INTERNATIONAL ECONOMIC REVIEW
Vol. 42, No. 4, November 2001

ON OPTIMAL INSTRUMENTAL VARIABLES ESTIMATION OF STATIONARY TIME SERIES MODELS*

By Kenneth D. West[1]

University of Wisconsin, U.S.A.

In many time series models, an infinite number of moments can be used for estimation in a large sample. I supply a technically undemanding proof of a condition for optimal instrumental variables use of such moments in a parametric model. I also illustrate application of the condition in estimation of a linear model with a disturbance that is serially uncorrelated and conditionally heteroskedastic.

1. INTRODUCTION

In many instrumental variables (IV) applications, an infinite number of moments is available for use in large sample estimation. This is in particular the case with most time series models. If a given variable, say z_t, is a legitimate instrument, so, too, are its lags $z_{t-1}, z_{t-2}, \dots$.

Several theoretical papers on optimal IV estimation have allowed for an infinite number of moments, including Hayashi and Sims (1983), Hansen (1985a), Hansen, Heaton and Ogaki (1988), Heaton and Ogaki (1991), and Bates and White (1990, 1993). Some other papers have shown that for some empirically realistic data generating processes, use of the entire set of available moments can result in large asymptotic efficiency gains relative to use of a small set of moments (Hansen and Singleton, 1991; West and Wilcox, 1996). Simulations in West and Wilcox (1996) showed that these gains are often realized in samples of size typically available.

Nevertheless, time series applications seem to rarely attempt to use anything but a small set of moments. If my experience is any guide, one reason is that much of the theoretical work is technically demanding. How one would construct an optimal estimator in a particular application is probably not obvious to the typical researcher.

In the next section of this article, I supply a technically undemanding proof of a condition characterizing an optimal IV estimator. This condition was first proved in Hansen (1985a). While the methods and assumptions I use do require some econometric sophistication, they will, I hope, be sufficiently familiar that many readers will find it straightforward to interpret my results. Indeed, in the final section of this article I do so by constructing the optimal IV estimator of the parameters of a

* Manuscript received April 1998; revised February 2000.

[1] This is a renamed and substantially revised version of West (1994). I thank Stanislav Anatolyev, Frank Diebold, Yuichi Kitamura, Michael McCracken, various students, and two anonymous referees for helpful comments and questions, and the National Science Foundation and the Graduate School of the University of Wisconsin for financial support. E-mail to: kdwest@facstaff.

univariate linear model with a serially uncorrelated and conditionally heteroskedastic disturbance, when the space of instruments is defined to be nonstochastic linear combinations of current and lagged values of a given variable.

To prevent confusion, let me note two limitations. First, in contrast to much of the theoretical literature, I do not consider what determines whether a given space of estimators includes one that satisfies the optimality condition. This simplifies the analysis. Second, I do not describe how to make an optimal estimator feasible. Rather, my aim is to show how to solve for what might be called a "population" optimal estimator, one that relies on moments or parameters that will be unknown in typical applications. In application one will have to estimate these moments and parameters. See West and Wilcox (1996) for an application that uses a parametric estimation technique; see Kuersteiner (1996) for one that uses a nonparametric technique.

2. AN OPTIMALITY CONDITION

The aim is to efficiently estimate the $(k \times 1)$ unknown parameter vector β_0 in the regression model

$$y_t = f_t(\beta_0) + u_t. \tag{2.1}$$

Here, y_t and u_t are each $l \times 1$, y_t is observed, and u_t is unobserved. The known function f_t: $\mathbf{R}^k \to \mathbf{R}^l$ is continuously differentiable with probability one in an open neighborhood around β_0, with a $k \times l$ matrix of derivatives

$$F_t = F_t(\beta_0) \equiv \frac{\partial f_t}{\partial \beta}(\beta_0). \tag{2.2}$$

Possible dependence of F_t on β_0 is suppressed for notational simplicity. The t subscripts on f and F indicate that they depend not only on β_0 but also on observable data, for example, $f_t(\beta_0) = X_t'\beta_0$, $F_t = X_t$ in a linear model.

