

# A Comparison of Alternative Instrumental Variables Estimators of a Dynamic Linear Model

**Kenneth D. WEST**

Dept. of Economics, University of Wisconsin, Madison, WI 53706

**David W. WILCOX**

Division of Monetary Affairs, Board of Governors of the Federal Reserve System, Washington, DC 20551

Using a dynamic linear equation that has a conditionally homoscedastic moving average disturbance, we compare two parameterizations of a commonly used instrumental variables estimator to one that is asymptotically optimal in a class of estimators that includes the conventional one. 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.

**KEY WORDS:** Asymptotic approximation; Efficient estimation; Optimal estimation; Simulation; Test statistics.

This article uses asymptotic theory and simulations to evaluate instrumental variables (IV) estimators of a scalar dynamic linear equation that has a conditionally homoscedastic moving average (MA) disturbance. Equations such as the one we consider arise frequently in empirical work (e.g., the inventory articles cited hereafter; Oliner, Rudebusch, and Sichel 1992; Rotemberg 1984); 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 heteroscedasticity 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 it 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 (DGP's). 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 article came from our own and others' empirical work with inventory models (indeed, the DGP's that we use in this article are calibrated to estimates from inventory data). A comparison of several empirical studies indicates that seemingly small changes in specification or estimation technique result in large changes in estimates (see West in press). But such problems do not seem

to be unique to inventory applications, as is indicated by the other articles in this symposium. Moreover, it is known that test statistics often are poorly sized in time series models (see the references following and the other articles 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 DGP 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 behavior of the Hansen (1982) estimator that is used in much work and to evaluate alternative IV estimators whose asymptotic or finite-sample behavior may be preferable. Work that has considered asymptotic properties includes that of Hayashi and Sims (1983), who found that, for some stylized DGP's, 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 (MA) (Popper 1992; Tauchen 1986; West and Wilcox 1994) or serially uncorrelated disturbances (Ferson and Foerster 1991; Kocherlakota 1990). This work has found that asymptotic approximations to the finite-sample distributions of parameter estimates and test statistics often, but not always, are

---

© 1996 American Statistical Association  
Journal of Business & Economic Statistics  
July 1996, Vol. 14, No. 3

reasonably accurate. The nature of such discrepancies as do arise varies from study to study 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 DGP's, 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 (1991) and Newey and West (1994), as well as some of the simulations of Kocherlakota (1990) and Ferson and Foerster (1991)], 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 1994), which used exactly the DGP's 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. We have selected for further analysis and comparison the best performing of the estimators that we previously studied. Taken by itself, then, this article probably gives too supportive an evaluation of the finite-sample behavior of our estimators. Second, we experiment with only a limited range of DGP's. The contrast between the results of Kocherlakota (1990) and Tauchen (1986), both of whom were motivated by the consumption-based capital asset-pricing model, suggests that results may be sensitive to changes in the DGP's. Finally, apart from a brief mention of asymptotic properties, we do not consider maximum likelihood estimation of the decision rule implied by our model. Although 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 article is organized as follows. Section 1 describes the model, solves for a reduced form that will be used to generate data, and describes our DGP's. Section 2 describes our three estimators. Section 3 displays simulation results. Section 4 presents an empirical example. Section 5 concludes. An appendix contains a brief overview of the economic model underlying our econometric work. An additional appendix, available on request, contains some material omitted from the published article to save space.

## 1. THE MODEL AND DATA-GENERATING PROCESSES

### 1.1 Model

As described in the Appendix, 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). This first-order condition may be written

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

$$\begin{aligned} X_{1t+2} \equiv & -b^2 H_{t+2} + (2b^2 + 2b)H_{t+1} + (2b + 2)H_{t-1} \\ & - H_{t-2} - b^2 S_{t+2} + (b^2 + 2b)S_{t+1} \\ & - (2b + 1)S_t + S_{t-1}; \end{aligned}$$

$$X_{2t+1} \equiv bH_{t+1} + H_{t-1} + bS_{t+1} - S_t. \quad (1.1)$$

In (1.1),  $S_t$  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  $t$ , and  $u_t$  an iid normal cost shock that is observable to a representative firm but unobservable to the econometrician;  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  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 DGP's and our econometric estimation but suppress these terms for the moment for notational economy.

Equation (1.1) is a first-order condition for optimality in inventory behavior. (See the Appendix and West in press.) To close the model, we must specify a process for sales. For simplicity, we specify that sales follow an exogenous second-order autoregressive [AR(2)] process,

$$S_t = \phi_1 S_{t-1} + \phi_2 S_{t-2} + \varepsilon_{St}, \quad (1.2a)$$

where  $\phi_1$  and  $\phi_2$  are such that  $S_t$  is  $I(0)$  around trend and  $\varepsilon_{St}$  is the iid 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,

$$\begin{aligned} H_t = & (\lambda_1 + \lambda_2)H_{t-1} - \lambda_1 \lambda_2 H_{t-2} \\ & + \pi_1 S_{t-1} + \pi_2 S_{t-2} + \varepsilon_{Ht}, \end{aligned} \quad (1.2b)$$

where  $\lambda_1$  and  $\lambda_2$  are roots of a certain fourth-order polynomial whose coefficients are functions of  $b$ ,  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  and

Table 1. Data-Generating Processes: Parameters of Cost Function

Mnemonic	$\beta_1$	$\beta_2$	$\beta_3$
A	.160	.016	.002
B	.126	-.252	.376
C	.099	.199	.01001
D	.197	-.099	.010

NOTE:  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  are the regression parameters in (2.1).

$\pi_1$  and  $\pi_2$  are certain nonlinear functions of  $b$ ,  $\lambda_1$ ,  $\lambda_2$ ,  $\phi_1$ , and  $\phi_2$ . See the Appendix.

## 1.2 Generating the Synthetic Data

To generate data, we need to specify (a) the cost parameters, which are imbedded in the  $\beta$ 's in (1.1); (b) the parameters of the forcing processes—that is, the autoregressive coefficients of the sales process ( $\phi_1$  and  $\phi_2$ ) and the variance-covariance matrix of  $(u_t, \varepsilon_{St})'$ ; (c) the coefficients on deterministic terms.

In all DGP's, the discount factor  $b$  was set to .995 (appropriate if the data are assumed to be monthly). We experiment with four sets of values of the  $\beta$ 's in (1.1); these are given in Table 1. All are based on studies using U.S. data of one sort or another. That the  $\beta$ 's typically have several nonzero digits (rather than just being, say, integers) should not be construed as indicating that it is a matter of substance that the  $\beta$ 's be exactly as indicated. Rather, the  $\beta$ 's are nonlinear functions of some underlying economically interpretable parameters, which we set to be round numbers. A paragraph in the Appendix, which likely will be of interest only to someone interested in reproducing the results of our study, gives these underlying parameters.

