the dummy variables for $\lambda=.9$ and $\gamma=.95$ and for $\lambda=.9$ and $\rho=.85$ are often large. In addi tion for λ and ρ, the value $\lambda=.9$ taken separately sometimes has an important impact on the standard errors. The coefficients δ_2, δ_{13}, δ_{16} and δ_{19} are significantly different from zero in most RF's for the four parameters α_1, λ, ρ and σ^2. Moreover, for ρ and σ^2, the coefficients δ_2 and δ_{13} are positive, whereas the coefficients δ_{16} and δ_{19} are negative. For ρ, the coefficients δ_3, δ_4 and δ_5 are significantly negative.

From the empirical findings, it is obvious that for λ and γ close to the unit circle, or for ρ close to 1, the finite sample properties of the three estimators considered here differ significantly from the large sample results for the corresponding unrestricted estimators, although the specification of the RF's implies that the small sample bias vanishes asymptotically. Of course, we have to be careful in drawing conclusions about the path of finite sample moments as a function of T, as only two values for T have been considered in the experiments. Similar results have been found by Morrison [1970] for the small sample properties of Liviatan's IV estimator, a time domain version of Hannan's [1965] two-step estimator and for the iterative Steiglitz and McBride [1965] estimator in a geometric distributed lag model with uncorrelated errors.

For the SE's, the estimate of δ_0 in (4.20) is usually small, but significantly different from zero. When we condition the specification on the value of λ, the estimate of δ_0 for $\lambda=.9$ increases up to one half. This result is not predicted by large sample theory for unrestricted estimators. In our opinion, it is partly due to the restrictions imposed on the estimators. Similar problems with parameter values close to the unit circle have been encountered for univariate ARIMA models (see e.g. Ansley and Newbold [1980]). The question of whether asymptotic theory for models where the true parameter lies on the boundary of the parameter space (see Gouriēroux et al. [1982] and references therein) offers any guidance in small samples for the properties of restricted estimates deserves more attention.

The predictive power of the RF's for the bias is reasonable as is indicated by the values of $Q_i(l)$. To compute $Q_i(l)$, we have used linear interpolation. For the bias of α_1, a few values are marginally significant at the 5 percent-level, whereas for the bias of λ and σ^2, no value is significant at the 5 percent-level. For the bias of ρ, some values of Q are significant at the 1 percetnt-level. The largest values of Q occur for the standard errors for the instrumental variables estimator of α_1 and λ.

A more detailed analysis of the prediction results shows that the experiments with $\lambda=.9$ and $T=40$ usually have the largest prediction error and make the $Q_i(l)$ significant. In this context, it is legitimate to ask whether a linear interpolation procedure is entirely appropriate. The reader can of course use his own interpolation scheme. In this respect, we also like to emphasize that the relationships between finite sample moments and the values for the parameters and for T are usually complex, as has been illustrated in small sample theory. Therefore, one should be very cautious when "extrapolating" the empirical findings to situations that have not been analyzed (see Maasoumi and Phillips [1982]). Finally, we should note that $Q_i(l)$-values based on predictions of the standardized outcome of the experiments in (4.15) and (4.16) using the asymptotic $N(0, 1)$ model are usually significantly different from zero at the 5 percent-level. This indicates again that part of the asymptotic theory for the unrestricted estimators is of limited value when the estimates are restricted and sample size is small.

An important conclusion from the specifications for the differences between estimates is that there are significant effects present for the differences between