A set of instruments, each of dimension $(k \times l)$, is available for estimation. (As illustrated below, the assumption that the row dimension of an instrument is the same as that of β_0 is not restrictive; if more than k moment conditions are exploited in estimation—as will be the case in the relevant applications—they are assumed to already have been combined.) Let $\mathbf{Z}$ denote this set (which may have an infinite number of elements). Let $\{Z_t\} \in \mathbf{Z}$ and $\{\widetilde{Z}_t\} \in \mathbf{Z}$ denote two typical instrument processes. I assume

(2.3) $\{\text{vec}(Z_t u_t)',\ \text{vec}(\widetilde{Z}_t u_t)'\}'$

is covariance stationary with summable autocovariances.

Covariance stationarity is assumed for algebraic simplicity; with some complication in notation, heterogeneity such as that allowed in Bates and White (1990, 1993) could be accommodated.

Let $\hat{\beta}$ be the estimate of β_0 associated with $\{Z_t\}$. (The dependence of $\hat{\beta}$ on observable data and on sample size T is suppressed here and throughout.) A given Z_t

satisfies

(2.4a) $Z_t F_t'$ is covariance stationary;

(2.4b) $EZ_t u_t = 0$, $EZ_t F_t'$ is of rank k;

(2.4c) $$T^{-1/2}\sum_{t=1}^{T} Z_t u_t \sim_A N(0, S), \qquad S \equiv \sum_{j=-\infty}^{\infty} EZ_t u_t u_{t-j}' Z_{t-j}',$$

S positive definite

(2.4d) $$\sqrt{T}(\hat{\beta} - \beta_0) - (EZ_t F_t')^{-1}\left(T^{-1/2}\sum_{t=1}^{T} Z_t u_t\right) \rightarrow_p 0.$$

Condition (2.4a) is again maintained for algebraic simplicity. The other conditions are also standard: (2.4b) is an identification condition, (2.4c) is a central limit result, and (2.4d) is a characteristic of IV estimators. Primitive conditions that insure (2.4c) and (2.4d) may be found in Hansen (1982, 1985a).

To clarify both the notation and the interpretation of (2.4d) it may be helpful to first review the generalized method of moments (GMM) interpretation of IV estimation and then consider a linear example. For the informal GMM interpretation, let $\hat{\beta}$ be chosen in light of the moment condition $0 = EZ_t u_t = EZ_t[y_t - f_t(\beta_0)]$, so that $0 = T^{-1}\sum_{t=1}^{T} Z_t[y_t - f_t(\hat{\beta})]$. A mean value expansion applied to each element of the right hand side of this first order condition yields

$$0 = T^{-1}\sum_{t=1}^{T} Z_t[y_t - f_t(\beta_0)] + T^{-1}\sum_{t=1}^{T} Z_t\left[-\frac{\partial f_t}{\partial \beta}(\mathring{\beta})\right](\hat{\beta} - \beta_0),$$

where $\mathring{\beta}$ is the mean value whose elements lie between those of $\hat{\beta}$ and β_0. This may be rewritten

$$\sqrt{T}(\hat{\beta} - \beta_0) - (EZ_t F_t')^{-1}\left(T^{-1/2}\sum_{t=1}^{T} Z_t u_t\right)$$
$$= \left[\left\{T^{-1}\sum_{t=1}^{T}\left[Z_t \frac{\partial f_t}{\partial \beta}(\mathring{\beta})\right]\right\}^{-1} - (EZ_t F_t')^{-1}\right]\left(T^{-1/2}\sum_{t=1}^{T} Z_t u_t\right),$$

which evidently will satisfy (2.4d) under suitable conditions.