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

Table 2 reports parameters for exogenous processes. The autoregressive coefficients of .7 and .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 .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  $\varepsilon_{St}$  were chosen so that, in conjunction with the cost parameters of parameter set A (Table 1), the implied ratio  $\text{var}(H_t)/\text{var}(S_t)$  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 a trend. Thus, the reduced form used to generate the data was not literally (1.2) but

$$S_t = \phi_1 S_{t-1} + \phi_2 S_{t-2} + \text{constant} + \text{trend} + \varepsilon_{St} \quad (1.3a)$$

and

Table 2. Data-Generating Processes: Parameters of Exogenous Processes

$\phi_1$	$\phi_2$	$\text{var}(\varepsilon_S)$	$\text{var}(u)$	$\rho(\varepsilon_S, u)$
.75	.20	.120833	3.5	-.5

NOTE:  $\phi_1$  and  $\phi_2$  are the autoregressive parameters of the sales process defined in (1.2);  $\varepsilon_S$  is the sales shock defined in (1.2);  $u$  is the cost shock defined in (1.1).

$$H_t = (\lambda_1 + \lambda_2)H_{t-1} - \lambda_1\lambda_2H_{t-2} + \pi_1 S_{t-1} + \pi_2 S_{t-2} + \text{constant} + \text{trend} + \varepsilon_{Ht}. \quad (1.3b)$$

Coefficients on trend terms in (1.3a) and (1.3b) were chosen so that the implied coefficients of variation of  $\Delta S_t$  and  $\Delta H_t$  were each .2, a figure that approximately matches estimates for monthly nondurables, 1967–1990. Because different choices of cost parameters imply different autoregressive coefficients in (1.3b), the coefficient on the trend term in (1.3b) varies from DGP to DGP. Note that, although the relevant empirical work typically models trends as deterministic, this decision may not be innocuous. We do not know the extent to which our results are applicable to systems estimated in error correction (e.g., Kashyap and Wilcox 1993) or differenced form (e.g., West 1990); either of these transformations would likely produce regressors distinctly less autocorrelated than in our simulations, which in turn might have a notable effect on small-sample behavior.

A complete 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 DGP's, we generate data as follows. As noted previously, the vector of shocks  $(u_t, \varepsilon_{St})$  is assumed to be iid 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  $4 \times 1$  vector  $(H_0, H_{-1}, S_0, S_{-1})'$ . We then use (1.3) 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 and observations 103/104 or 303/304 for leads and discard the final 10,000 – 104 or 10,004 – 304 observations. These 9,700 additional observations were reserved for some additional experiments. One thousand samples were generated for each DGP. 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 United States. The sample size of 100 was chosen for comparison.

Table 3 displays the implied values of the parameters of the inventory equation (1.2b) for each of our DGP's. The

Table 3. Data-Generating Processes: Implied Coefficients of Inventory Equation

DGP	$\lambda_1 + \lambda_2$	$-\lambda_1\lambda_2$	$\pi_1$	$\pi_2$
A	1.22	-.42	.14	-.12
B	.24	-.14	.38	.05
C	1.07	-.22	.10	-.09
D	1.43	-.69	.33	-.15

NOTE:  $\lambda_1 + \lambda_2$ ,  $-\lambda_1\lambda_2$ ,  $\pi_1$ , and  $\pi_2$  are the coefficients of the reduced-form inventory equation (1.3b).

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 DGP's A and D, mild serial correlation for DGP C, and little serial correlation for DGP B.

## 2. 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.

### 2.1 Conventional Instrumental Variables

We make the first-order condition (1.1) estimable by replacing expected with realized values and moving all variables but  $H_t$  to the right side:

$$\begin{aligned} H_t &= \beta_1 X_{1t+2} + \beta_2 X_{2t+1} + \beta_3 S_{t+1} + v_{t+2} \\ &\equiv X_t' \beta + v_{t+2}, \\ v_{t+2} &\equiv u_t - \beta_1 (X_{1t+2} - E_t X_{1t+2}) \\ &\quad - \beta_2 (X_{2t+1} - E_t X_{2t+1}) - \beta_3 (S_{t+1} - E_t S_{t+1}); \\ X_t &\equiv (X_{1t+2}, X_{2t+1}, S_{t+1})'; \quad \beta \equiv (\beta_1, \beta_2, \beta_3)'. \end{aligned} \quad (2.1)$$

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,  $b = .995$ . Our conventional IV estimator calculates  $\beta$  linearly as follows. Let  $Z_t$  be a  $q \times 1$  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  $IV_q$  denote the estimator (2.3) defined later when there are  $q$  instruments:

$$\begin{aligned} IV4: \quad Z_t &\equiv (H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2})' \\ IV12: \quad Z_t &\equiv (H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}, H_{t-3}, S_{t-3}, \\ &\quad H_{t-4}, S_{t-4}, H_{t-5}, S_{t-5}, H_{t-6}, S_{t-6})'. \end{aligned} \quad (2.2)$$

[Note that the presence of the cost shock  $u_t$  invalidates the use of  $H_t$  and  $S_t$  as instruments; see (1.2).] See Section 2.2 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  $T \times q$  matrix whose  $t$ th row is  $Z_t'$ , and, similarly let  $X = [X_t']$  be the  $T \times 3$  matrix of right-side variables and  $Y = [H_t]$  the  $T \times 1$  vector of the left-side variable. Given  $Z_t$ , the IV estimator that has the smallest possible asymptotic variance-covariance matrix is

$$\hat{\beta} = (X' Z \hat{W} Z' X)^{-1} X' Z \hat{W} Z' Y, \quad (2.3)$$

where  $\hat{W}$  is a  $q \times q$  matrix that is an estimate of the inverse of the spectral density at frequency 0 of the  $q \times 1$  vector  $Z_t v_{t+2}$ —that is, the inverse of  $\sum_{j=-\infty}^{\infty} E Z_t Z_{t-j}' v_{t+2} v_{t+2-j}$ . Because the cost shock  $u_t$  is

iid,  $Z_t v_{t+2}$  is MA(2) and this infinite sum collapses to

$$\begin{aligned} W^{-1} &= \sum_{j=-2}^2 E Z_t Z_{t-j}' v_{t+2} v_{t+2-j} \\ &= \sum_{j=-2}^2 E Z_t Z_{t-j}' E v_{t+2} v_{t+2-j} \\ &\equiv \Gamma_0 + (\Gamma_1 + \Gamma_1') + (\Gamma_2 + \Gamma_2') \\ &\equiv \gamma_0 C_0 + \gamma_1 (C_1 + C_1') + \gamma_2 (C_2 + C_2'), \\ \gamma_j &\equiv E v_t v_{t-j}; \quad C_j \equiv E Z_t Z_{t-j}'. \end{aligned} \quad (2.4)$$

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

$$V = (E X_t Z_t' W E Z_t X_t')^{-1}, \quad (2.5)$$

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

$$\hat{\Gamma}_j = T^{-1} \sum_{t=j+1}^T Z_t Z_{t-j}' \hat{v}_{t+2} \hat{v}_{t+2-j} \quad (2.6)$$

for  $j \geq 0$ . Let  $m = \min(10, [\hat{\gamma} T^{1/3}])$ , where

$$\begin{aligned} \hat{\gamma} &= 1.1447 (\hat{s}^{(1)} / \hat{s}^{(0)})^{2/3}; \quad \hat{s}^{(1)} = 2\hat{\sigma}_1 + 4\hat{\sigma}_2 \\ \hat{s}^{(0)} &= \hat{\sigma}_0 + 2\hat{\sigma}_1 + 2\hat{\sigma}_2; \quad \hat{\sigma}_j = w' \hat{\Gamma}_j w, \quad w = (1, 1, 1, 1)'. \end{aligned}$$