TABLE 2

RESPONSE FUNCTIONS

estimator \ coefficient δ *)		δ_1 $\lambda=.6$	δ_2 $\lambda=.9$	δ_3 $\rho=-.5$	δ_4 $\rho=.5$	δ_5 $\rho=.85$	δ_6 $\gamma=.7$	δ_7 $\gamma=.95$	δ_8 $\sigma^2=5$	δ_9 $\sigma^2=10$	δ_{10} $\lambda=.6$ $\gamma=.7$	δ_{11} $\lambda=.6$ $\gamma=.95$
bias: bcv_t model 1 equation (4.17)	IV	−.008 .283	−.109 3.566				−.035 1.232	−.066 2.339	.135 6.398	.108 5.053	.003 .066	.026 .654
	2S	−.187 2.099	−.177 1.813				.040 .443	.061 .683	.120 1.796	.128 1.887	.122 .969	.381 3.024
	IGN	.016 .196	.210 2.370				−.021 .263	.016 .191	.110 1.815	.130 2.101	.101 .887	.330 2.883
	IV–IGN	−.006 .055	−.260 2.100				−.054 .475	−.152 1.333	−.065 .762	.170 1.966	−.093 .581	−.248 1.547
	2S–IGN	−.203 2.606	−.387 4.537				.061 .781	.046 .583	.010 .170	−.001 .023	.021 .187	.051 .465
bias: bcv_t model 2 equation (4.17)	IV	.022 .524	.018 .378	.128 3.721	.223 6.500	.225 6.483	−.000 .008	−.013 .447	−.085 2.738	−.112 3.603	.022 .532	.047 1.111
	2S	−.338 5.246	.419 5.696	.131 2.510	.195 3.734	.143 2.710	.041 .890	.018 .394	−.071 1.505	−.044 .942	.095 1.477	.245 3.820
	IGN	.028 .498	.257 4.033	.012 .273	.052 1.148	.034 .737	−.026 .649	−.027 .681	.092 2.249	.131 3.213	.051 .919	.153 2.767
	IV–IGN	.184 2.243	.089 .956	−.088 1.327	−.163 2.443	−.138 2.043	.040 .686	.006 .108	.007 .119	.221 3.678	−.070 .862	−.114 1.394
	2S–IGN	−.366 6.449	.162 2.504	.119 2.583	.143 3.113	.110 2.354	.066 1.649	.045 1.117	−.163 3.922	−.176 4.230	.044 .774	.091 1.681
standard error ln ESE_t model 1 equation (4.20)	IV	.052 1.037	−.065 1.153				−.019 .373	−.037 .734	−.035 .593	−.103 1.907	−.067 .940	−.050 .696
	IV	.055 1.128	.069 .708				−.019 .380	−.037 .752	−.074 1.200	−.142 2.481	−.067 .976	−.050 .724
	2S	−.016 .090	.326 1.657				−.035 .199	−.037 .210	−.437 2.472	−.396 2.383	.053 .215	.125 .506
	2S	−.006 .063	−.765 4.526				−.016 .170	−.003 .037	.007 .063	.018 .183	.122 .949	.240 1.841
	IGN	.001 .003	.205 .909				−.043 .213	−.035 .177	−.447 2.214	−.451 2.373	.024 .085	.129 .457
	IGN	.011 .085	−.919 3.852				−.023 .176	−.001 .009	.010 .063	−.024 .169	.095 .524	.247 1.345
standard error ln ESE_t model 2 equation (4.20)	IV	−.074 1.148	−.081 1.074	.444 8.285	.549 9.931	.274 5.071	−.053 1.165	−.003 .069	−.415 7.625	−.516 10.060	.047 .731	.032 .501
	IV	−.068 1.077	.020 .235	.428 8.085	.525 9.527	.259 4.863	−.053 1.184	−.003 .062	−.447 8.153	−.548 10.567	.047 .749	.032 .513
	2S	.296 2.128	1.495 9.078	−.032 .285	−.063 .557	−.057 .500	−.000 .004	.016 .166	−.405 3.122	−.344 2.781	.040 .287	.071 .517
	2S	.348 3.453	−.026 .142	−.067 .826	−.154 1.859	−.142 1.695	.021 .296	.047 .663	.043 .415	.081 .830	.076 .760	.133 1.330
	IGN	.076 .677	.868 6.561	.008 .085	.002 .022	.003 .028	−.019 .247	.005 .064	−.277 2.655	−.285 2.873	.016 .141	.066 .600
	IGN	.110 1.198	−.145 .854	−.016 .211	−.058 .771	−.053 .701	−.005 .080	.026 .395	.021 .226	−.002 .027	.040 .437	.108 1.178

*) For instance, δ_1 is the regression coefficient of a dummy variable which takes respectively otherwise. Similarly, ASE_t denotes a variable with value $(m-1)^{1/2}(30/T)^{1/3}$ ln ASE_t, and zero otherwise. The t-values are given below the coefficient estimates.

