A COMPARISON OF ALTERNATIVE INSTRUMENTAL VARIABLES ESTIMATORS OF A DYNAMIC LINEAR MODEL

Kenneth D. West
University of Wisconsin

David W. Wilcox
Board of Governors of the Federal Reserve System

May 1993
Last Revised April 1994

ABSTRACT

Using a dynamic linear equation that has a conditionally homoskedastic moving average disturbance, we compare two parameterizations of a commonly used instrumental variables estimator (Hansen (1982)) to one that is asymptotically optimal in a class of estimators that includes the conventional one (Hansen (1985)). We find that for some plausible data generating processes, the optimal one is distinctly more efficient asymptotically. Simulations indicate that in samples of size typically available, asymptotic theory describes the distribution of the parameter estimates reasonably well, but that test statistics sometimes are poorly sized.

We thank Wouter den Haan and participants at the February 1994 NBER meeting on Impulses and Propagation Mechanisms for helpful comments and discussion. West thanks the National Science Foundation, the Sloan Foundation and the University of Wisconsin Graduate School for financial support. The views expressed here are those of the authors and not necessarily those of the Board of Governors of the Federal Reserve System or of other members of its staff.

1. Introduction

This paper uses asymptotic theory and simulations to evaluate instrumental variables estimators of a scalar dynamic linear equation that has a conditionally homoskedastic moving average disturbance. Equations such as the one we consider arise frequently in empirical work (e.g., the inventory papers cited below, Rotemberg (1984), Oliner et al. (1992)), as do related nonlinear equations (e.g., Epstein and Zin (1991)).

The conventional approach to estimating such equations is to specify a priori an instrument vector of fixed and finite length and select the linear combination of the instruments that is asymptotically efficient in light of the serial correlation and (when relevant) conditional heteroskedasticity of the disturbance (Hansen (1982)). We examine two versions of this estimator, the two differing only in the specification of instrument vector. We also consider a single version of an estimator that begins by defining a wide space of possible instrument vectors, and uses a data-dependent method to choose the instrument vector that is asymptotically efficient in that space (Hansen (1985)). In our application, we define this space in such a fashion that it includes the first two instrument vectors. So this estimator by definition must be at least as efficient as the other two, and in our application is strictly more efficient.

Our aims are threefold. The first is to quantify the asymptotic efficiency gains from using the optimal estimator, for some plausible data generating processes. The second is to supply simulation evidence on the finite-sample behavior of the estimators, with regard to both parameter estimates and test statistics. The third is to illustrate the implementation of the optimal estimator.

The initial impetus for this paper came from our own and others' empirical work with inventory models (indeed, the data generating processes that we use in this paper are calibrated to estimates from inventory data). A comparison of several empirical papers indicates that seemingly small changes in specification or estimation technique result in large changes in estimates (see West (1993)). But such problems do not seem to be unique to inventory applications, as is indicated by the other papers in this symposium. Also, it is known that test statistics often are poorly sized in time series models (see the references below, and the other papers in this symposium).

In some applications, it is possible to use bootstrapping rather than conventional asymptotic theory to construct test statistics. But in many applications, nonlinearities or an inability or unwillingness to simultaneously model all endogenous variables make it difficult or impossible to solve for decision rules or reduced forms; the absence of a tractable data generating process then makes such bootstrapping problematic. In any case, the quality of parameter estimates is important even in applications in which bootstrapping of test statistics is straightforward.

There is therefore a critical need to understand the finite-sample behavior of the Hansen (1982) estimator that is used in much work, and to evaluate alternative instrumental variables estimators whose asymptotic or finite-sample behavior may be preferable. Work that has considered asymptotic properties includes Hayashi and Sims (1982), who found that for some stylized data generating processes, an alternative estimator sometimes yielded dramatic asymptotic efficiency gains relative to that of Hansen (1982). Hansen and Singleton (1988) found the same, for the optimal estimator that we, too, consider.

Some earlier work has evaluated the finite-sample performance of the Hansen (1982) estimator (as well as that of another estimator that we do not consider ["iterated GMM"]), in nonlinear and linear equations with moving average (Tauchen (1986), Popper (1992), West and Wilcox (1993)) or serially uncorrelated disturbances (Kocherlakota (1990), Ferson and Forster (1991)). This work has found that asymptotic approximations to the finite-sample distributions of parameter estimates and test statistics often but not always are reasonably accurate. The nature of such discrepancies as do arise varies from paper to paper, and seems not to be easy to characterize in general terms. To our knowledge, there is no evidence on the finite-sample behavior of the other estimator that we consider.

We find that for a sample size of 300, asymptotic theory generally provides a tolerably good approximation to the finite-sample distribution of parameter estimates for all three of our estimators. For the most part, estimates are only slightly more dispersed than asymptotic theory predicts, and are centered correctly. For a sample size of 100, dispersion is rather greater and centering more erratic, but the theory still provides at least a rough guide.

In particular, then, the parameter estimates of the optimal estimator tend to be more tightly concentrated around the true parameter values than are those of the conventional one. In some but not all data generating processes, the efficiency gains from the optimal estimator are dramatic, with this estimator having asymptotic standard errors and finite-sample confidence intervals that are smaller by a factor of two than those of the conventional estimator whose instruments are the variables in the reduced form of the model.

Asymptotic theory is somewhat less successful in approximating the behavior of test statistics. Consistent with the simulations in some recent work on estimation of covariance matrices in the presence of serially correlated disturbances (e.g., Andrews (1990), Newey and West (1994)), as well as some of the simulations in Kocherlakota (1990) and Ferson and Forster (1992), we find that tests sometimes are badly sized. In one extreme case, a nominal .05 test for the conventional estimator has an actual size of about .01 even in samples of size 10,000. Overall, test statistics for the optimal estimator are sized as well (or poorly) as are those of the conventional estimator.

Three important limitations of our study should be noted. The first is that our own previous work (West and Wilcox (1993)), which used exactly the data generating processes we use here, generally gave a more pessimistic picture than do the simulations here on the distribution of the parameter estimates of one of our two versions of the conventional estimator.[2] We have selected for further analysis and comparison the best performing of the estimators that we previously studied. Taken by itself, then, this paper probably gives too supportive an evaluation of the finite sample behavior of our estimators. Second, we experiment with only a limited range of data generating processes. The contrast between the results in Kocherlakota (1990) and Tauchen (1986), both of which were motivated by the consumption-based CAPM, suggests that results may be sensitive to changes in the data generating processes. Finally, apart from a brief mention of asymptotic properties, we do not consider maximum likelihood estimation of the decision rule implied by our model. While such a technique is feasible and perhaps desirable in the context of our simple linear model, nonlinearities or an inability to model all endogenous variables makes maximum likelihood infeasible in many applications; we use our model for simplicity, but would like to develop lessons that may be applicable in much broader contexts.

The paper is organized as follows. Part 2 describes the model, solves for a reduced form that will be used to generate data and describes our data generating processes. Part 3 describes our three estimators. Part 4 displays simulation results. Part 5 presents an empirical example. Part 6 concludes. An Appendix contains some technical details. An additional appendix available on request contains some material omitted from the published paper to save space.

2. The Model and Data Generating Processes

We consider estimation of a first order condition from an inventory model studied by (among others) West (1986a), Eichenbaum (1989), Ramey (1991), Krane and Braun (1991) and Kashyap and Wilcox (1993).[2] This first order condition may be written

$$(2.1)\quad E_t\{H_t - \beta_1 X_{1t+2} - \beta_2 X_{2t+1} - \beta_3 S_{t+1} - u_t\} = 0,$$

$$X_{1t+2} \equiv -b^2 H_{t+2} + (2b^2+2b)H_{t+1} + (2b+2)H_{t-1} - H_{t-2}$$

$$- b^2 S_{t+2} + (b^2+2b)S_{t+1} - (2b+1)S_t + S_{t-1},$$

$$X_{2t+1} \equiv bH_{t+1} + H_{t-1} + bS_{t+1} - S_t.$$

In (2.1), S_t is real sales, H_t real end of period inventories, b a discount factor, $0 \leq b < 1$, E_t mathematical expectations conditional on information known at time t, u_t an i.i.d. normal cost shock that is observable to a representative firm but unobservable to the econometrician; β_1, β_2 and β_3 are parameters whose estimation is the subject of our study. In line with some of the empirical work just cited, we include deterministic terms in both our data generating processes and our econometric estimation, but suppress these terms for the moment for notational economy.