We set

$$\hat{W} = \left\{ \hat{\Gamma}_0 + \sum_{j=1}^m [1 - j/(m+1)] (\hat{\Gamma}_j + \hat{\Gamma}_j') \right\}^{-1}. \quad (2.7)$$

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

### 2.2 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 (GLS) [here, we interpret ordinary least squares (OLS) and GLS as IV estimators in which the instruments are the right-side variables or transformations of those variables]. Hayashi and Sims (1983) pointed out that, although the usual GLS estimator is inconsistent in models with MA 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

Table 4. Parameters of the MA(2) Disturbance

DGP	$\theta_1$	$\theta_2$	Modulus of larger root
A	1.27	-.45	.67
B	.50	-.19	.43
C	.93	-.18	.67
D	1.44	-.71	.85

NOTE:  $\theta_1$  and  $\theta_2$  are the parameters of the MA(2) disturbance  $v_{t+2}$ ; see (2.1) and (2.9). The modulus presented is that of the larger of the two roots to  $z^2 - \theta_1 z - \theta_2 = 0$ .

IV 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, 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  $S_t$  are used as instruments in estimation of (2.3). 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 homoscedastic MA process, such as ours, Hansen (1985) provided 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  $IV_\infty$ .

Let  $R_t^*$  be the  $(4 \times 1)$  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 vector autoregression (VAR) (1.2) as a first-order VAR in  $R_t^*$ ,

$$R_t^* = F^* R_{t-1}^* + \varepsilon_t, \quad \varepsilon_t = (\varepsilon_{H_{t-1}}, \varepsilon_{S_{t-1}}, 0, 0)', \quad (2.8)$$

where, for example,  $F^*(1, 1) = \lambda_1 + \lambda_2$  and  $F^*(2, 2) = \phi_1$ . Write the MA representation of Equation (2.1)'s disturbance  $v_{t+2}$  as

$$v_{t+2} = \eta_{t+2} - \theta_1 \eta_{t+1} - \theta_2 \eta_t, \quad \eta_t \equiv v_t - E(v_t | v_{t-1}, v_{t-2}, \dots). \quad (2.9)$$

Let  $P^*$  be the  $3 \times 4$  matrix of coefficients of the projection of  $X_t$  on  $R_t^*$ ,  $E(X_t | R_t^*) = EX_t R_t^{*'} (ER_t^* R_t^{*'})^{-1} R_t^* = P^* R_t^*$ . In our particular case, application of the general formula supplied by Hansen (1985) indicates that an optimal set of instruments  $Z_t^*$  satisfies

$$Z_t^* = \theta_1 Z_{t-1}^* + \theta_2 Z_{t-2}^* + P^* (I_4 - \theta_1 F^* - \theta_2 F^{*2})^{-1} R_t^*. \quad (2.10)$$

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

$$(EZ_t^* X_t')^{-1} (W^*)^{-1} (EX_t Z_t^{*'})^{-1}, \quad (W^*)^{-1} = \gamma_0 G_0 + \gamma_1 (G_1 + G_1') + \gamma_2 (G_2 + G_2'), \quad \gamma_j = E v_t v_{t-j}; \quad G_j \equiv EZ_t^* Z_{t-j}^{*'} \quad (2.11)$$

Asymptotic standard errors may be computed from (2.10) and (2.11) in straightforward but tedious fashion.

Observe from (2.10) that, if  $\theta_1$  and  $\theta_2$  were 0, as would be the case if  $v_{t+2}$  were iid, the optimal instrument list would be the usual two-stage least squares one,  $Z_t^* = EX_t R_t^{*'} (ER_t^* R_t^{*'})^{-1} R_t^*$ , and there would be no point in using as instruments any variables other than those in the reduced form (1.2). It follows that, if  $\theta_1$  and  $\theta_2$  are close to 0, efficiency gains from using instruments other than those in the reduced form will be small, but if  $\theta_1$  and  $\theta_2$  are large, in the sense that one or both of the roots of  $z^2 - \theta_1 z - \theta_2$  are near unity in modulus, such efficiency gains potentially will be large. For each of our four DGP's, Table 4 presents  $\theta_1$  and  $\theta_2$ , along with the modulus of the larger root of  $z^2 - \theta_1 z - \theta_2$ . 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 5 presents the ratio of the standard errors of (1) IV $q$  for various  $q$ 's to (2) IV\* for each of our four DGP's. 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% of those of IV\*. On the other hand, for DGP's A, C, and D, there are substantial gains to using instruments beyond the 4 in the reduced form. Table 5 persuaded us to include IV12 in our simulation analysis: In DGP's 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 many parameters estimated in the first-stage regression.

[Note that 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  $\beta_1$ ,

Table 5. Asymptotic Standard Errors, IV $q$  Relative to IV\*

DGP	Estimator	Parameter		
		$\beta_1$	$\beta_2$	$\beta_3$
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

NOTE: IV $q$  is the conventional IV estimator described in (2.2) and (2.3), where  $Z_t$  consists of  $q$  instruments ( $q = 4, 6, 8, \text{ or } 12$ ); IV\* is the optimal estimator described in (2.12) and (2.13). The table presents the ratio of the square roots of the diagonal elements of (a) the variance-covariance matrix of IV $q$  [computed according to (2.5)], to (b) the variance-covariance matrix of IV\* [computed according to (2.11)].

Table 6. Distributions of Standardized Parameter Estimates, From Simulations,  $T = 100$

DGP estimator	$\hat{\beta}_1 - \beta_1$			$\hat{\beta}_2 - \beta_2$			$\hat{\beta}_3 - \beta_3$		
	50% CI	Median	Trimmed MSE	50% CI	Median	Trimmed MSE	50% CI	Median	Trimmed MSE
<b>A</b>									
IV4	(-.6, .8)	.07	1.14	(-.8, .6)	-.10	1.20	(-3.1, 1.6)	-.68	2.51
IV12	(-.3, .9)	.29	.92	(-.9, .3)	-.30	.98	(-4.3, .8)	-1.26	2.52
IV*	(-.4, .4)	.04	.61	(-.5, .4)	-.06	.63	(-2.7, 1.6)	-.39	2.27
asy*	(-.3, .3)	.00	.21	(-.3, .3)	.00	.20	(-.5, .5)	.00	.51
<b>B</b>									
IV4	(-.5, 1.0)	.27	1.13	(-1.1, .4)	-.38	1.24	(-.5, 1.2)	.23	1.47
IV12	(-.2, 1.0)	.47	.94	(-1.4, -.0)	-.63	1.34	(-.2, 1.7)	.67	1.79
IV*	(-.5, .7)	.14	.93	(-.9, .4)	-.22	1.09	(-.5, 1.1)	.24	1.33
asy*	(-.6, .6)	.00	.79	(-.6, .6)	.00	.82	(-.7, .7)	.00	.97
<b>C</b>									
IV4	(-.7, .8)	.11	1.14	(-.8, .6)	-.13	1.17	(-1.1, 2.8)	.66	2.41
IV12	(.0, 1.0)	.54	.98	(-1.1, -.1)	-.59	1.06	(-2.3, 1.0)	-.59	2.18
IV*	(-.8, .5)	-.06	.91	(-.5, .8)	.03	.91	(-.4, 3.9)	1.19	2.06
asy*	(-.5, .5)	.00	.45	(-.4, .4)	.00	.44	(-.5, .5)	.00	.59
<b>D</b>									
IV4	(-.6, .9)	.11	1.23	(-.9, .6)	-.11	1.28	(-3.1, .5)	-1.02	2.30
IV12	(-.4, .8)	.15	1.05	(-.8, .5)	-.15	1.06	(-3.6, .3)	-1.39	2.38
IV*	(-.3, .4)	.05	.53	(-.5, .3)	-.07	.55	(-2.6, .5)	-.84	2.11
asy*	(-.2, .2)	.00	.11	(-.2, .2)	.00	.11	(-.5, .5)	.00	.58
asy4	(-.7, .7)	.00	1.00	(-.7, .7)	.00	1.00	(-.7, .7)	.00	1.00