To further clarify, and to touch briefly on asymptotic properties of feasible estimators, consider a textbook two stage least squares (2SLS) example (although the results of this article do not yield any special insight under textbook conditions). Suppose $y_t = X_t'\beta + u_t$, so that $F_t' = X_t'$, where X_t' is $1 \times k$ and y_t and u_t are scalars. Let Q_t be the $q \times 1$ vector of variables appearing in the reduced form, $q \geq k$. Let $\widehat{Z}_t = \widehat{A}Q_t$, $\widehat{A} = (T^{-1}\sum_{t=1}^{T} X_t Q_t')(T^{-1}\sum_{t=1}^{T} Q_t Q_t')^{-1}$, so that $\widehat{A}$ is $k \times q$ and $\widehat{Z}_t$ is $k \times 1$. The 2SLS estimator is $\hat{\beta}_{2SLS} = (\sum_{t=1}^{T} \widehat{Z}_t X_t')^{-1}\sum_{t=1}^{T} \widehat{Z}_t y_t$. Define $A \equiv EX_t Q_t(EQ_t Q_t')^{-1}$, $Z_t \equiv AQ_t$, $\hat{\beta} = (\sum_{t=1}^{T} Z_t X_t')^{-1}\sum_{t=1}^{T} Z_t y_t$. Suppose $Z_t u_t$ is stationary and Z_t, $X_t\,(= F_t)$, u_t, and $\hat{\beta}$ satisfy (2.4). Then under the mild additional conditions that $T^{-1}\sum_{t=1}^{T} X_t Q_t' \to_p EX_t Q_t'$ and $(T^{-1}\sum_{t=1}^{T} Q_t Q_t')^{-1} \to_p (EQ_t Q_t')^{-1}$, $\sqrt{T}(\hat{\beta}_{2SLS} - \hat{\beta}) \to_p 0$, and the asymptotic distribution of $\hat{\beta}$ is also that of $\hat{\beta}_{2SLS}$.

This example is representative of overidentified models, in which a feasible combination of instruments used in estimation will depend on sample information. In such models, the Z_t referenced in my formal assumptions is an asymptotic combination that depends on population rather than sample moments. To extend the analysis to feasible estimators, one must verify that feasible estimators and the ones discussed here have the same asymptotic distribution. For such verification in models with an infinite set of underlying moment conditions, see, among others, Newey (1988), Kuersteiner (1996) and West et al. (1998).

To return to the general discussion: Let $\hat{\beta}$ and $\hat{\beta}^{\sim}$ be the estimates associated with $\{Z_t\}$ and $\{\widetilde{Z}_t\} \in \mathbf{Z}$. Let "acov" denote the asymptotic covariance between the two estimators:

$$(2.5)\quad \operatorname{acov}(\hat{\beta}, \hat{\beta}^{\sim})$$

$$\equiv \lim_{T\to\infty} E\left\{(EZ_t F_t')^{-1}\left(T^{-1/2}\sum_{t=1}^{T} Z_t u_t\right)\right\}\left\{(E\widetilde{Z}_t F_t')^{-1}\left(T^{-1/2}\sum_{t=1}^{T} \widetilde{Z}_t u_t\right)\right\}'$$

$$= (EZ_t F_t')^{-1}\left\{\lim_{T\to\infty} T^{-1}\sum_{j=-T+1}^{T-1}(T-|j|)EZ_t u_t u_{t-j}'\widetilde{Z}_{t-j}'\right\}(EF_t\widetilde{Z}_t')^{-1}$$

$$= (EZ_t F_t')^{-1}\sum_{j=-\infty}^{\infty} EZ_t u_t u_{t-j}'\widetilde{Z}_{t-j}'(EF_t\widetilde{Z}_t')^{-1}.$$

The final equality follows since $\sum_{j=-\infty}^{\infty} EZ_t u_t u_{t-j}'\widetilde{Z}_{t-j}' < \infty$ by (2.3) (Anderson, 1971: p. 460). Similarly, define

$$(2.6)\quad \operatorname{acov}(\hat{\beta}, \hat{\beta} - \hat{\beta}^{\sim}) \equiv \operatorname{acov}(\hat{\beta}, \hat{\beta}) - \operatorname{acov}(\hat{\beta}, \hat{\beta}^{\sim}) \equiv \operatorname{avar}(\hat{\beta}) - \operatorname{acov}(\hat{\beta}, \hat{\beta}^{\sim}),$$

and analogously for the asymptotic covariance between arbitrary finite linear combinations of estimators. In (2.6), "avar" denotes "asymptotic variance." From (2.4) it follows that

$$(2.7)\qquad \sqrt{T}(\hat{\beta} - \beta_0) \sim_A N[0, \operatorname{avar}(\hat{\beta})],\ \operatorname{avar}(\hat{\beta}) = (EZ_t F_t')^{-1} S (EF_t Z_t')^{-1}.$$

The aim is to characterize the Z_t that results in the smallest possible $\operatorname{avar}(\hat{\beta})$. As usual, one positive definite matrix is said to be no larger than another if the difference between them is positive semidefinite.

Through the rest of this section, $\{Z_t\}$ and $\{Z_t^*\}$ refer to elements of $\mathbf{Z}$, with corresponding estimates $\hat{\beta}$ and $\hat{\beta}^*$.