Equation (2.1) is a first order condition for optimality in inventory behavior. (See West (1993).) To close the model, we must specify a process for sales. For simplicity, we specify that sales follow an exogenous AR(2),

$$\text{(2.2a)} \qquad S_t = \phi_1 S_{t-1} + \phi_2 S_{t-2} + \epsilon_{St},$$

where ϕ_1 and ϕ_2 are such that S_t is I(0) around trend, and ϵ_{St} is the i.i.d. normal innovation in the S_t process. Application of standard techniques for solving linear rational expectations models then yields the reduced form equation for inventories (2.2b):

$$\text{(2.2b)} \qquad H_t = (\lambda_1+\lambda_2)H_{t-1} - \lambda_1\lambda_2 H_{t-2} + \pi_1 S_{t-1} + \pi_2 S_{t-2} + \epsilon_{Ht},$$

where λ_1 and λ_2 are roots of a certain fourth-order polynomial whose coefficients are functions of b, β_1, β_2 and β_3; π_1 and π_2 are certain nonlinear functions of b, λ_1, λ_2, ϕ_1 and ϕ_2. See the Appendix.

B. Generating the Synthetic Data

To generate data, we need to specify (a)the cost parameters, which are imbedded in the β's in (2.1); (b)the parameters of the forcing processes, i.e., the autoregressive coefficients of the sales process (ϕ_1 and ϕ_2) and the variance-covariance matrix of $(u_t,\epsilon_{St})'$; (c)the coefficients on deterministic terms.

In all data generating processes, the discount factor b was set to 0.995

(appropriate if the data are assumed to be monthly). We experiment with four sets of values of the β's in (2.1); these are given in Table 1A. All are based on studies using U.S. data of one sort or another. That the β's typically have a number of non-zero digits (rather than just being, say, integers) should not be construed as indicating that it is a matter of substance that the β's be exactly as indicated. Rather, the β's are nonlinear functions of some underlying economically interpretable parameters, which we set to be round numbers. A footnote, which likely will be of interest only to someone interested in reproducing the results of our study, gives these underlying parameters.[3]

In Table 1, parameter set A is roughly consistent with the estimates for post-war aggregate data in West (1990) and those for automobile data in Blanchard and Melino (1985), parameter sets B and C with those for post-war two-digit manufacturing in Ramey (1991) and West (1986a) respectively, parameter set D with those for auto data from the 1920s and 1930s in Kashyap and Wilcox (1993).

Table 1B reports parameters for exogenous processes. The autoregressive coefficients of 0.7 and 0.2 were chosen to match roughly the estimates of an AR(2) around trend fit to real sales of nondurable goods manufacturing industries, monthly, 1967-1990. The sales innovation variance of 0.120833 was chosen so that the implied unconditional variance of sales is 1 (a harmless normalization). The variance of the cost shock u_t and its correlation with the sales shock ϵ_{St} were chosen so that, in conjunction with the cost parameters of parameter set A (Table 1A), the implied ratio $var(H_t)/var(S_t)$ and the implied correlation of H_t and S_t approximately matched that of monthly nondurables manufacturing industries, 1967-1990, with H_t total inventories.

All regressions and instrument lists included a constant and trend. Thus, the reduced form used to generate the data was not literally (2.2) but

(2.3a) $S_t = \phi_1 S_{t-1} + \phi_2 S_{t-2} + \text{constant} + \text{trend} + \epsilon_{St},$

(2.3b) $H_t = (\lambda_1+\lambda_2)H_{t-1} - \lambda_1\lambda_2 H_{t-2} + \pi_1 S_{t-1} + \pi_2 S_{t-2} +$

$$\text{constant} + \text{trend} + \epsilon_{Ht}.$$

Coefficients on trend terms in (2.3a) and (2.3b) were chosen so that the implied coefficients of variation of ΔS_t and ΔH_t were each 0.2, a figure that approximately matches estimates for monthly nondurables, 1967-1990. Because different choices of cost parameters imply different autoregressive coefficients in (2.3b), the coefficient on the trend term in (2.3b) varies from data generating process to data generating process.

A complete data generating process (DGP) is specified by combining a given set of cost parameters (A, B, C or D) with the sales and cost shock processes. Given one of our four DGPs, we generate data as follows. As noted above, the vector of shocks (u_t, ϵ_{St}) is assumed to be i.i.d. normal. This implies that H_t and S_t are normally distributed. We begin by drawing a vector of initial values from the unconditional distribution of the 4x1 vector $(H_0, H_{-1}, S_0, S_{-1})'$. We then use (2.3) to generate 10,004 observations recursively.

Most of our experiments used a sample size of either 100 or 300, in which case we use observations 1 and 2 for lags, observations 103/104 or 303/304 for leads, and discard the final 10,000-104 or 10,004-304 observations. These 9700 additional observations were reserved for some additional experiments. 1000 samples were generated for each data generating process. A sample size of 300 was chosen because there are currently about 300 monthly observations available on manufacturing inventories at the two-digit level in the U.S.. The sample size of 100 was chosen for comparison.

Table 1C displays the implied values of the parameters of the inventory equation (2.2b) for each of our DGPs. The values of $\lambda_1+\lambda_2$ and $-\lambda_1\lambda_2$, the coefficients on inventories lagged once and twice, imply considerable serial correlation in inventories conditional on sales (i.e., slow adjustment of inventories to sales shocks) for DGPs A and D, mild serial correlation for DGP C, little serial correlation for DGP B.

3. Estimating the Parameters

For algebraic simplicity, we ignore constant and trend terms throughout this section. In the simulations, all equations and instrument lists included a constant and a trend.

A. Conventional Instrumental Variables

We make the first order condition (2.1) estimable by replacing expected with realized values and moving all variables but H_t to the right-hand side:

$$
\begin{aligned}
(3.1)\quad H_t &= \beta_1 X_{1t+2} + \beta_2 X_{2t+1} + \beta_3 S_{t+1} + v_{t+2},\\
&\equiv X_t'\beta + v_{t+2},\\
v_{t+2} &\equiv u_t - \beta_1(X_{1t+2}-E_tX_{1t+2}) - \beta_2(X_{2t+1}-E_tX_{2t+1}) - \beta_3(S_{t+1}-E_tS_{t+1}),\\
X_t &\equiv (X_{1t+2},X_{2t+1},S_{t+1})',\ \beta \equiv (\beta_1,\beta_2,\beta_3)'.
\end{aligned}
$$

As is typical in empirical work, we impose a value of b, which allows us to construct X_{1t} and X_{2t}; the value chosen is that used in generating the data, b=0.995. Our conventional instrumental variables (IV) estimator calculates β linearly as follows. Let Z_t be a qx1 vector of instruments. Apart from deterministic terms, q=4 or q=12 in the Monte Carlo experiments, and Z_t consists of (q/2) lags of H_t and of S_t. We let "IVq" denote the estimator (3.3) defined below, when there are q instruments:

$$
\begin{aligned}
(3.2)\quad &\text{IV4: } Z_t \equiv (H_{t-1},S_{t-1},H_{t-2},S_{t-2})',\\
&\text{IV12: } Z_t \equiv (H_{t-1},S_{t-1},H_{t-2},S_{t-2},H_{t-3},S_{t-3},H_{t-4},S_{t-4},H_{t-5},S_{t-5},H_{t-6},S_{t-6})'.
\end{aligned}
$$

(Note that the presence of the cost shock u_t invalidates the use of H_t and S_t as instruments; see (2.2).) See section B below on the rationale for use of lags beyond those in the reduced form.