NOTE: The estimating equations are IV4 (2.3), IV12 (2.3), IV\* (2.13). The difference between estimated and population parameter is standardized by dividing by asymptotic standard error for IV4. The "50% CI" is a 50% confidence interval constructed using the 250th and 750th largest of the 1,000 estimates; "Median" is the 500th 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. "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 5 and the text). "asy4" presents the asymptotic values for IV4.

for example, the ratios of the asymptotic standard errors of FIML to IV\* are A: .90; B: .84; C: .69; D: .90. The ratios for the other parameters are comparable. For our DGP's, then, the efficiency gains associated with FIML relative to IV\* are modest, a result consistent with West (1986b).]

We operationalize (2.10) for a given artificial sample of size  $T$  as follows, leaving for future research evaluation of other ways of making IV\* feasible. (a) We estimate four 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-side variables in the two equations:  $H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}; H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}; H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}, H_{t-3}, S_{t-3}; H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}, H_{t-3}, S_{t-3}, H_{t-4}, S_{t-4}$ . 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 DGP.) (b) We then obtain  $\hat{P}$  from an OLS regression of  $X_t$  on  $R_t$  and  $\hat{\theta}_1$  and  $\hat{\theta}_2$  by maximum likelihood applied to residuals obtained from IV4. (c) Next, we use (2.10) to construct an estimate of the unconditional variance-covariance matrix of  $(\hat{Z}_t^*, \hat{Z}_{t-1}^*)'$  and then draw  $(\hat{Z}_0^*, \hat{Z}_{-1}^*)'$  from its unconditional normal distribution. (d) Then we use (2.10) to generate  $\hat{Z}_t^*$  recursively forward from  $t = 1$  to  $t = T$ ,

$$\hat{Z}_t^* = \hat{\theta}_1 \hat{Z}_{t-1}^* + \hat{\theta}_2 \hat{Z}_{t-2}^* + \hat{P}(I - \hat{\theta}_1 \hat{F} - \hat{\theta}_2 \hat{F}^2)^{-1} R_t. \quad (2.12)$$

(e) Finally, we estimate  $\beta$  as

$$\hat{\beta} = \left( \sum_{t=1}^T \hat{Z}_t^* X_t' \right)^{-1} \sum_{t=1}^T \hat{Z}_t^* H_t. \quad (2.13)$$

### 2.3 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 (2.5)] is constructed as

$$\hat{V} = \left[ \left( \sum_{t=1}^T X_t Z_t' / T \right) \hat{W} \left( \sum_{t=1}^T Z_t X_t' / T \right) \right]^{-1} \quad (2.14)$$

for  $\hat{W}$  defined in (2.7). 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

$$(\hat{\beta}_1 - \beta_1) / [\hat{V}(1, 1) / T]^{1/2}. \quad (2.15)$$

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

$$T^{-1} \left( \sum_{t=1}^T \hat{v}_{t+2} Z_t' \right) \hat{W} \left( \sum_{t=1}^T Z_t \hat{v}_{t+2} \right) \overset{A}{\sim} \chi^2(q - 3), \quad (2.16)$$

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

Table 7. Distributions of Standardized Parameter Estimates, From Simulations,  $T = 300$

DGP estimator	$\hat{\beta}_1 - \beta_1$			$\hat{\beta}_2 - \beta_2$			$\hat{\beta}_3 - \beta_3$		
	50% CI	Median	Trimmed MSE	50% CI	Median	Trimmed MSE	50% CI	Median	Trimmed MSE
<b>A</b>									
IV4	(-.7, .8)	.14	1.14	(-.8, .6)	-.15	1.14	(-1.2, .9)	-.17	1.76
IV12	(-.2, .7)	.29	.57	(-.7, .2)	-.29	.57	(-1.3, .6)	-.32	1.64
IV*	(-.3, .4)	.08	.36	(-.4, .3)	-.09	.35	(-.9, .6)	-.14	1.38
asy*	(-.3, .3)	.00	.21	(-.3, .3)	.00	.20	(-.5, .5)	.00	.51
<b>B</b>									
IV4	(-.6, .9)	.21	1.10	(-.9, .5)	-.30	1.16	(-.6, .9)	.14	1.21
IV12	(-.3, 1.0)	.41	.98	(-1.1, .2)	-.49	1.15	(-.4, 1.2)	.37	1.38
IV*	(-.5, .7)	.14	.84	(-.8, .4)	-.22	.92	(-.5, .9)	.18	1.11
asy*	(-.6, .6)	.00	.79	(-.6, .6)	.00	.82	(-.7, .7)	.00	.97
<b>C</b>									
IV4	(-.7, .8)	.10	1.13	(-.8, .7)	-.11	1.14	(-.6, 1.5)	.36	1.66
IV12	(-.2, .9)	.44	.78	(-.9, .1)	-.45	.78	(-.8, .9)	-.07	1.33
IV*	(-.7, .4)	-.06	.70	(-.4, .7)	.04	.70	(-.1, 1.7)	.58	1.39
asy*	(-.5, .5)	.00	.45	(-.4, .4)	.00	.44	(-.5, .5)	.00	.59
<b>D</b>									
IV4	(-.6, .8)	.15	1.14	(-.8, .6)	-.15	1.16	(-1.3, .4)	-.34	1.53
IV12	(-.3, .6)	.16	.48	(-.6, .3)	-.15	.49	(-1.3, .4)	-.43	1.56
IV*	(-.2, .3)	.06	.25	(-.3, .2)	-.06	.26	(-1.1, .4)	-.29	1.28
asy*	(-.2, .2)	.00	.11	(-.2, .2)	.00	.11	(-.5, .5)	.00	.58
asy4	(-.7, .7)	.00	1.00	(-.7, .7)	.00	1.00	(-.7, .7)	.00	1.00

NOTE: See note to Table 6.