PROPOSITION 1. *Suppose that*

$$EF_t Z_t' = \sum_{j=-\infty}^{\infty} EZ_t^* u_t u_{t-j}' Z_{t-j}', \tag{2.8}$$

$$EF_t Z_t^{*\prime} = \sum_{j=-\infty}^{\infty} EZ_t^* u_t u_{t-j}' Z_{t-j}^{*\prime}. \tag{2.9}$$

Then $\text{avar}(\hat{\beta}^*) \leq \text{avar}(\hat{\beta})$.

PROOF. Conditions (2.8) and (2.9) imply that

$$\sum_{j=-\infty}^{\infty} EZ_t^* u_t u_{t-j}' Z_{t-j}^{*\prime} (EF_t Z_t^{*\prime})^{-1} = \sum_{j=-\infty}^{\infty} EZ_t^* u_t u_{t-j}' Z_{t-j}' (EF_t Z_t')^{-1}$$

$$\Longleftrightarrow (EZ_t^* F_t')^{-1} \sum_{j=-\infty}^{\infty} EZ_t^* u_t u_{t-j}' Z_{t-j}^{*\prime} (EF_t Z_t^{*\prime})^{-1}$$

$$= (EZ_t^* F_t')^{-1} \sum_{j=-\infty}^{\infty} EZ_t^* u_t u_{t-j}' Z_{t-j}' (EF_t Z_t')^{-1}$$

$$\Longleftrightarrow \text{acov}(\hat{\beta}^*, \hat{\beta}^*) = \text{acov}(\hat{\beta}^*, \hat{\beta}) \Longleftrightarrow$$

$$\text{acov}(\hat{\beta}^*, \hat{\beta}^* - \hat{\beta}) = 0 \Longrightarrow \tag{2.10}$$

$$\begin{aligned} \text{avar}(\hat{\beta}) \equiv \text{acov}(\hat{\beta}, \hat{\beta}) &= \text{acov}(\hat{\beta} - \hat{\beta}^* + \hat{\beta}^*, \hat{\beta} - \hat{\beta}^* + \hat{\beta}^*) \\ &= \text{acov}(\hat{\beta} - \hat{\beta}^*, \hat{\beta} - \hat{\beta}^*) + \text{acov}(\hat{\beta}^*, \hat{\beta}^*) \\ &\geq \text{acov}(\hat{\beta}^*, \hat{\beta}^*) \equiv \text{avar}(\hat{\beta}^*) \end{aligned}$$

The last equality follows from (2.10); the inequality follows since $\text{acov}(\hat{\beta} - \hat{\beta}^*, \hat{\beta} - \hat{\beta}^*)$ is positive semidefinite.

PROPOSITION 2. *Suppose that* Z_t^* *satisfies* (2.8) *for all* $\{Z_t\} \in \mathbf{Z}$. *Then* $\text{avar}(\hat{\beta}^*) \leq \text{avar}(\hat{\beta})$ *for all* $\hat{\beta}$.

PROOF. Note that if Z_t^* satisfies (2.8) for all possible Z_t, it does so for $Z_t = Z_t^*$ as well $\Longrightarrow$ (2.9) must hold as well. The result then follows from Proposition 1.

REMARKS. (1) If Z_t^* satisfies Proposition 2, then the efficiency bound $\text{avar}(\hat{\beta}^*)$ can also be obtained using as an instrument matrix CZ_t^* for any nonsingular $(k \times k)$ matrix C. (2) Equation (2.8) is Equation (4.9) in Hansen (1985a).[2] (3) We see

[2] Related results may be found in Breusch et al. (1999).

in (2.10) that the optimal estimator satisfies the familiar efficiency condition that it is uncorrelated with the difference between it and any other estimator. (4) It is possible that there is no $\{Z_t^*\} \in \mathbf{Z}$ that satisfies Proposition 2. Hansen (1985a) discusses conditions on $\mathbf{Z}$ that insure that an optimal estimator exists. (5) Equation (2.8) can sometimes be used to solve explicitly for Z_t^*, thereby constructively verifying that $\mathbf{Z}$ contains Z_t^* satisfying Proposition 2. This is illustrated in the next section.