Let T be the sample size (T=100 or T=300 in most of the Monte Carlo experiments). Let Z be a Txq matrix whose t'th row is Z_t', and, similarly let X =[X_t'] be the Tx3 matrix of right-hand-side variables, Y = [H_t] the Tx1 vector of the left-hand-side variable. Given Z_t, the instrumental variables estimator that has the smallest possible asymptotic variance-covariance matrix is

(3.3) $\hat{\beta} = (X'Z\hat{W}Z'X)^{-1}X'Z\hat{W}Z'Y$,

where $\hat{W}$ is a qxq matrix that is an estimate of the inverse of the spectral density at frequency zero of the qx1 vector $Z_t v_{t+2}$, i.e., the inverse of $\Sigma_{j=-\infty}^{\infty} EZ_t Z_{t-j}' v_{t+2} v_{t+2-j}$. Since the cost shock u_t is iid, $Z_t v_{t+2}$ is MA(2) and this infinite sum collapses to

$$
\begin{aligned}
(3.4)\quad W^{-1} &= \Sigma_{j=-2}^{2} EZ_t Z_{t-j}' v_{t+2} v_{t+2-j} \\
&= \Sigma_{j=-2}^{2} EZ_t Z_{t-j}' Ev_{t+2} v_{t+2-j} \\
&\equiv \Gamma_0 + (\Gamma_1 + \Gamma_1') + (\Gamma_2 + \Gamma_2') \\
&\equiv \gamma_0 C_0 + \gamma_1 (C_1 + C_1') + \gamma_2 (C_2 + C_2'),\ \gamma_j \equiv Ev_t v_{t-j},\ C_j \equiv EZ_t Z_{t-j}'.
\end{aligned}
$$

The scalar γ_j's are the same for any choice of Z_t; the matrix C_j's and Γ_j's change with different Z_t's. The asymptotic variance-covariance matrix of $T^{1/2}(\hat{\beta}-\beta)$ is then

(3.5) $V = (EX_t Z_t' W E Z_t X_t')^{-1}$,

We construct $\hat{W}$ as follows. Let $\mathring{v}_{t+2}$ be the two stage least squares residual, and let

(3.6) $\hat{\Gamma}_j = T^{-1}\Sigma_{t=j+1}^{T} Z_t Z_{t-j}' \mathring{v}_{t+2} \mathring{v}_{t+2-j}$

for $j \geq 0$. Let

$m = \min(10, [\hat{\gamma} T^{1/3}])$

where

$\hat{\gamma} = 1.1447(\hat{S}^{(1)}/\hat{S}^{(0)})^{2/3}$, $\hat{S}^{(1)} = 2\hat{\sigma}_1 + 4\hat{\sigma}_2$, $\hat{S}^{(0)} = \hat{\sigma}_0 + 2\hat{\sigma}_1 + 2\hat{\sigma}_2$,

$\hat{\sigma}_j = w'\hat{\Gamma}_j w$, $w = (1,1,1,1)'$.

We set

(3.7) $\hat{W} = \{\hat{\Gamma}_0 + \Sigma_{j=1}^{m}[1-j/(m+1)](\hat{\Gamma}_j+\hat{\Gamma}_j')\}^{-1}$.

The weights 1-j/(m+1) guarantee that $\hat{W}$ is positive definite. Newey and West (1994) provide analytical and simulation evidence on this technique for estimating W (although that paper did not consider truncating m at 10 or at any bound less than the sample size; we do that here to speed computation).

B. The Optimal Instrumental Variables Estimator

In the textbook simultaneous equations model, in which regression disturbances are iid, use of instruments other than those in the reduced form would yield no asymptotic gain and possibly a finite sample penalty. That this is not true when disturbances are serially correlated is implicitly recognized in textbook discussions of generalized least squares (here, we interpret OLS and GLS as IV estimators where the instruments are the right-hand-side variables or transformations of those variables). Hayashi and Sims (1982) pointed out that while the usual GLS estimator is inconsistent in models with moving average errors and predetermined but not exogenous instruments, a transformation similar to that of GLS can yield an estimator more efficient than that of Hansen (1982). More generally, Hansen (1985) established conditions for optimality of an instrumental variables estimator in models in which instruments are predetermined but not exogenous, and the orthogonality condition is potentially infinite dimensional. Such is the case in many time series models, including ours, in that any and all lags of predetermined variables (in our case, H_t and S_t) are legitimate instruments.

In our application, a smaller asymptotic variance-covariance matrix is obtained when a larger number of lags of H_t and S_t is used as instruments in estimation of (3.3). Thus, the asymptotic variance-covariance matrix of, say, IV6 is smaller than that of IV4, and that of IV12 is smaller still. For models in which the disturbance follows a conditionally homoskedastic moving average process, such as ours, Hansen (1985) provides a closed form expression for the linear combination of instruments that emerges asymptotically as the number of instruments used approaches infinity. Because this estimator is optimal in the class of estimators that use linear combinations of lags of H_t and S_t as instruments, we call it IV* rather than IV∞.

Let R^*_t be the (4x1) vector of reduced form variables, $R^*_t = (H_{t-1},S_{t-1},H_{t-2},S_{t-2})'$ (= the vector of instruments used in IV4). Write the second order VAR (2.2) as a first order VAR in R^*_t,

$$(3.8)\quad R^*_t = F^*R^*_{t-1} + \epsilon_t,\ \epsilon_t = (\epsilon_{Ht-1},\epsilon_{St-1},0,0)',$$

where, e.g., $F^*(1,1) = \lambda_1+\lambda_2$, $F^*(2,2) = \phi_1$. Write the moving average representation of equation (3.1)'s disturbance v_{t+2} as

$$(3.9)\quad v_{t+2} = \eta_{t+2} - \theta_1\eta_{t+1} - \theta_2\eta_t,\ \eta_t \equiv v_t-E(v_t|v_{t-1},v_{t-2},\ldots).$$

Let P^* be the 3x4 matrix of coefficients of the projection of X_t on R^*_t, $E(X_t|R^*_t) = EX_tR^{*\prime}_t(ER^*_tR^{*\prime}_t)^{-1}R^*_t = P^*R^*_t$. In our particular case, application of the general formula supplied by Hansen (1985) indicates that an optimal set of instruments Z^*_t satisfies

$$(3.10)\qquad Z^*_t = \theta_1Z^*_{t-1} + \theta_2Z^*_{t-2} + P^*(I_4-\theta_1F^*-\theta_2F^{*2})^{-1}R^*_t.$$

Any instrument vector obtained by a nonsingular linear transformation of Z^*_t is of course optimal as well. The population variance-covariance matrix resulting from use of an optimal instrument vector is

$$(3.11)\quad (EZ^*_tX_t')^{-1}(W^*)^{-1}(EX_tZ^{*\prime}_t)^{-1},$$
$$(W^*)^{-1} = \gamma_0G_0 + \gamma_1(G_1+G_1') + \gamma_2(G_2+G_2'),\ \gamma_j = Ev_tv_{t-j},\ G_j \equiv EZ^*_tZ^{*\prime}_{t-j}.$$

Asymptotic standard errors may be computed from (3.10) and (3.11) in straightforward but tedious fashion.