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

$$\hat{V}^* = \left( \sum_{t=1}^T \hat{Z}_t^* X_t' / T \right)^{-1} (\hat{W}^*)^{-1} \left( \sum_{t=1}^T X_t \hat{Z}_t^{*'} / T \right)^{-1} \tag{2.17}$$

In (2.17), we initially computed  $\hat{W}^*$  in a fashion analogous to (2.7). But the resulting test statistics sometimes were very poorly sized. So for  $\hat{\beta}, \hat{\theta}_1,$  and  $\hat{\theta}_2$  defined in (2.12) and (2.13), we instead followed West (1994) and estimated  $\hat{W}^*$  as

$$\begin{aligned} \hat{W}^* &= \left( T^{-1} \sum \hat{d}_t \hat{d}_t' \right)^{-1} \\ \hat{d}_t &\equiv \hat{\eta}_{t+2} (\hat{Z}_t^* - \hat{\theta}_1 \hat{Z}_{t+1}^* - \hat{\theta}_2 \hat{Z}_{t+2}^*) \\ \hat{v}_{t+2} &\equiv \hat{\eta}_{t+2} - \hat{\theta}_1 \hat{\eta}_{t+1} - \hat{\theta}_2 \hat{\eta}_t, \\ \hat{v}_{t+2} &\equiv H_t - X_t' \hat{\beta}; \quad \hat{\eta}_0 = \hat{\eta}_{-1} = 0. \end{aligned} \tag{2.18}$$

$\hat{W}^*$  is positive semidefinite by construction. It may be shown that  $\hat{W}^*$  so defined is consistent. In estimating  $\hat{W}$ , it might be of interest to apply a computation like (2.18) [or, as a referee has pointed out, to impose conditional homoscedasticity, or to iterate once more so that IV $q$  residuals are used in estimation of  $\beta$  (see Kocherlakota 1990)]. To keep the scope of the study manageable, however, we did not, and we limited our analysis of the conventional IV estimator to the version that seems to us to be most widely used.

### 3. SIMULATION EVIDENCE ON DISTRIBUTION OF PARAMETER ESTIMATES

Tables 6 and 7 present some Monte Carlo results on the distribution of the parameter estimates, Table 6 for a sample size  $T = 100$ , Table 7 for  $T = 300$ . They are organized by DGP. For each DGP, the tables present results for three estimators—IV4 (2.2), IV12 (2.3), and IV\* [(2.12), (2.13)]. The “asy\*” rows in each panel give asymptotic quantities for IV\*, and the “asy4” row at the bottom of the table does the same for IV4; in light of Table 5, 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 the 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\* and  $N(0, (\sigma_{12}/\sigma_4)^2)$  for IV12, where  $\sigma^*/\sigma_4$  and  $\sigma_{12}/\sigma_4$  are computed from the relevant rows of Table 5. For example, in DGP A, the asymptotic theory implies that standardizing the IV\* estimate of  $\beta_1$  in this fashion produces an  $N(0, .45^2)$  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 labeled “50% CI” gives a 50% confidence interval constructed by dropping the largest 250 and smallest 250 of the 1,000 standardized parameter estimates, or, for “asymptotic,” the values appropriate for an  $N(0, 1)$  variable. The difference between the upper and lower bounds of these confidence intervals is the interquartile range. “Median” gives the median of the 1,000 estimates.

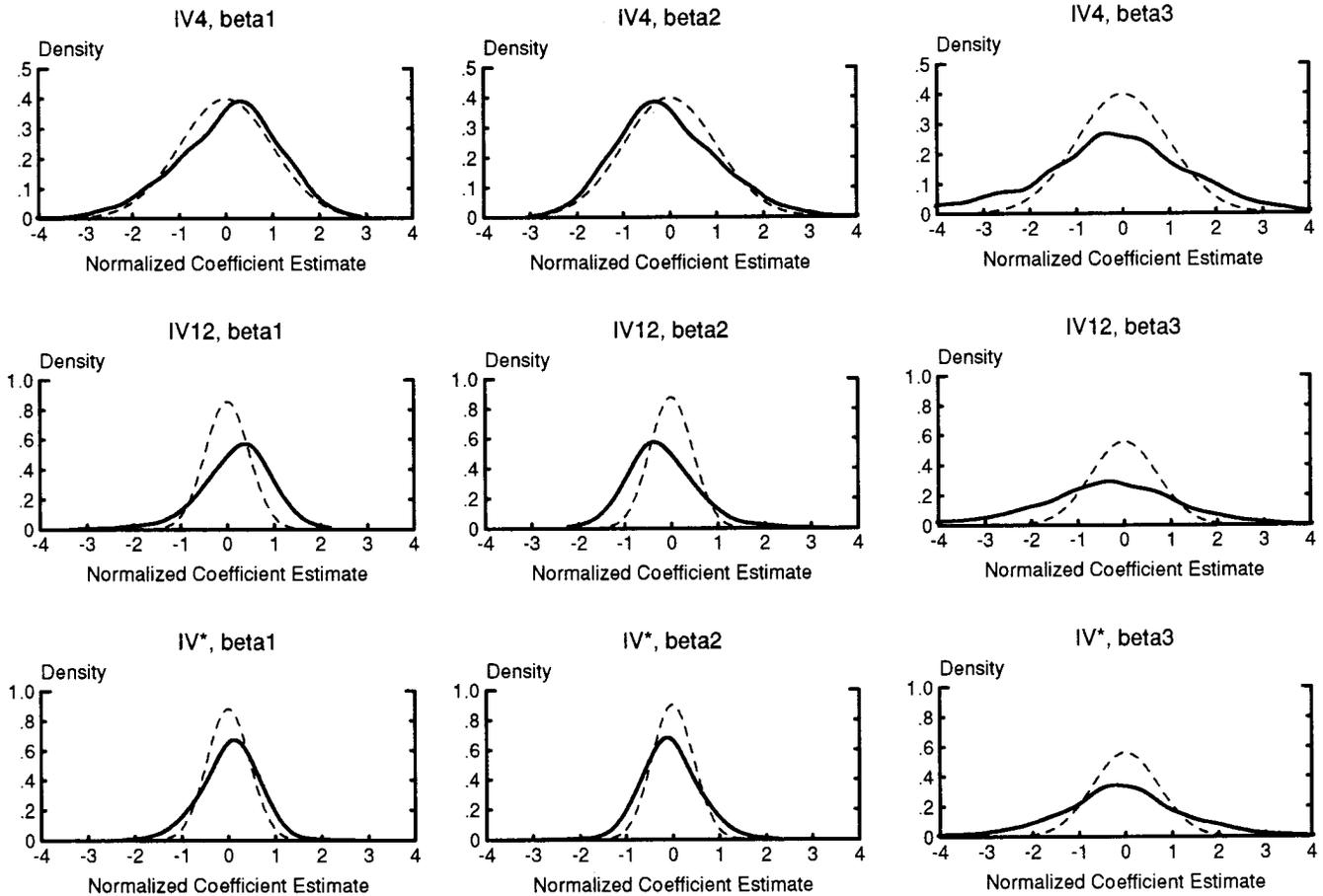


Figure 1. Estimated and Theoretical Densities of Parameter Estimates, DGP A,  $T = 300$ . The figure presents densities for three different estimators applied to data drawn from a single DGP with sample size equal to 300. The three panels in the first row describe the density of the parameter estimates from the conventional estimator that uses two lags each of inventories and sales as instruments (four stochastic variables total). Similarly, the three panels in the second row describe the density of the conventional estimator that uses six lags each of inventories and sales as instruments, and the panels in the third row describe the density of the estimator that is optimal in the class of all estimators that use linear combinations of lags of inventories and sales as instruments. Within each panel, the solid line describes the estimated small-sample density of the parameter estimate, and the dashed line describes the theoretical normal density suggested by standard asymptotic theory.