3. EXAMPLE

Here I apply the results of the previous section to estimation of a univariate linear regression with an innovation that is a conditionally heteroskedastic martingale difference. Let

$$y_t = X_t\beta_0 + u_t \equiv f_t(\beta_0) + u_t, \tag{3.1}$$

where u_t, y_t, and X_t are scalars ($l = k = 1$). Lags of a scalar variable z_t are available as instruments. I assume that $(z_t, X_t, u_t)'$ is fourth-order stationary and invertible, and that u_t is mean zero and serially uncorrelated:

$$E(u_t \mid z_t, u_{t-1}\, z_{t-1}, \ldots) = 0. \tag{3.2}$$

Let ϵ_t be the Wold innovation in z_t, assumed to be a martingale difference sequence, $\epsilon_t = z_t - E(z_t \mid z_{t-1}, z_{t-2}, \ldots)$.

Let us suppose that the space of instruments is well-behaved distributed lags on z_t, or equivalently and more conveniently, ϵ_t. Thus, allowable Z_t's may be written as

$$Z_t = \sum_{j=0}^{\infty} G_j\epsilon_{t-j}, \qquad \sum_{j=0}^{\infty}|G_j| < \infty, \qquad EZ_tX_t \neq 0, \tag{3.3}$$

where the G_js are scalars. Since the data have been assumed to be fourth-order stationary, conditions (2.3) and (2.4a, b) now follow; assume that all $\{Z_t\} \in \mathbf{Z}$ satisfy (2.4c, d) as well. (Primitive conditions to insure this when, for example, $z_t = X_t = y_{t-1}$ may be found in Weiss (1984) among others.) I will let $\{G_j^*\}$ denote the distributed lag weights for the optimal Z_t^*.

For clarity in solving for $\{G_j^*\}$, I assume as well

$$Eu_t^2\epsilon_{t-j}\epsilon_{t-i} = 0 \qquad \text{when } i \neq j. \tag{3.4}$$

This assumption greatly simplifies the algebra and in many applications is probably harmless (the condition is maintained in GARCH models with conditionally symmetric distributions, for example).

Since u_t is conditionally serially uncorrelated, the right hand side of (2.8) reduces to $\sum_{j=-\infty}^{\infty} EZ_t^*u_tu_{t-j}Z_{t-j} = EZ_t^*u_t^2Z_t$. Condition (2.8) holds for arbitrary Z_t if and only if it holds for arbitrary (1×1) ϵ_{t-i} for $i \geq 0$:

$$EX_t\epsilon_{t-i} = E\left(\sum_{j=0}^{\infty} G_j^*\epsilon_{t-j}\right)u_t^2\epsilon_{t-i}. \tag{3.5}$$

From (3.4), this is simply

$$EX_t\epsilon_{t-i} = G_i^* E\epsilon_{t-i}^2 u_t^2 \Longrightarrow G_i^* = EX_t\epsilon_{t-i}/(E\epsilon_{t-i}^2 u_t^2). \tag{3.6}$$

REMARKS. (1) If $y_t = \beta_0 y_{t-1} + u_t$ and $z_t = y_{t-1} (\Rightarrow \epsilon_t = u_{t-1})$, for example, $EX_t\epsilon_{t-i} = \beta_0^i E u_t^2$, $G_i^* = \beta_0^i E u_t^2/(E u_t^2 u_{t-i-1}^2)$, and $Z_t^* = \sum_{j=0}^{\infty} G_j^*(y_{t-j-1} - \beta_0 y_{t-j-1})$. Kuersteiner (1996: p. 13) independently presents the form of the optimal weights for an autoregression of any finite order and develops as well a sophisticated nonparametric feasible estimator. (2) If u_t is conditionally homoskedastic, so that $E\epsilon_{t-i}^2 u_t^2 = E\epsilon_{t-i}^2 E u_t^2$, a little algebra reveals $Z_t^* = E(X_t \mid z_t, z_{t-1}, \ldots)/Eu_t^2$ (a result easily verified by replacing $\sum_{j=0}^{\infty} G_j^* \epsilon_{t-j}$ with $(EX_t \mid z_t, z_{t-1}, \ldots)/Eu_t^2$ on the right hand side of (3.5)). (3) Extensions to multiple equation multivariate models are straightforward if notationally complex. (4) While this estimator is efficient within the assumed space, when u_t is conditionally heteroskedastic a yet more efficient estimator may be obtained if one considers a broader space that includes stochastic functions of lagged z's (see Hansen, 1985b: pp. 33–34). I use the smaller class for clarity and because it allows for relatively straightforward extensions to more general environments. (5) Such extensions usually entail more complicated computations to solve for $\{G_i^*\}$ in terms of moments of the data. Suppose first that u_t follows a moving average process of known finite order n but is conditionally homoskedastic. Then application of the technique sketched in this section—i.e., use (2.8) with $Z_t = \epsilon_{t-i}$, $i = 0, 1, 2, \ldots$—will yield a constant coefficient linear difference equation in $\{G_i^*\}$ of order $2n$, which may be solved in routine fashion (this provides an alternative derivation of results in Hayashi and Sims, 1983; Hansen, 1985a: Section 5.2). If, as well, u_t is conditionally heteroskedastic, or one drops the symmetry assumption (3.4), the resulting difference equations will generally not have constant coefficients and may be infinite order, and thus they will be difficult to solve explicitly. Under regularity conditions, one can then proceed as follows: Use (2.8) with $Z_t = \epsilon_{t-i}$, $i = 0, 1, \ldots, J-1$, for some integer J, in each equation setting $G_i^* = 0$ for all $i \geq J$. This yields a set of J linear equations that can be solved for $G_0^*, \ldots, G_{J-1}^*$. The resulting values will not quite be those of the optimal ones, since they are solved by setting $G_i^* = 0$ for $i \geq J$. But since the optimal weights on distant ϵ's are declining (i.e., $G_i^* \to 0$ as $j \to \infty$), for large J this solution will well approximate the optimal one.