Observe from (3.10) that if θ_1 and θ_2 were zero, as would be the case if v_{t+2} were iid, the optimal instrument list would be the usual two stage least squares one, $Z_t^* = EX_tR_t^{*\prime}(ER_t^*R_t^{*\prime})^{-1}R_t^*$, and there would be no point in using as instruments any variables other than those in the reduced form (2.2). It follows that if θ_1 and θ_2 are close to zero, efficiency gains from using instruments other than those in the reduced form will be small, while if θ_1 and θ_2 are large, in the sense that one or both of the roots of $z^2-\theta_1z-\theta_2$ are near unity in modulus, such efficiency gains potentially will be large. For each of our four DGPs, Table 2 presents θ_1 and θ_2 along with the modulus of the larger root of $z^2-\theta_1z-\theta_2$. It may be seen that this root is smallest for DGP B, suggesting that the efficiency gains from use of IV* will be relatively small with that DGP.

Table 3 presents the ratio of the standard errors of (1)IVq for various q's to (2)IV*, for each of our four DGPs. Diminishing returns to use of instruments beyond those in the reduced form set in fairly rapidly; indeed, when 12 instruments (6 lags each of H_t and S_t) are used, the asymptotic standard errors in all cases are within 8 percent of those of IV*. On the other hand, for DGPs A, C and D, there are substantial gains to using instruments beyond the 4 in the reduced form. Table 3 persuaded us to include IV12 in our simulation analysis: in DGPs A, C and D it is much more efficient than IV4; it is roughly as efficient as IV* asymptotically, but may (and in fact does) perform worse in samples of typical size than IV*, presumably because of the large number of parameters that is estimated in the first stage regression.[3]

We operationalize (3.10) for a given artificial sample of size T as follows. (1)We estimate 4 different autoregressive systems in $(H_t,S_t)'$ by OLS, and use the Schwarz criterion to choose one of them. The specifications differ in terms of right-hand-side variables in the two equations: H_{t-1},S_{t-1},H_{t-2}; $H_{t-1},S_{t-1},H_{t-2},S_{t-2}$; $H_{t-1},S_{t-1},H_{t-2},S_{t-2},H_{t-3},S_{t-3}$; $H_{t-1},S_{t-1},H_{t-2},S_{t-2},H_{t-3},S_{t-3},H_{t-4},S_{t-4}$. Once we have chosen the order of the autoregression, we write the system as a vector AR(1). Let $\hat{F}$ be the estimated autoregressive coefficients of that system, R_t the associated variables. (Note that $\hat{F}$ has the same dimension as F^*, and $R_t=R_t^*$, only when the Schwarz criterion chooses the correct data generating process.) (2)We then obtain $\hat{P}$ from an OLS regression of X_t on R_t, $\hat{\theta}_1$ and $\hat{\theta}_2$ by maximum likelihood applied to residuals obtained from IV4. (3)Next, we use (3.10) to construct an estimate of the unconditional variance-covariance matrix of $(\hat{Z}_t^{*\prime},\hat{Z}_{t-1}^{*\prime})'$, and then draw $(\hat{Z}_0^{*\prime},\hat{Z}_{-1}^{*\prime})'$ from its unconditional normal distribution. (4)Then we use (3.10) to generate $\hat{Z}_t^*$ recursively forward from t=1 to t=T,

(3.12) $\hat{Z}_t^* = \hat{\theta}_1\hat{Z}_{t-1}^* + \hat{\theta}_2\hat{Z}_{t-2}^* + \hat{P}(I-\hat{\theta}_1\hat{F}-\hat{\theta}_2\hat{F}^2)^{-1}R_t$.

(5)Finally, we estimate β as

(3.13) $\hat{\beta} = (\Sigma_{t=1}^T\hat{Z}_t^*X_t')^{-1}\Sigma_{t=1}^T\hat{Z}_t^*H_t$.

C. Test Statistics

For our first two estimators, we construct covariance matrices and compute test statistics in a familiar way. For example, for the conventional IV estimator, an estimate of V (defined in (3.5)) is constructed as

(3.14) $\hat{V} = [(\Sigma_{t=1}^T X_tZ_t'/T)\hat{W}(\Sigma_{t=1}^T Z_tX_t'/T)]^{-1}$

for $\hat{W}$ defined in (3.7). Let $\hat{V}(i,j)$ be the (i,j) element of $\hat{V}$. The t-statistic for H_0: $\hat{\beta}_1=\beta_1$, for example, is then computed as

(3.15) $\hat{\beta}_1-\beta_1 / [\hat{V}(1,1)/T]^{1/2}$.

J-statistics, or tests of instrument-residual orthogonality, were computed for our first two estimators as

(3.16) $T^{-1}(\Sigma_{t=1}^T\hat{V}_{t+2}Z_t')\hat{W}(\Sigma_{t=1}^T Z_t\hat{V}_{t+2}) \overset{A}{\sim} \chi^2(q-3)$,

where $\hat{W}$ was constructed as described above. The test of instrument-residual orthogonality is not applicable for the optimal estimator, since the dimension of the instrument vector is identical to that of the right-hand-side variables.

For the optimal estimator, the variance-covariance matrix used in computing test statistics was

$$(3.17)\quad \hat{V}^* = (\Sigma_{t=1}^{T}\hat{Z}_t^* X_t'/T)^{-1}(\hat{W}^*)^{-1}(\Sigma_{t=1}^{T} X_t \hat{Z}_t^{*\prime}/T)^{-1}.$$

In (3.17), we initially computed $\hat{W}^*$ in a fashion analogous to (3.7). But the resulting test statistics sometimes were very poorly sized. So for $\hat{\beta}$, $\hat{\theta}_1$ and $\hat{\theta}_2$ defined in (3.12) and (3.13), we estimated $\hat{W}^*$ instead as

$$(3.18)\quad \hat{W}^* = (T^{-1}\Sigma\hat{d}_t\hat{d}_t')^{-1},$$
$$\hat{d}_t \equiv \hat{\eta}_{t+2}(\hat{Z}_t^* - \hat{\theta}_1\hat{Z}_{t+1}^* - \hat{\theta}_2\hat{Z}_{t+2}^*),$$
$$\hat{v}_{t+2} \equiv \hat{\eta}_{t+2} - \hat{\theta}_1\hat{\eta}_{t+1} - \hat{\theta}_2\hat{\eta}_t,\ \hat{v}_{t+2} \equiv H_t - X_t'\hat{\beta},\ \hat{\eta}_0 = \hat{\eta}_{-1} = 0.$$

$\hat{W}^*$ is positive semidefinite by construction. A similar estimator was suggested in Hodrick (1991). It may be shown that $\hat{W}^*$ so defined is consistent. A computation like (3.18) might also be used to estimate $\hat{W}$ or $\hat{\Omega}$. We followed conventional practice and did not: (3.17) requires estimates of the θ's, which are already available for our optimal estimator but not for the other two.

4. Simulation Evidence on Distribution of Parameter Estimates

Tables 4 and 5 present some Monte Carlo results on the distribution of the parameter estimates, Table 4 for a sample size T=100, Table 5 for T=300. They are organized by DGP. For each DGP the tables present results for three estimators, IV4 (equations (3.2)), IV12 ((3.3)) and IV* (3.12, 3.13). The "asy*" rows in each panel give asymptotic quantities for IV*, while the "asy4" row at the bottom of the table does the same for IV4; in light of Table 3, the "asy*" row applies approximately to IV12 as well.

Each estimated parameter was standardized by subtracting the population parameter value and then dividing by the IV4 population asymptotic standard error. The population rather than estimated standard error was used because our interest at the moment is in the distribution of parameter estimates rather than the distribution of test statistics. According to the asymptotic theory, the resulting quantity should be approximately N(0,1) for IV4, $N(0,(\sigma^*/\sigma_4)^2)$ for IV*, $N(0,(\sigma_{12}/\sigma_4)^2)$ for IV12, where σ^*/σ_4 and σ_{12}/σ_4 are computed from the relevant rows of Table 3. For example, in DGP A, the asymptotic theory implies that standardizing the IV* estimate of β_1 in this fashion produces a $N(0,.45^2)$ variable, where .45 = 1/(2.21); the comparable variance for IV12 is $.47^2 = (1.03/2.21)^2$.