“Trimmed MSE” gives a mean squared error computed by (a) dropping all entries greater than 3.0 in absolute value, (b) calculating the average squared value of the remaining observations, and (c) dividing by .9735, which is the variance of an  $N(0, 1)$  variable doubly truncated at  $-3$  and  $+3$  (Johnson and Kotz 1970, p. 83). We trimmed before computing the MSE because the simultaneous-equations literature indicates that second moments of our estimator may not exist because our equation has only one more instrument than right-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 1994), we found little sensitivity to the exact point of truncation.

In conjunction with Table 5, we read Tables 6 and 7 as follows. First, as measured by either interquartile range (width of the 50% CI's) or trimmed MSE, the asymptotic theory underpredicts the variability of all three estimators. The discrepancies between asymptotics and simulation are larger for  $T = 100$  than  $T = 300$  (no surprise) and larger for  $\beta_3$  than for  $\beta_1$  or  $\beta_2$  (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 approx-

imate 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 6), 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. Although 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 its asymptotic standard deviation rather than the IV4 standard deviation used in Tables 6 and 7, all 12 estimates have a median value of  $\hat{\beta} - \beta$  that is more than .4 asymptotic standard deviations, and 10 are greater than .5. For all three estimators, asymmetry in the 50% CI's is also evident in Table 6. On the other hand, Table 7 indicates that, although some problems remain, par-

Table 8. Size of Nominal .05 T Tests, From Simulations

DGP	Estimator	T = 100			T = 300		
		$\beta_1$	$\beta_2$	$\beta_3$	$\beta_1$	$\beta_2$	$\beta_3$
A	IV4	.061	.056	.004	.063	.060	.001
	IV12	.142	.134	.020	.080	.067	.004
	IV*	.080	.077	.068	.076	.073	.087
B	IV4	.073	.072	.065	.065	.053	.061
	IV12	.193	.260	.231	.103	.104	.106
	IV*	.115	.073	.059	.056	.052	.066
C	IV4	.073	.071	.010	.075	.076	.011
	IV12	.264	.264	.031	.158	.154	.014
	IV*	.115	.119	.053	.080	.082	.050
D	IV4	.020	.018	.002	.037	.028	.002
	IV12	.047	.044	.026	.017	.015	.006
	IV*	.083	.086	.091	.118	.117	.114

NOTE: In each of 1,000 simulations, we computed *t* statistics testing whether each of the three  $\beta_i$ s equals its Table 1 population value. This table 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.

ticularly 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  $.27 \approx 1.06(1,000)^{-1/5} = 1.06(\text{sample size})^{-1/5}$  (e.g., see Silverman 1986). Note that, although the horizontal scales are the same on all nine 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 dispersed than is IV4. The IV4 and IV\* simulation densities

are noticeably better centered than are those of IV12. An appendix, available on request, has comparable figures for the other DGP's and for  $T = 100$ . These tell a qualitatively similar story. So too do estimates that set the bandwidth at  $(.75) \times (\text{bandwidth in Fig. 1})$  and  $(1.25) \times (\text{bandwidth in Fig. 1})$ .

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  $\beta_3$  for DGP's 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.

Tables 8 and 9 present information on the size of test statistics. Table 8 presents the size of nominal .05 tests of the hypothesis  $H_0: \beta_i = \text{population value}, i = 1, 2, 3$ , 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 1,000 simulations for which the computed statistic was greater than 3.84 [the .05 critical value for a  $\chi^2(1)$  random variable]. The asymptotic standard error on a given fraction is  $[(.05)(.95)/1,000]^{1/2} \approx .007$ .

For IV4 and IV\*, tests for  $\beta_1$  and  $\beta_2$  typically were well behaved for both  $T = 100$  and  $T = 300$  [at least by the standards of recent work such as that of Newey and West (1994)!]; actual sizes ranged from about .02 to about .12. For IV12, test statistics of  $\beta_1$  and  $\beta_2$  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 6 and 7 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  $\beta_3$  than  $\beta_1$  or  $\beta_2$ . Tables 8 and 9 do indeed show that tests on  $\beta_3$  were generally more problematic than those on  $\beta_1$  and  $\beta_2$ , but in

Table 9. Size of Nominal .05 J Tests, From Simulations

DGP	Estimator	T = 100	T = 300
		J size	J size
A	IV4	.041	.056
	IV12	.001	.001
B	IV4	.052	.061
	IV12	.004	.023
C	IV4	.039	.051
	IV12	.003	.001
D	IV4	.042	.055
	IV12	.001	.000

NOTE: In each of 1,000 simulations, tests of instrument-residual orthogonality were computed as in (2.16). This table 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.