REFERENCES

ANDERSON, T. W., *The Statistical Analysis of Time Series* (New York: Wiley, 1971).

BATES, C. A., AND H. WHITE, "Efficient Instrumental Variables Estimation for Systems of Implicit Heterogeneous Nonlinear Dynamic Equations With Nonspherical Errors," in E. Berndt and H. White, eds., *Dynamic Econometric Modeling* (New York: Cambridge University Press, 1990), pp. 3–25.

——— AND ———, "Determination of Estimators With Minimum Asymptotic Covariance Matrices," *Econometric Theory* (1993), 633–48.

BREUSCH, T., HIALONG, P. SCHMIDT, AND D. WYHOWSKI, "Redundancy of Moment Conditions," *Journal of Econometrics* 91 (1999), 89–111.

HANSEN, L. P., "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50 (1982), 1029–54.

———, "A Method for Calculating Bounds on the Asymptotic Variance–Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics* 30 (1985a), 203–228.

———, "Using Martingale Difference Approximations to Obtain Covariance Matrix Bounds for Generalized Method of Moments Estimators," manuscript, University of Chicago, 1985b.

———, J. C. HEATON, AND M. OGAKI, "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," *Journal of the American Statistical Association* 83 (1988), 863–71.

——— AND K. J. SINGLETON, "Computing Semi-Parametric Efficiency Bounds for Linear Time Series Models," in W. Barnett, J. Powell, and G. Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics* (Cambridge: Cambridge University Press, 1991), pp. 387–412.

HAYASHI, F., AND C. A. SIMS, "Nearly Efficient Estimation of Time Series Models with Predetermined, but not Exogenous, Instruments," *Econometrica* 51 (1983), 783–98,

HEATON, J. C., AND M. OGAKI, "Efficiency Bounds for a Time Series Model With Conditional Heteroskedasticity," *Economics Letters* 35 (1991), 167–71.

KUERSTEINER, G. M., "Efficient IV Estimation for Autoregressive Models with Conditional Heterogeneity," manuscript, Yale University, 1996.

NEWEY, W. K., "Adaptive Estimation of Regression Models via Moment Restrictions," *Journal of Econometrics* 38 (1988), 301–39.

WEISS, A. A., "ARMA Models with ARCH Errors," *Journal of Time Series Analysis* 5 (1984), 129–43.

WEST, K. D., "Optimal Instrumental Variables Estimation," manuscript, University of Wisconsin, 1994.

——— AND D. W. WILCOX, "A Comparison of Alternative Instrumental Variables Estimators of a Dynamic Linear Model," *Journal of Business and Economic Statistics* 14 (1996), 281–93.

———, K.-F. WONG, AND STANISLAV ANATOLYEV, "Optimal Instrumental Variables Estimation of Linear Models in the Presence of Heteroskedasticity," manuscript, University of Wisconsin, 1998.