**) The number of degrees of freedom l is 4 and 8 respectively for models 1 and 2.

FOR α_1

δ_{12} $\lambda=.9$ $\gamma=.7$	δ_{13} $\lambda=.9$ $\gamma=.95$	δ_{14} $\lambda=.6$ $\rho=-.5$	δ_{15} $\lambda=.6$ $\rho=.5$	δ_{16} $\lambda=.6$ $\rho=.85$	δ_{17} $\lambda=.9$ $\rho=-.5$	δ_{18} $\lambda=.9$ $\rho=.5$	δ_{19} $\lambda=.9$ $\rho=.85$	δ_0 ASE_l	δ_0' ASE_l $\lambda=.9$	δ_0'' ASE_l $\lambda\neq.9$	S_l^2	$Q_l(l)$**)
.183 4.316	.232 5.346										.260	2.504
.403 2.987	1.201 8.722										.827	9.754
.391 3.187	1.164 9.309										.751	3.987
−.164 .954	−.656 3.746										1.051	11.783
.013 .106	.037 .309										.722	3.612
.262 5.553	.292 5.910	−.026 .532	−.065 1.336	−.072 1.481	−.129 2.290	−.214 3.832	−.263 4.655				.543	3.768
.262 3.648	.567 7.545	.107 1.447	.235 3.184	.245 3.287	−.380 4.423	−.475 5.601	−.585 6.804				.826	4.103
.280 4.501	.846 13.013	.003 .044	−.069 1.083	−.107 1.655	−.088 1.185	−.210 2.860	−.436 5.855				.716	8.744
−.128 1.406	−.292 3.053	−.078 .830	−.178 1.898	−.166 1.752	−.120 1.099	−.212 1.966	−.162 1.483				1.052	16.719
−.018 .283	−.279 4.228	.104 1.601	.304 4.682	.351 5.361	−.292 3.859	−.265 3.549	−.149 1.970				.728	1.578
−.143 1.843	−.026 .332							−.085 4.439			.422	67.861
−.146 1.945	−.026 .344								−.037 1.067	−.103 4.801	.408	75.130
.567 2.100	1.436 4.930							−.226 4.569			1.448	11.329
.319 2.198	.906 5.451								−.451 11.909	−.034 .953	.761	8.099
.443 1.436	1.588 4.768							−.241 4.259			1.656	8.288
.188 .917	1.042 4.441								−.473 8.839	−.043 .860	1.874	7.338
−.053 .724	−.031 .404	.018 .239	.144 1.949	.193 2.578	−.099 1.133	.096 1.106	.327 3.661	.010 .504			.742	25.811
−.057 .796	−.034 .457	.017 .236	.137 1.903	.180 2.443	−.030 .336	.224 2.256	.454 4.484		.087 2.398	−.016 .755	.727	26.177
.520 3.289	1.123 6.574	−.100 .629	−.346 2.183	−.379 2.355	−.512 2.729	−1.484 7.942	−2.038 10.560	−.246 8.554			1.598	17.104
.267 2.286	.590 4.423	−.105 .911	−.408 3.549	−.463 3.971	−.399 2.934	−1.023 7.207	−1.403 9.242		−.532 15.676	−.080 3.054	1.156	19.064
.437 3.439	1.240 9.034	.001 .005	−.122 .958	−.149 1.156	−.198 1.317	−.960 6.396	−1.423 9.181	−.178 7.693			1.283	16.339
.269 2.518	.885 7.272	−.003 .024	−.163 1.555	−.206 1.932	−.124 .995	−.653 5.043	−1.001 7.225		−.368 11.889	−.067 2.847	1.055	16.792

the value $(30\,m/T)^{1/2}$ in (4.17) and $(m-1)^{1/2}(30/T)$ in (4.20) when $\lambda=.6$, and which is zero whereas ASE_l, $\lambda=.9$, denotes a variable with value $(m-1)^{1/2}(30/T)^{1/3}$ In ASE_l when $\lambda=.9$

the biases. High values of λ appear to be quite important in this respect. But the applied econometrician should evaluate the expected gain in small sample bias and the expected reduction of the small sample variance before deciding whether to choose a consistent estimator or whether to iterate an asymptotically efficient estimator. And he should not forget that the effects of specific parameter values on the finite sample first and second moments become smaller as sample size increases.