For each of the three parameters, the columns labelled "50% CI" gives a 50 percent confidence interval constructed by dropping the largest 250 and smallest 250 of the 1000 standardized parameter estimates, or, for "asymptotic", the values appropriate for a N(0,1) variable. The difference between the upper and lower bounds of these confidence intervals is the interquartile range. "Median" gives the median of the 1000 estimates.

"Trimmed MSE" gives a mean squared error computed by (1)dropping all entries greater than 3.0 in absolute value, (2)calculating the average squared value of the remaining observations, and (3)dividing by .9735, which is the variance of a N(0,1) variable doubly truncated at -3 and +3 (Johnson and Kotz (1970, p83)). We trimmed before computing the MSE because the simultaneous equations literature indicates that second moments of our estimator may not exist, since our equation has only one more instrument than right hand side variable (e.g., Phillips (1983)). The decision to truncate at 3.0 was arbitrary; in related work, which only considered a sample size of 300 (West and Wilcox (1993)) we found little sensitivity to the exact point of truncation.

In conjunction with Table 3, we read Tables 4 and 5 as follows. First, as measured by either interquartile range (width of the 50% CIs) or trimmed MSE, the asymptotic theory underpredicts the variability of all three estimators. The discrepancies between asymptotic and simulation are larger for T=100 rather than T=300 (no surprise) and larger for β_3 than for β_1 or β_2 (for reasons that are not clear to us). Of the three estimators, the asymptotic approximation predicts variability most poorly for IV12. The trimmed MSE for this estimator is generally more than twice the approximate theoretical figure in the asy* row, as is the width of the interquartile range. By the same measures, the theory does moderately better for IV*, but better still for IV4.

On the other hand, the measures of dispersion that are probably most relevant in practice are the raw figures themselves rather than those figures relative to asymptotic theory. Our second point is that in this light IV12 is less variable than IV4, slightly so with T=100 (Table 4), more notably so with T=300. But IV* is notably less variable than IV4 and IV12 for both sample sizes (although there are occasional exceptions).

Our third point pertains to bias. For T=100, centering of parameters is a little erratic. While there does not appear to be a persistent tendency for median bias to be of a particular sign, median bias is often substantial from the point of view of asymptotic theory. In particular, if one standardizes the IV12 estimate by the corresponding asymptotic standard deviation, rather than the IV4 standard errors used in Tables 4 and 5, all 12 estimates have a median value of $\hat{\beta}_{.}-\beta_{.}$ that is more than 0.4 asymptotic standard deviations, and 10 are greater than 0.5. For all three estimators, asymmetry in the 50% CIs is also evident in Table 4. On the other hand, Table 5 indicates that while some problems remain, particularly with IV12, by and large the estimators are centered correctly for T=300.

Once again, however, the measure of bias that is more relevant in practice is that reported in the Tables, in which all parameter estimates are normalized by the same asymptotic standard error. Our fourth point, then, is that the IV12 estimator shows the most median bias, IV* the least.

Some of these points are clearly illustrated in Figure 1. For T=300, DGP A, this plots estimates of the density of the simulation estimates of the parameters (solid lines) along with the theoretical normal density suggested by the asymptotic theory (dashed lines). We constructed the simulation densities using a normal kernel and a bandwidth of .27 ≈ $1.06(1000)^{-1/5}$ = $1.06(\text{sample size})^{-1/5}$ (see, e.g., Silverman (1986)). Note that while the horizontal scales are the same on all 9 plots, the vertical scale for IV4 (row 1) is different from that for IV12 (row 2) and IV* (row 3).

The Figure illustrates that the asymptotic approximation works best for IV4, worst for IV12. The IV12/IV* discrepancies between simulation results and asymptotic theory are, however, sufficiently small that either is less disperse than is IV4. The IV4 and IV* simulation densities are noticeably better centered than are those of IV12. The appendix available on request has comparable figures for the other DGPs and for T=100. These tell a qualitatively similar story. So, too, do estimates that set the bandwidth at (.75)x(bandwidth in Figure 1) and (1.25)x(bandwidth in Figure 1).

Overall, then, IV12 probably shows the sharpest departures from asymptotic theory, perhaps because of overfitting in the first stage regression; IV4 shows the least. By our measures of variability, IV4 is worst, IV* best; by our measures of bias, IV12 is worst, IV* best. Regardless of how one trades off variability versus bias, then, IV* seems the best performing estimator. With the exception of β_3 for DGPs B and C for T=100 and T=300, IV* is better than either of the other two estimators as measured by median bias, trimmed MSE or width of 50% CI.

Table 6 presents information on the size of test statistics. Panel A presents the size of nominal .05 tests of the hypothesis H_0:β_i = population value, i=1,2,3, computed as the square of the usual t-statistic. Asymptotically each test statistic is $\chi^2(1)$, and the table reports the fraction of the 1000 simulations for which the computed statistic was greater than 3.84 (the .05 critical value for a $\chi^2(1)$ random variable). For IV4 and IV*, tests for β_1 and β_2 typically were well-behaved for both T=100 and T=300 (at least by the standards of recent work such as Newey and West (1994)!); actual sizes ranged from about .02 to about .12. For IV12, test statistics of

β_1 and β_2 were more poorly sized, especially for T=100 (e.g., in DGP C, both sizes were about .26). This latter result is perhaps unsurprising, in that Tables 4 and 5 and Figure 1 indicated that the asymptotic approximation works more poorly for parameter estimates of IV12 than for IV4 or IV*.

Those tables and that figure also indicated that all three estimators had greater difficulty estimating β_3 than β_1 or β_2. Table 6 does indeed show that tests on β_3 were generally more problematic than those on β_1 and β_2, but in a fashion that surprised us: IV4 and IV12 tend to reject not too much but too infrequently. Presumably this indicates that even though the parameter estimates are too spread out (Tables 4 and 5), the relevant entries of the variance-covariance matrices are even more spread out, the result being egregious under-rejection rather than egregious over-rejection. IV* suffers from no such problem, we conjecture due more to the way the covariance matrix was estimated (see section 3D) than to something inherent in the way the parameters were estimated.[5]

Once again, some of these points are clearly reflected in a figure, this time the Figure 2 plot of actual versus nominal sizes for DGP A, T=300. All three estimators reject slightly too much for β_1 and β_2 (columns 1 and 2), as does IV* for β_3. This applies not only to nominal .05 tests (the focus of Table 6) but to tests of nominal sizes ranging from .01 to .25. IV4 and IV12 reject much too infrequently for β_3 (column 3, rows 1 and 2). Analogous plots for other DGPs and for T=100 are available in the additional appendix, and are also consistent with Table 6.

IV4 is sufficiently simple computationally that we repeated our simulation exercise for DGP A with samples of size 10000 (relaxing the constraint on the maximum value of the bandwidth m [defined below (3.6)] in a fashion that insured consistency). Even here there was evidence of missizing for one hypothesis test: the nominal .05 t-tests on β_3 had an actual size of .007. (The comparable figures for β_1 and β_2 were .048 and .048.) It seems that for test statistics, the asymptotic approximation may work poorly even for samples that are very large relative to those of most economic time series.

Panel B of Table 6 indicates that J-tests are approximately correctly sized for IV4, poorly sized for IV12.

5. Empirical Example

Here, we apply the IV4, IV12 and IV* estimators to aggregate inventories and sales of nondurables manufacturing industries, monthly, seasonally adjusted. After accounting for lags and leads, the sample was 1967:3-1992:10. In applying IV*, we used the procedure described in section 2, including use of the Schwarz criterion, which happened to choose $R_t = (H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2})'$. Our aim is not to provide a reinterpretation or even a refinement of the existing inventory literature, but merely to underscore the ease with which the IV* estimator can be applied.