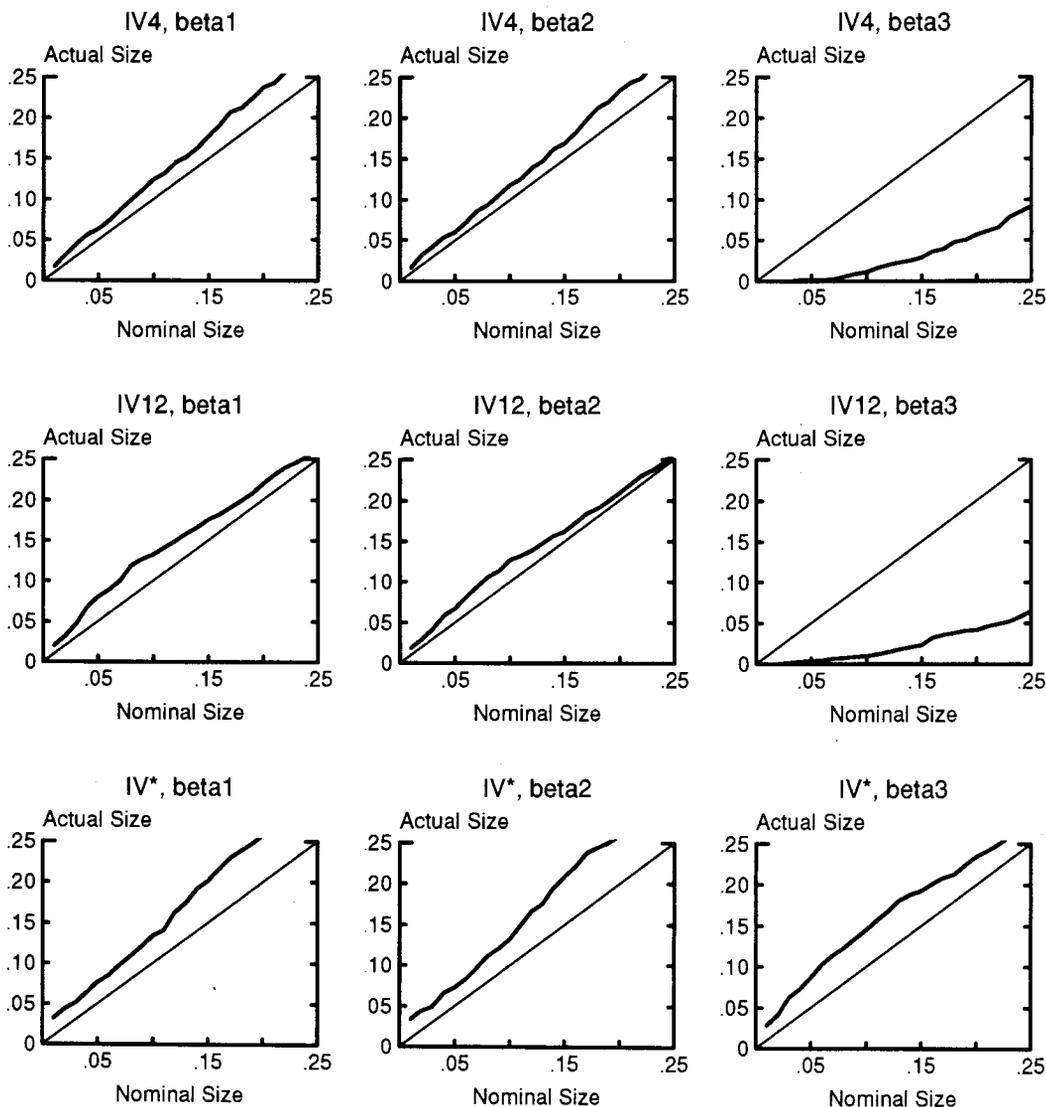


Figure 2. Actual and Nominal Sizes of  $t$  Tests, DGP A,  $T = 300$ . The figure compares the actual (plotted on the  $y$  axis) and nominal (plotted on the  $x$  axis) sizes of  $t$  tests, for the same three estimators and the same data-generating process as were used in Figure 1. If the tests were correctly sized, all curves would lie on the 45-degree line. In cases in which the curve is above the 45-degree line, the  $t$  test rejects the null hypothesis too frequently, and conversely for cases in which the curve is below the 45-degree line.

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 6 and 7), the relevant entries of the variance-covariance matrices are even more spread out, the result being egregious underrejection rather than egregious overrejection. IV\* suffers from no such problem. We conjecture that this is due more to the way the covariance matrix was estimated (see Sec. 2.4) than to something inherent in the way the parameters were estimated: We repeated the calculation of Table 8 for  $T = 300$  for IV\*, calculating Equation (2.17)'s  $\hat{W}^*$  in a fashion analogous to that described in Equations (2.6) and (2.7). The size of nominal .05 tests of Table 8 on  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  were DGP A: .001, .000, .001; DGP B: .040, .037, .044; DGP C: .023, .022, .003; DGP D: .000, .000, .001. The evidence of West (1994) suggests that use of the estimator (2.18) might result in a similar improvement of the sizes of the test statistics for IV4 and IV12.

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  $\beta_1$  and  $\beta_2$  (the first two columns of the figure), as does IV\* for  $\beta_3$  (last column, last row). IV4 and IV12 reject much too infrequently for  $\beta_3$  (last column, first two rows). The figure shows that this applies not only to nominal .05 tests (the focus of Tables 8 and 9) but to tests of nominal sizes ranging from .01 to .25. (A referee has noted that such statements should be interpreted with caution because we do not provide a  $p$  value for statements concerning the joint behavior of tests over a range of nominal sizes.) Analogous plots for other DGP's and for  $T = 100$  are available in the additional appendix and are also consistent with Table 8.

IV4 is sufficiently simple computationally that we repeated our simulation exercise for DGP A with samples of size 10,000 [relaxing the constraint on the maximum value of the bandwidth  $m$  [defined following (2.6)] in a fashion

Table 10. Estimates of Aggregate Nondurables in Manufacturing, 1967-1992

(1) Estimator	(2) $\hat{\beta}_1$	(3) $\hat{\beta}_2$	(4) $\hat{\beta}_3$	(5) $\hat{\theta}_1$	(6) $\hat{\theta}_2$	(7) Modulus of larger root
IV4	.114 (.044)	.160 (.134)	.004 (.008)	.84	-.35	.42
IV12	.155 (.016)	.036 (.048)	-.044 (.004)			
IV*	.145 (.024)	.068 (.071)	.001 (.006)			

NOTE: The table presents estimates of IV4, IV12, and IV\*, computed according to (2.3) and (2.13). The vector  $\hat{R}_t$  [defined above (2.12)] was the set of lags that maximized the Schwarz criterion, where the following four sets were considered:  $H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}; H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}; H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}; H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}, H_{t-3}, S_{t-3}; H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}, H_{t-3}, S_{t-3}, H_{t-4}, S_{t-4}$ . All four also included intercept and trend. Columns (5)-(7) are as described in Table 4 and are estimated from the two-stage least squares residuals.

that ensured consistency}. Even here there was evidence of missizing for one hypothesis test: The nominal .05  $t$  tests on  $\beta_3$  had an actual size of .007. (The comparable figures for  $\beta_1$  and  $\beta_2$  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.

Table 9 indicates that  $J$  tests are approximately correctly sized for IV4, poorly sized for IV12.

#### 4. 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 1, 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 10 has the IV4 estimates. The estimates of the MA parameter yield an implied larger root of .42, about the same size as that for DGP B (see Table 4). Accordingly, dramatic efficiency gains in going to either IV12 or IV\* are not to be expected. Lines 2 and 3 bear out this expectation. Although the  $t$  statistic on  $\beta_1$  becomes noticeably larger, that on  $\beta_2$  falls and that on  $\beta_3$  falls for IV\*, rises for IV12. On transforming the estimates of the  $\beta_i$ 's to the underlying economic parameters, 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. See the Appendix for details and discussion.

#### 5. CONCLUSIONS

This article has compared several estimators of a dynamic linear model. For all of 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 sug-

gested by Hansen (1985), which for three of our four DGP's yielded substantial asymptotic and finite-sample benefits relative to conventional IV estimators.

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

#### ACKNOWLEDGMENTS

We thank Wouter den Haan, an anonymous referee, and participants at the February 1994 National Bureau of Economic Research 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 ours and not necessarily those of the Board of Governors of the Federal Reserve System or of other members of its staff.

#### APPENDIX: SPECIFICATION AND SOLUTION OF THE MODEL

The model and DGP's were also used by West and Wilcox (1994), so some of the prose in this Appendix and in Section 1 and some of the entries in the tables are also found in that article.