5. CONCLUSIONS

In this paper, we have investigated the finite sample behavior of three estimators for the geometric distributed lag model using Monte Carlo experiments. We have tried to increase the precision of the outcome of the experiments by using control variates derived from the asymptotic distribution of the estimators. While the CV's yield a reduction of the variance of the results, the form and the point estimates of the RF's for the CV estimates of the bias are quite similar to those for the direct simulation results. Approximately the same gain in precision has been achieved by modeling the differences between the simulation outcome for alternative estimators. Certainly, the gain in precision is lower than the increase in precision obtained by Hendry and Srba [1977]. However, a major difference between the two studies is the nonlinearity in the parameter λ of our model.

The results obtained in this simulation study enable an investigator to determine whether it is appropriate to use an efficient estimator instead of the consistent intrumental variables estimator and whether it pays to iterate the efficient estimator. However, as the effect of trended data on the properties of parameter estimates, which has recently received much attention in the literature, can be important, the conclusions for the stationary models cannot be simply carried over to nonstationary processes. The nonstationary model needs further investigation. Nevertheless, our results should in particular be relevant for situations where the series to be analyzed have been made stationary. This paper also provides an illustration of a very convenient and flexible way to summarize and explain the outcome of simulation experiments.

Our results do not give much evidence about the possible non-existence of finite sample moments of the three estimators that we have considered. Perhaps the restrictions imposed on λ and ρ assure the existence of moments in finite samples. Possibly, we obtained good estimates of the Nagar approximations to the moments (see Sargan [1978]). Finally, the response functions enable us to answer questions such as "What is a large sample?", "How large is large?" for a sample size close to the values for T considered in our study. That the answer to these questions depends on the true parameter values (or what one might think as being the true parameter values) should be obvious.

Vrije Universiteit, Amsterdam, and Katholieke Universiteit Nijmegen, The Netherlands

APPENDIX

We shall give the elements of the matrix $E(P'P)=A$ as functions of the parameters of the model (3.1). Summation goes from $t=3$ to T. Denoting the i, j th element of the symmetric matrix A by a_{ij}, we have

$$a_{11} = (T-2)(1-\rho)^2$$
$$a_{12} = E[\sum (x_t^* - \rho x_{t-1}^*)(1-\rho)] = 0$$
$$a_{13} = 0$$
$$a_{14} = 0$$
$$a_{22} = E[\sum (x_t^* - \rho x_{t-1}^*)^2] = (T-2)[(1+\rho^2)E(x_t^{*2}) - 2\rho E(x_t^* x_{t-1}^*)]$$
$$a_{23} = E[\sum \alpha(x_t^* - \rho x_{t-1}^*)(x_{t-1}^{**} - \rho x_{t-2}^{**})]$$
$$= (T-2)\alpha_1[(1+\rho^2)E(x_t^* x_{t-1}^{**}) - \rho E(x_t^* x_t^{**}) - \rho E(x_t^* x_{t-2}^{**})]$$
$$a_{24} = E[\sum (x_t^* - \rho x_{t-1}^*)u_{t-1}] = 0$$
$$a_{33} = E[\alpha_1^2 \sum (x_{t-1}^{**} - \rho x_{t-2}^{**})^2]$$
$$= (T-2)\alpha_1^2[(1+\rho^2)E(x_t^{**2}) - 2\rho E(x_t^{**} x_{t-1}^{**})]$$
$$a_{34} = E[\alpha_1 \sum (x_{t-1}^{**} - \rho x_{t-2}^{**})u_{t-1}] = 0$$
$$a_{44} = E[\sum u_{t-1}^2] = \frac{(T-2)\sigma^2}{1-\rho^2}.$$

Next, we must express the second-order moment of x_t^* and x_t^{**} as functions of the parameters λ, γ and σ_η^2. Notice that x_t^* and x_t^{**} are generated by a second and third order autoregressive process respectively with mean zero

$$x_t^* = \frac{1}{(1-\lambda L)(1-\gamma L)}\eta_t, \; x_t^{**} = \frac{1}{(1-\lambda L)^2(1-\gamma L)}\eta_t.$$