The first row of Table 7 has the IV4 estimates. The estimates of the moving average parameter yield an implied larger root of 0.42, about the same size as that for DGP B (see Table 2). Accordingly, dramatic efficiency gains in going to either IV12 or IV* are not to be expected. Lines 2 and 3 bear out this expectation. While the t-statistic on β_1 becomes noticeably larger, that on β_2 falls and that on β_3 falls for IV*, rises for IV12. Upon transforming the estimates of the β_i's to the underlying economic parameters (see footnote 1), we find that two of the four underlying parameters are positive for IV4, three of four for IV12 and IV*. Some (but not all) investigators have argued that all four underlying parameters should be positive.[6]

6. Conclusions

This paper has compared several estimators of a dynamic linear model. For all our estimators, the asymptotic theory characterizes the distribution of parameter estimates tolerably well. But test statistics occasionally are very poorly sized. The recommended estimator would seem to be the one that is most efficient. This is the estimator suggested by Hansen (1985), which for three of our four data generating processes yielded substantial asymptotic and finite-sample benefits relative to conventional instrumental variables estimators.

Because earlier, related work has found sensitivity of results to choice of data generating processes, one priority for future work includes experimentation with additional data generating processes. Other priorities include development of alternative methods of computing test statistics and of refined asymptotics to better characterize finite sample distributions.

References

Andrews, Donald W. K., 1991, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," Econometrica 59, 817-858.

Blanchard, Olivier J., and Angelo Melino, 1986, "The Cyclical Behavior of Prices and Quantities: The Case of the Automobile Market," Journal of Monetary Economics 17, 379-408.

Eichenbaum, Martin S., 1989, "Some Empirical Evidence on the Production Level and Production Cost Smoothing Models of Inventory Investment," American Economic Review 79, 853-864.

Epstein, Larry G. and Stanley E. Zin, 1991, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," Journal of Political Economy 99, 263-286.

Ferson, Wayne E. and Stephen E. Foerster, 1991, "Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models," manuscript, University of Chicago.

Hansen, Lars Peter, 1982, "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica 50, 1029-54.

Hansen, Lars Peter, 1985, "A Method for Calculating Bounds on the Asymptotic Variance-Covariance Matrices of Generalized Method of Moments Estimators," Journal of Econometrics 30, 203-228.

Hansen, Lars Peter, and Kenneth J. Singleton, 1988, "Computing Semi-Parametric Efficiency Bounds for Linear Time Series Models," manuscript, University of Chicago.

Hansen, Lars Peter, and Kenneth J. Singleton, 1990, "Efficient Estimation of Linear Asset Pricing Models with Moving-Average Errors," NBER Technical

Working Paper No. 86.

Hayashi, Fumio and Christopher A. Sims, 1983, "Nearly Efficient Estimation of Time Series Models with Predetermined, but not Exogenous, Instruments", _Econometrica_ 51, 783-798,

Hillier, Grant H., 1990, "On the Normalization of Structural Equations: Properties of Direction Estimators," _Econometrica_ 58, 1181-1194.

Hodrick, Robert A., 1991, "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," NBER Technical Working Paper No. 108.

Johnson, Norman L., and Samuel Kotz, 1970, _Continuous Univariate Distributions-1_, Boston: Houghton Mifflin Company.

Kashyap, Anil K and David W. Wilcox, 1993, "Production and Inventory Control at the General Motors Corporation During the 1920s and 1930s," _American Economic Review_ 83, 383-401.

Kocherlakota, Narayana, 1990, "On Tests of Representative Consumer Asset Pricing Models," _Journal of Monetary Economics_ 26, 285-304.

Krane, Spencer D. and Steven N. Braun, 1991, "Production Smoothing Evidence from Physical Product Data," _Journal of Political Economy_ 99, 558-581.

Newey, Whitney K., and Kenneth D. West, 1994, "Automatic Lag Selection in Covariance Matrix Estimation," forthcoming, _Review of Economic Studies_.

Oliner, Stephen, Glenn Rudebusch and Daniel Sichel, 1992, "The Lucas Critique Revisited: Assessing the Stability of Empirical Euler Equations," manuscript, Federal Reserve Board of Governors.

Phillips, Peter C. B., 1983, "Exact Small Sample Theory in the Simultaneous Equations Model," 449-516 in Z. Griliches and M.D. Intriligator (eds.) Handbook of Econometrics Volume 1, Amsterdam: North Holland.

Popper, Helen, 1992, "Formulating Nonlinear Moment Conditions for Generalized Method of Moments," manuscript, Santa Clara University.

Ramey, Valerie A., 1991, "Nonconvex Costs and the Behavior of Inventories," Journal of Political Economy 99, 306-334.

Rotemberg, Julio, 1984, "Interpreting the Statistical Failures of Some Rational Expectations Macroeconomic Models," American Economic Review 74 (May), 188-193.

Silverman, B.W., 1986, Density Estimation for Data Analysis, London: Chapman Hall.

Tauchen, George, 1986, "Statistical Properties of Generalized Method of Moments Estimators of Structural Parameters Obtained from Financial Markets Data," Journal of Business and Economic Statistics 4, 397-425.

Theil, Henri, 1972, Principles of Econometrics, New York: John Wiley and Sons.

West, Kenneth D., 1986a, "A Variance Bounds Test of the Linear Quadratic Inventory Model," Journal of Political Economy 94, 374-401.

West, Kenneth D., 1986b, "Full versus Limited Information Estimation of a Rational Expectations Model," Journal of Political Economy 94, 374-401.

West, Kenneth D., 1990, "The Sources of Fluctuations in Aggregate Inventories and GNP," Quarterly Journal of Economics CV 939-972.

West, Kenneth D., 1993, "Inventory Models," NBER Technical Working Paper No.

143.

West, Kenneth D., and David W. Wilcox, 1993, "Some Evidence on Finite Sample Distributions of Instrumental Variables Estimators of the Linear Quadratic Inventory Model," forthcoming in Inventory Cycles and Monetary Policy, R. Fiorito (ed.), Berlin: Springer-Verlag.

Appendix

The reduced form parameters in (2.2b) relate to the underlying cost parameters as follows: Let a_0, a_1, a_2 and a_3 be defined as in footnote 1. Then λ_1 and λ_2 are the two smallest (in modulus) roots of:

$$\lambda^4 - b^{-2}a_0^{-1}[ba_1+2a_0b(1+b)]\lambda^3 + b^{-2}a_0^{-1}[a_0(1+4b+b^2)+a_1(1+b)+ba_2]\lambda^2$$
$$- b^{-2}a_0^{-1}[a_1+2a_0(1+b)]\lambda + b^{-2} = 0.$$

Define the scalars ρ_1, ρ_2, w_1, w_2, w_3, and w_4, the (1x2) vector e' and the (2x2) matrices Φ and D as

$$\rho_1 = \lambda_1+\lambda_2,\ \rho_2 = -\lambda_1\lambda_2,\ w_1 = b^2\rho_2,\ w_2 = -\rho_2[b^2+2b+b(a_1/a_0)+(ba_2a_3/a_0)],$$

$$w_3 = \rho_2[2b+1+(a_1/a_0),\ w_4 = -\rho_2,\ e' = (1\ 0),$$

$$\Phi = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix},\qquad D = [I-b\rho_1\Phi-b\rho_2\Phi^2]^{-1}.$$

Then

$$(\pi_1,\pi_2)' = e'D(w_1\Phi^3+w_2\Phi^2+w_3\Phi+w_4I),$$

$$\epsilon_{Ht} = (\rho_2/a_0)u_t + (\pi_2/\phi_2)\epsilon_{St}.$$

Footnotes

1. That paper only considered one of our two versions of the conventional estimator, for one of the two sample sizes we consider here, and did not consider test statistics at all.