The model underlying (1.1) follows Holt, Modigliani, Muth, and Simon (1960). A representative firm maximizes the expected present discounted value of future cash flows, with a cost function that includes linear and quadratic costs of production and of changing production and holding inventories. As in (1.1), let  $S_t$  be real sales,  $Q_t$  real production,  $H_t$  real end-of-period inventories,  $b$  a discount factor,  $0 \leq b < 1$ ,  $E_t$  mathematical expectations conditional on information known at time  $t$ , assumed equivalent to linear projections; also let  $C_t$  be real costs,  $p_t$  real price, and  $u_t$  a cost shock that is observable to the firm but unobservable to the econometrician. The objective function is

$$\max \lim_{T \rightarrow \infty} E_t \sum_{j=0}^T b^j (p_{t+j} S_{t+j} - C_{t+j}),$$

where

$$Q_{t+j} = S_{t+j} + H_{t+j} - H_{t+j-1}$$

and

$$C_{t+j} = .5a_0\Delta Q_{t+j}^2 + .5a_1Q_{t+j}^2 + .5a_2(H_{t+j-1} - a_3s_{t+j})^2 + H_{t+j}u_{t+j} + \text{linear terms} + (\text{linear} \times \text{trend}) \text{ terms.}$$

The  $a_i$ 's are the parameters of interests. Omission of shocks that shift the marginal cost of production or of changing production (i.e., terms of the form shock  $\times Q_{t+j}$  or shock  $\times \Delta Q_{t+j}$ ) is for notational economy and without economic substance. An optimizing firm will not be able to cut costs by increasing production by one unit this period, storing the unit in inventory, and producing one less unit next period, holding revenue unchanged throughout. Formally, differentiating (1.1) with respect to  $H_t$  gives

$$E_t\{a_0(\Delta Q_t - 2b\Delta Q_{t+1} + b^2\Delta Q_{t+2}) + a_1(Q_t - bQ_{t+1}) + ba_2(H_t - a_3S_{t+1}) + \text{deterministic terms} + u_t\} = 0,$$

where the deterministic terms result from the linear and (linear  $\times$  trend) terms in the cost function.

Let  $c \equiv a_0(1 + 4b + b^2) + a_1(1 + b) + ba_2$ . Then (1.1) follows from the preceding equation with  $\beta_1 = a_0/c$ ,  $\beta_2 = a_1/c$ , and  $\beta_3 = ba_2a_3/c$ . The values of the  $a_i$ 's in each DGP are as follows (see the text for a reference to an empirical paper that suggests such values): A:  $a_0 = 1., a_1 = .1, a_2 = .1, a_3 = .1$ ; B:  $a_0 = 1., a_1 = -2.0, a_2 = 6., a_3 = 1.0$ ; C:  $a_0 = 1., a_1 = .1, a_2 = 2., a_3 = .1$ ; D:  $a_0 = 1., a_1 = -.5, a_2 = .1, a_3 = .5$ .

The reduced-form parameters in (1.2b) relate to the underlying cost parameters as follows: Let  $\lambda_1$  and  $\lambda_2$  be 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  $\rho_1, \rho_2, w_1, w_2, w_3$ , and  $w_4$ , the  $(1 \times 2)$  vector  $e'$ , and the  $(2 \times 2)$  matrices  $\Phi$  and  $D$  as

$$\rho_1 = \lambda_1 + \lambda_2, \quad \rho_2 = -\lambda_1\lambda_2, \quad w_1 = b^2\rho_2,$$

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

$$w_3 = \rho_2[2b + 1 + (a_1/a_0)], \quad w_4 = -\rho_2,$$

$$e' = (1 \quad 0), \quad \Phi = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix},$$

$$D = [I - b\rho_1\Phi - b\rho_2\Phi^2]^{-1}.$$

Then

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

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

With regard to the empirical results discussed in Section 4, for each estimator, at least one of  $a_0, a_1$ , or  $a_2$  must be positive by construction. The parameter estimates that were positive for both IV\* and IV4 were the costs of production  $a_1$  and of changing production  $a_0$ . The inventory holding cost estimate  $a_2$  was positive for IV\* as well. The estimate of the parameter  $a_3$  that determines the target inventory-sales ratio was negative. Because the simulations found it particularly difficult to get a reliable estimate of  $\beta_3$ , it may be noteworthy that the two parameter estimates that were positive may be inferred from the estimates of  $\beta_1$  and  $\beta_2$ , without use of the estimate of  $\beta_3$ , but the estimates that were negative relied in part on the estimate of  $\beta_3$ .

[Received April 1994. Revised March 1995.]

### REFERENCES

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

Blanchard, O. J., and Melino, A. (1986), "The Cyclical Behavior of Prices and Quantities: The Case of the Automobile Market," *Journal of Monetary Economics*, 17, 379-408.

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

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

Ferson, W. E., and Foerster, S. E. (1991), "Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models," unpublished manuscript, University of Chicago, Graduate School of Business.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

— (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, L. P., and Singleton, K. J. (1988), "Computing Semi-Parametric Efficiency Bounds for Linear Time Series Models," unpublished manuscript, University of Chicago, Dept. of Economics.

Hayashi, F., and Sims, C. A. (1983), "Nearly Efficient Estimation of Time Series Models With Predetermined, but not Exogenous, Instruments," *Econometrica*, 51, 783-798.

Holt, C. C., Modigliani, F., Muth, J., and Simon, H. (1960), *Planning, Production, and Inventories*, Englewood Cliffs, NJ: Prentice-Hall.

Johnson, N. L., and Kotz, S. (1970), *Continuous Univariate Distributions—1*, Boston: Houghton Mifflin.

Kashyap, A. K., and Wilcox, D. W. (1993), "Production and Inventory Control at the General Motors Corporation During the 1920s and 1930s," *American Economic Review*, 83, 383-401.

Kocherlakota, N. (1990), "On Tests of Representative Consumer Asset Pricing Models," *Journal of Monetary Economics*, 26, 285-304.

Krane, S. D., and Braun, S. N. (1991), "Production Smoothing Evidence From Physical Product Data," *Journal of Political Economy*, 99, 558-581.

Newey, W. K., and West, K. D. (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631-654.

Oliner, S., Rudebusch, G., and Sichel, D. (1992), "The Lucas Critique Revisited: Assessing the Stability of Empirical Euler Equations," *Journal of Econometrics*, 70, 291-316.

- Phillips, P. C. B. (1983), "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics Volume 1*, eds. Z. Griliches and M. D. Intriligator, Amsterdam: North-Holland, pp. 449–516.
- Popper, H. (1992), "Formulating Nonlinear Moment Conditions for Generalized Method of Moments," unpublished manuscript, Santa Clara University, Dept. of Economics.
- Ramey, V. A. (1991), "Nonconvex Costs and the Behavior of Inventories," *Journal of Political Economy*, 99, 306–334.
- Rotemberg, J. (1984), "Interpreting the Statistical Failures of Some Rational Expectations Macroeconomic Models," *American Economic Review*, 74, 188–193.
- Silverman, B. W. (1986), *Density Estimation for Data Analysis*, London: Chapman and Hall.
- Tauchen, G. (1986), "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained From Financial Markets Data," *Journal of Business & Economic Statistics*, 4, 397–425.
- West, K. D. (1986a), "A Variance Bounds Test of the Linear Quadratic Inventory Model," *Journal of Political Economy*, 94, 374–401.
- (1986b), "Full Versus Limited Information Estimation of a Rational Expectations Model: Some Numerical Comparisons," *Journal of Econometrics*, 33, 367–386.
- (1990), "The Sources of Fluctuations in Aggregate Inventories and GNP," *Quarterly Journal of Economics*, CV 939–972.
- (1994), "Another Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," unpublished manuscript.
- (in press), "Inventory Models," in *Handbook of Applied Econometrics—Volume 1 (Macroeconometrics)*, eds. M. Pesaran and M. Wickens, Oxford: Basil Blackwell.
- West, K. D., and Wilcox, D. W. (1994), "Some Evidence on Finite Sample Distributions of Instrumental Variables Estimators of the Linear Quadratic Inventory Model," in *Inventory Cycles and Monetary Policy*, ed. R. Fiorito, Berlin: Springer-Verlag, pp. 253–282.