The variance of the AR(2) process x_t is given by

$$E(x_t^{*2}) = \frac{\sigma_\eta^2(1+\gamma\lambda)}{1+(\gamma\lambda)^2-\gamma\lambda-\gamma^2-\lambda^2+\gamma^3\lambda+\gamma\lambda^3-(\gamma\lambda)^3}.$$

The first order autocovariance is

$$E(x_t^* x_{t-1}^*) = \frac{\sigma_\eta^2(\gamma+\lambda)}{1+(\gamma\lambda)^2-\gamma\lambda-\gamma^2-\lambda^2+\gamma^3\lambda+\gamma\lambda^3-(\gamma\lambda)^3}.$$

The variance of x_t^{**} is

$$E(x_t^{**2}) = \frac{\sigma_\eta^2}{1-\psi_1\rho_1-\psi_2\rho_2-\psi_3\rho_3}$$

with $\psi_1 = \gamma + 2\lambda$

$$\psi_2 = -(\lambda^2 + 2\gamma\lambda)$$

$$\psi_3 = \lambda^2\gamma$$

$$\rho_1 = \frac{\psi_1}{1-\psi_2} + \frac{\psi_3}{1-\psi_2}\left[\frac{\psi_1^2+\psi_1\psi_3+\psi_2-\psi_2^2}{1-\psi_2-\psi_3(\psi_1+\psi_3)}\right]$$

$$\rho_2 = \frac{\psi_1^2+\psi_1\psi_3+\psi_2-\psi_2^2}{1-\psi_2-\psi_3(\psi_1+\psi_3)}$$

$$\rho_3 = \psi_1\rho_2 + \psi_2\rho_1 + \psi_3.$$

The first order autocovariance of x_t^{**} is

$$E(x_t^{**}x_{t-1}^{**}) = \rho_1 E(x_t^{**2}).$$

The cross-covariances are

$$E(x_t^{*}x_t^{**}) = \frac{B_1}{1-\lambda^2} + \frac{B_2}{1-\gamma\lambda}$$

where $B_1 = \dfrac{\sigma_\eta^2\lambda(1-\gamma^2)}{(\lambda-\gamma)[1+(\gamma\lambda)^2-\gamma\lambda-\gamma^2-\lambda^2+\gamma^3\lambda+\gamma\lambda^3-(\gamma\lambda)^3]}$

$$B_2 = \frac{-B_1\gamma}{\lambda}$$

$$E(x_t^{*}x_{t-1}^{**}) = \frac{B_1\lambda}{1-\lambda^2} + \frac{B_2\gamma}{1-\gamma\lambda},$$

$$E(x_t^{*}x_{t-2}^{**}) = \frac{B_1\lambda^2}{1-\lambda^2} + \frac{B_2\gamma^2}{1-\lambda\gamma}.$$

Finally notice that the matrix $E(Z'X)$ for the control variates of the IV estimator is obtained in a similar way.

REFERENCES

ANSLEY, C. F. AND P. NEWBOLD, "Finite Sample Properties of Estimators for Autoregressive Moving Average Models," *Journal of Econometrics*, 3 (1980), 159–183.

BEACH, C. M. AND J. G. MACKINNON, "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors," *Econometrica*, 46 (1978), 51–58.

BOX, G. E. P. AND G. M. JENKINS, *Time Series Analysis, Forecasting and Control* (San Francisco: Holden-Day, 1970).

CAMPOS, J., "Properties of Instrumental Variables Estimator in Dynamic Economic Systems with Autocorrelated Errors," paper presented at the Econometric Society Meeting in Amsterdam, (1981).

DHRYMES, P. J., *Distributed Lags: Problems of Estimation and Formulation* (San Francisco: Holden-Day, 1971).

———, " A note on an Efficient Two-Step Estimator," *Journal of Econometrics*, 2 (1974), 301–304.

———, KLEIN, L. R., AND K. STEIGLITZ, "Estimation of Distributed Lag Models," *International Economic Review*, 11 (1970), 235–250.