2. This model and data generating processes were also used in West and Wilcox (1993), so some of the entries in the Tables, and some of the prose in this section, are also found in that paper.

3. The underlying parameters are denoted a_0, a_1, a_2 and a_3. Let $c \equiv a_0(1+4b+b^2)+a_1(1+b)+ba_2$. Then $\beta_1=a_0/c$, $\beta_2=a_1/c$, $\beta_3=ba_2a_3/c$. The values of the a_i's in each DGP: A: a_0=1., a_1=0.1, a_2=0.1, a_3=0.1; B: a_0=1., a_1=-2.0, a_2=6., a_3=1.0; C: a_0=1., a_1=0.1, a_2=2., a_3=0.1; B: a_0=1., a_1=-0.5, a_2=0.1, a_3=0.5.

4. The class of estimators in which IV* is optimal does not include full information maximum likelihood (FIML), which gains additional efficiency by exploiting the cross-equation restrictions of the $(H_t,S_t)'$ process. For β_1, for example, the ratios of the asymptotic standard errors of FIML to IV* are: A: 0.90; B: 0.84; C:0.69; D: 0.90. The ratios for the other parameters are comparable. For our DGPs, then, the efficiency gains associated with IV* relative to FIML are modest.

5. We repeated the Table 6 calculation for T=300 for IV*, calculating equation 3.17's $\hat{W}^*$ in a fashion analogous to that described in equations 3.6 and 3.7. The size of Table 6's nominal .05 tests on β_1, β_2 and β_3 were: DGP A, .001, .000, .001; DGP B, .040, .037, .044; DGP C, .023, .022, .003; DGP D, .000, .000, .001. We do not know, however, whether use of estimator (3.18) would result in a similar improvement of the sizes of the test statistics for IV4 and IV12.

6. For each estimator, at least one of the four must be positive by construction. Additional information, of interest only to aficionados of the linear-quadratic inventory model: the parameter estimates that were positive for both IV* and IV4 were the costs of production and of changing production. The inventory holding cost estimate was positive for IV* as well. The estimate of the parameter that determines the target inventory-sales ratio was

negative. Since the simulations found it particularly difficult to get a reliable estimate of β_3, it may be noteworthy that the two parameter estimates that were positive may be inferred from the estimates of β_1 and β_2, without use of the estimate of β_3, while the estimates that were negative relied in part on the estimate of β_3.

Table 1

Data Generating Processes

A. Parameters of Cost Function

Mnemonic	β_1	β_2	β_3
A	0.160	0.016	0.002
B	0.126	-0.252	0.376
C	0.099	0.199	0.01001
D	0.197	-0.099	0.010

B. Parameters of Exogenous Processes

ϕ_1	ϕ_2	var(ϵ_S)	var(u)	$\rho(\epsilon_S,u)$
.75	.20	.120833	3.5	-.5

C. Implied Coefficients of Inventory Equation

DGP	$\lambda_1+\lambda_2$	$-\lambda_1\lambda_2$	π_1	π_2
A	1.22	-0.42	0.14	-0.12
B	0.24	-0.14	0.38	0.05
C	1.07	-0.22	0.10	-0.09
D	1.43	-0.69	0.33	-0.15

Notes:

1. β_1, β_2 and β_3 are the regression parameters in (3.1); ϕ_1 and ϕ_2 are the autoregressive parameters of the sales process defined in (2.2); ϵ_S is the sales shock defined in (2.2); u is the cost shock defined in (2.1); $\lambda_1+\lambda_2$, $-\lambda_1\lambda_2$, π_1, and π_2 are the coefficients of the reduced form inventory equation (2.3b).

Table 2

Parameters of the MA(2) Disturbance

DGP	θ_1	θ_2	Modulus of Larger Root
A	1.27	-0.45	0.67
B	0.50	-0.19	0.43
C	0.93	-0.18	0.67
D	1.44	-0.71	0.85

Notes:

1. θ_1 and θ_2 are the parameters of the MA(2) disturbance v_{t+2}; see (3.1) and (3.9). The modulus presented is that of the larger of the two roots to $z^2-\theta_1 z-\theta_2 = 0$.

Table 3

Asymptotic Standard Errors, IVq Relative to IV*

DGP	Estimator	Parameter β_1	β_2	β_3
A	IV4	2.21	2.26	1.40
A	IV6	1.46	1.47	1.13
A	IV8	1.19	1.20	1.05
A	IV12	1.03	1.03	1.01
B	IV4	1.12	1.10	1.02
B	IV6	1.00	1.00	1.00
B	IV8	1.00	1.00	1.00
B	IV12	1.00	1.00	1.00
C	IV4	1.49	1.51	1.31
C	IV6	1.16	1.17	1.10
C	IV8	1.06	1.07	1.04
C	IV12	1.01	1.01	1.01
D	IV4	3.02	2.99	1.31
D	IV6	1.67	1.63	1.07
D	IV8	1.23	1.22	1.03
D	IV12	1.08	1.08	1.03

Notes:

1. IVq is the conventional instrumental variables estimator described in (3.2) and (3.3), where Z_t consists of q instruments (q=4,6,8, or 12); IV* is the

optimal estimator described in (3.12) and (3.13). The table presents the ratio of the square roots of the diagonal elements of (1)the variance-covariance matrix of IVq (computed according to (3.5)), to (2)the variance-covariance matrix of IV* (computed according to (3.11)).

Table 4

Distributions of Standardized Parameter Estimates, From Simulations, T=100

DGP/ Estimator	$\hat{\beta}_1-\beta_1$ 50% CI	Median	Trimmed MSE	$\hat{\beta}_2-\beta_2$ 50% CI	Median	Trimmed MSE	$\hat{\beta}_3-\beta_3$ 50% CI	Median	Trimmed MSE
A									
IV4	(-0.6,0.8)	0.07	1.14	(-0.8,0.6)	-0.10	1.20	(-3.1,1.6)	-0.68	2.51
IV12	(-0.3,0.9)	0.29	0.92	(-0.9,0.3)	-0.30	0.98	(-4.3,0.8)	-1.26	2.52
IV*	(-0.4,0.4)	0.04	0.61	(-0.5,0.4)	-0.06	0.63	(-2.7,1.6)	-0.39	2.27
asy*	(-0.3,0.3)	0.00	0.21	(-0.3,0.3)	0.00	0.20	(-0.5,0.5)	0.00	0.51
B									
IV4	(-0.5,1.0)	0.27	1.13	(-1.1,0.4)	-0.38	1.24	(-0.5,1.2)	0.23	1.47
IV12	(-0.2,1.0)	0.47	0.94	(-1.4,-0.0)	-0.63	1.34	(-0.2,1.7)	0.67	1.79
IV*	(-0.5,0.7)	0.14	0.93	(-0.9,0.4)	-0.22	1.09	(-0.5,1.1)	0.24	1.33
asy*	(-0.6,0.6)	0.00	0.79	(-0.6,0.6)	0.00	0.82	(-0.7,0.7)	0.00	0.97
C									
IV4	(-0.7,0.8)	0.11	1.14	(-0.8,0.6)	-0.13	1.17	(-1.1,2.8)	0.66	2.41
IV12	(0.0,1.0)	0.54	0.98	(-1.1,-0.1)	-0.59	1.06	(-2.3,1.0)	-0.59	2.18
IV*	(-0.8,0.5)	-0.06	0.91	(-0.5,0.8)	0.03	0.91	(-0.4,3.9)	1.19	2.06
asy*	(-0.5,0.5)	0.00	0.45	(-0.4,0.4)	0.00	0.44	(-0.5,0.5)	0.00	0.59
D									
IV4	(-0.6,0.9)	0.11	1.23	(-0.9,0.6)	-0.11	1.28	(-3.1,0.5)	-1.02	2.30
IV12	(-0.4,0.8)	0.15	1.05	(-0.8,0.5)	-0.15	1.06	(-3.6,0.3)	-1.39	2.38
IV*	(-0.3,0.4)	0.05	0.53	(-0.5,0.3)	-0.07	0.55	(-2.6,0.5)	-0.84	2.11
asy*	(-0.2,0.2)	0.00	0.11	(-0.2,0.2)	0.00	0.11	(-0.5,0.5)	0.00	0.58
asy4	(-0.7,0.7)	0.00	1.00	(-0.7,0.7)	0.00	1.00	(-0.7,0.7)	0.00	1.00

Notes:

1. The estimating equations are: IV4, (3.3); IV12, (3.3); IV*, (3.13).

2. The difference between estimated and population parameter is standardized by dividing by asymptotic standard error for IV4.

3. The "50% CI" is a 50 percent confidence interval constructed using the 250'th and 750'th largest of the 1000 estimates; "Median" is the 500'th largest such entry; "Trimmed MSE" is a mean squared error computed after dropping observations greater than 3.0 in absolute value, and is expressed relative to the MSE for a standard normal similarly trimmed.

4. "asy*" presents the asymptotic values for IV* and (approximately) IV12, which vary from DGP to DGP because the ratio of standard errors of IV* to IV4 varies from DGP to DGP (see Table 3 and the text). "asy4" presents the asymptotic values for

IV4.

Table 5

Distributions of Standardized Parameter Estimates, From Simulations, T=300

DGP/ Estimator	$\hat{\beta}_1-\beta_1$ 50% CI	Median	Trimmed MSE	$\hat{\beta}_2-\beta_2$ 50% CI	Median	Trimmed MSE	$\hat{\beta}_3-\beta_3$ 50% CI	Median	Trimmed MSE
A									
IV4	(-0.7,0.8)	0.14	1.14	(-0.8,0.6)	-0.15	1.14	(-1.2,0.9)	-0.17	1.76
IV12	(-0.2,0.7)	0.29	0.57	(-0.7,0.2)	-0.29	0.57	(-1.3,0.6)	-0.32	1.64
IV*	(-0.3,0.4)	0.08	0.36	(-0.4,0.3)	-0.09	0.35	(-0.9,0.6)	-0.14	1.38
asy*	(-0.3,0.3)	0.00	0.21	(-0.3,0.3)	0.00	0.20	(-0.5,0.5)	0.00	0.51
B									
IV4	(-0.6,0.9)	0.21	1.10	(-0.9,0.5)	-0.30	1.16	(-0.6,0.9)	0.14	1.21
IV12	(-0.3,1.0)	0.41	0.98	(-1.1,0.2)	-0.49	1.15	(-0.4,1.2)	0.37	1.38
IV*	(-0.5,0.7)	0.14	0.84	(-0.8,0.4)	-0.22	0.92	(-0.5,0.9)	0.18	1.11
asy*	(-0.6,0.6)	0.00	0.79	(-0.6,0.6)	0.00	0.82	(-0.7,0.7)	0.00	0.97
C									
IV4	(-0.7,0.8)	0.10	1.13	(-0.8,0.7)	-0.11	1.14	(-0.6,1.5)	0.36	1.66
IV12	(-0.2,0.9)	0.44	0.78	(-0.9,0.1)	-0.45	0.78	(-0.8,0.9)	-0.07	1.33
IV*	(-0.7,0.4)	-0.06	0.70	(-0.4,0.7)	0.04	0.70	(-0.1,1.7)	0.58	1.39
asy*	(-0.5,0.5)	0.00	0.45	(-0.4,0.4)	0.00	0.44	(-0.5,0.5)	0.00	0.59
D									
IV4	(-0.6,0.8)	0.15	1.14	(-0.8,0.6)	-0.15	1.16	(-1.3,0.4)	-0.34	1.53
IV12	(-0.3,0.6)	0.16	0.48	(-0.6,0.3)	-0.15	0.49	(-1.3,0.4)	-0.43	1.56
IV*	(-0.2,0.3)	0.06	0.25	(-0.3,0.2)	-0.06	0.26	(-1.1,0.4)	-0.29	1.28
asy*	(-0.2,0.2)	0.00	0.11	(-0.2,0.2)	0.00	0.11	(-0.5,0.5)	0.00	0.58
asy4	(-0.7,0.7)	0.00	1.00	(-0.7,0.7)	0.00	1.00	(-0.7,0.7)	0.00	1.00

See notes to Table 4.

Table 6

A. Size of Nominal .05 T-Tests, from Simulations

		T=100			T=300		
DGP	Estimator	β_1	β_2	β_3	β_1	β_2	β_3
A	IV4	0.061	0.056	0.004	0.063	0.060	0.001
	IV12	0.142	0.134	0.020	0.080	0.067	0.004
	IV*	0.080	0.077	0.068	0.076	0.073	0.087
B	IV4	0.073	0.072	0.065	0.065	0.053	0.061
	IV12	0.193	0.260	0.231	0.103	0.104	0.106
	IV*	0.115	0.073	0.059	0.056	0.052	0.066
C	IV4	0.073	0.071	0.010	0.075	0.076	0.011
	IV12	0.264	0.264	0.031	0.158	0.154	0.014
	IV*	0.115	0.119	0.053	0.080	0.082	0.050
D	IV4	0.020	0.018	0.002	0.037	0.028	0.002
	IV12	0.047	0.044	0.026	0.017	0.015	0.006
	IV*	0.083	0.086	0.091	0.118	0.117	0.114

B. Size of Nominal .05 J-Tests, from Simulations

		T=100	T=300
DGP	Estimator	J-size	J-size
A	IV4	0.041	0.056
	IV12	0.001	0.001
B	IV4	0.052	0.061
	IV12	0.004	0.023
C	IV4	0.039	0.051
	IV12	0.003	0.001
D	IV4	0.042	0.055
	IV12	0.001	0.000

Notes:

1. In each of 1000 simulations, we computed t-statistics testing whether each of the three β_is equals its Table 1A population value. Panel A presents the fraction of simulations in which the square of the t-statistic exceeded 3.84, which is the .05 critical value for a $\chi^2(1)$ random variable.

2. In each of 1000 simulations, tests of instrument-residual orthogonality were computed as in (3.16). Panel B presents the fraction of simulations in which the resulting statistic was greater than 3.84 (IV4) or 16.92 (IV12), which are the .05 critical values for $\chi^2(1)$ and $\chi^2(9)$ random variables.

Table 7

Estimates of Aggregate Nondurables in Manufacturing, 1967-1992

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Estimator	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\theta}_1$	$\hat{\theta}_2$	Modulus of Larger Root
IV4	0.114	0.160	0.004	0.84	-0.35	0.42
	(0.044)	(0.134)	(0.008)			
IV12	0.155	0.036	-0.004			
	(0.016)	(0.048)	(0.004)			
IV*	0.145	0.068	0.001			
	(0.024)	(0.071)	(0.006)			

Notes:

1. The table presents estimates of IV4, IV12 and IV*, computed according to (3.3) and (3.13). The vector $\hat{R}_t$ (defined above (3.12)) was the set of lags that maximized the Schwarz criterion, where the following four sets were considered: H_{t-1},S_{t-1},H_{t-2}; $H_{t-1},S_{t-1},H_{t-2},S_{t-2}$; $H_{t-1},S_{t-1},H_{t-2},S_{t-2},H_{t-3},S_{t-3}$; $H_{t-1},S_{t-1},H_{t-2},S_{t-2},H_{t-3},S_{t-3},H_{t-4},S_{t-4}$. All four also included intercept and trend.

2. Columns (5)-(7) are as described in Table 2, and are estimated from the two stage least squares residuals.