———, AND J. B. TAYLOR, "On an Efficient Two-Step Estimator for Dynamic Simultaneous

Equations Models with Autoregressive Errors," *International Economic Review*, 17 (1976), 362–376.

FISHER, R. A., *Statistical Methods for Research Workers* (Edinburgh: Oliver and Boyd, 1925).

FOMBY, T. B. AND D. K. GUILKEY, "An Examination of Two-Step Estimators for Models with Lagged Dependent Variables and Autocorrelated Errors," *Journal of Econometrics*, 22 (1983), 291–300.

GOURIÉROUX, C., HOLLY, A. AND A. MONFORT, "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on Regression Parameters," *Econometrica*, 50 (1982), 63–80.

HAMMERSLEY, J. M. AND D. C. HANDSCOMB, *Monte Carlo Methods* (London: Methuen, 1964).

HANNAN, E. J. "The Estimation of Relationships involving Distributed Lags," *Econometrica*, 33 (1965), 206–224.

HARVEY, A. C., *The Econometric Analysis of Time Series* (Oxford: Philip Allan, 1981).

——— AND I. D. MCAVINCHEY, "On the Relative Efficiency of Various Estimators of Regression Models with Moving Average Disturbances," in, Charatsis, E. D., ed., *Proceedings of the Econometric Society Meeting* (Amsterdam: North-Holland Publishing Co., 1979).

HATANAKA, M., "An Efficient Two-Step Estimator for the Dynamic Adjustment Model with Autoregressive Errors," *Journal of Econometrics*, 2 (1974), 199–220.

HENDRY, D. F., "Monte Carlo Experimentation in Econometrics," University of Louvain, CORE Discussion Paper 8035, forthcoming in, Griliches, Z., and M. D. Intriligator, eds., *Handbook of Econometrics* (Amsterdam: North-Holland Publishing Co., 1980).

——— AND F. SRBA, "The Properties of Autoregressive Instrumental Variables Estimators in Dynamic Systems," *Econometrica*, 45 (1977), 969–990.

KOYCK, L. M., *Distributed Lags and Investment Analysis* (Amsterdam: North-Holland Publishing Co., 1954).

MAASOUMI, E. AND P. C. B. PHILLIPS, "On the Behavior of Inconsistent Instrumental Variable Estimators," *Journal of Econometrics*, 19 (1982), 183–202.

MAESHIRO, A., "Small Sample Properties of Estimators of Distributed Lag Models," *International Economic Review*, 21 (1980), 721–734.

MIKHAIL, W. M., "Simulating the Small Sample Properties of Econometric Estimators," *Journal of the American Statistical Association*, 67 (1972), 620–624.

———, "A Comparative Monte Carlo Study of the Properties of Econometric Estimators," *Journal of the American Statistical Association*, 70 (1975), 94–104.

MORRISON, JR., J. L., "Small Sample Properties of Selected Distributed Lag Estimators," *International Economic Review*, 11 (1970), 13–23.

PALM, F. C., E. VOGELVANG, AND D. A. KODDE, "Efficient Estimation of the Geometric Distributed Lag Model; Some Monte Carlo Results on Small Sample Properties," Vrije Universiteit, Amsterdam, Economische Faculteit, R. M. 1980–12 (1980).

PARK, R. E. AND B. M. MITCHELL, "Estimating the Autocorrelated Error Model with Trended Data," *Journal of Econometrics*, 13 (1980), 185–201.

RAO, C. R., *Advanced Statistical Methods in Biometric Research* (New York: John Wiley and Sons, 1982).

RAO, P. AND Z. GRILICHES, "Small-Sample Properties of Several Two-Stage Regression Methods in the Context of Autocorrelated Errors," *Journal of the American Statistical Association*, 64 (1969), 253–272.

SARGAN, J. D., "On Monte Carlo Estimators of Moments which are Infinite," mimeograph, London School of Economics, (1978).

STEIGLITZ, K. AND L. B. MCBRIDE, "A Technique for Identification of Linear Systems," *IEEE Transacontions on Automatic Control*, AC–10 (1965), 461–464.