NOTE: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

THE SMALL-DISTURBANCE-ASYMPTOTIC MOMENTS OF THE INSTRUMENTAL VARIABLES AND ORDINARY LEAST SQUARES ESTIMATORS FOR A DYNAMIC EQUATION WITH CORRELATED ERRORS

Jon K. Peck

July 19, 1976
THE SMALL-DISTURBANCE-ASYMPTOTIC MOMENTS OF THE INSTRUMENTAL
VARIABLES AND ORDINARY LEAST SQUARES ESTIMATORS FOR
A DYNAMIC EQUATION WITH CORRELATED ERRORS*

by

Jon K. Peck

This paper presents an analysis of some finite-sample properties
of the instrumental variables (IV) and ordinary least squares (OLS) esti-
mators for a single equation that includes the lagged dependent variable
as a regressor. The disturbances in the equation are assumed to be nor-
mally distributed with an arbitrary nonsingular covariance matrix. Ap-
proximations are found for the mean and mean squared error of the esti-
mators. The approximations improve in accuracy as the disturbance vari-
ance becomes small. Comparisons are presented with the conventional
large-sample-asymptotic approximations for this model; some comparisons
of IV and OLS are presented.

The model to be estimated is

(1) \( y = y_{-1} \alpha + X \beta + \sigma u \)

where \( y = (y_1, \ldots, y_T)' \), \( y_{-1} = (y_0, y_1, \ldots, y_{T-1})' \),

*The research described in this paper was undertaken by grants from the
National Science Foundation and Ford Foundation. The author would like
to thank without implicating David F. Hendry and J. Kadane for their
assistance.
\[ X = (x_1, x_2, \ldots, x_K), \quad X_t = (x_{1t}, \ldots, x_{Kt})', \quad \text{and} \quad u = (u_1, \ldots, u_T). \]

The disturbances, \( u \), are assumed to be normally distributed independently of \( X \) with mean zero and a normalized nonsingular covariance matrix \( \Omega \). The regressors are of full rank, and \( X \) is assumed to be nonstochastic.

Since finite-sample results are to be obtained, some assumption must be made about the initial observation on \( y, y_0 \). Perhaps the most natural assumption would be that it is stochastic and drawn from the same distribution as the other observations, \( y_t \), but this is a conditional distribution depending on the regressor values in the pre-sample period. These values are unknown by assumption. If a distribution were assumed for the exogenous variables, a marginal distribution could be computed for \( y_0 \) if the equation were stable, but this is not consistent with the assumptions made about the exogenous variables. It is, therefore assumed that \( y_0 \) is fixed and the results obtained below are thus conditional on the value of \( y_0 \).

The approach taken in this paper is to determine the approximate bias and mean squared error of the estimators where the approximation is accurate up to terms of order \( \sigma^{2k} \), the disturbance variance, rather than the customary approximation up to terms of order \( T^k \), where \( T \) is the sample size. \( T \) is a parameter in the sigma expansions. This approach has been applied in Brown [2], Kadane [4], and Peck [6].

The approximations to be presented are referred to as the (small-sigma)-asymptotic moments of the estimator, but it is not guaranteed that these approximations converge to the exact finite-sample moments any more than the moments of a large-sample limiting distribution are necessarily the limits of the finite-sample moments. Indeed, the exact finite sample moments of IV do not even exist. (See e.g. Hatanaka [3]).
Even in such a case it can be argued that the (finite) limiting distribution moments are a useful approximate characterization of the behavior of the distribution in finite samples and are indications of where probability is concentrated. They are, perhaps, more useful than the infinite "exact" moments. The validity of this type of expansion is analyzed further in Ramage [8] and Sargan [11].

The calculation of the small-sigma moments is performed by expressing the error and squared error of the estimator, \( e = (\hat{\alpha} - \alpha, \hat{\beta} - \beta) \), as infinite series in powers of \( \sigma \). Expectations of these series are taken term by term until sufficient accuracy of the approximation has been achieved. These approximations, then, are more accurate as the disturbance variance is smaller.\(^1\)

In deriving the small disturbance moments of the estimators it is useful to distinguish between the original equation (1) and the "final form" of that equation in which \( y_{-1} \) does not appear. The final form is given by

\[
(2) \quad y = (W + \sigma V)\alpha + X\beta + \sigma u
\]

\[
= (Z + \sigma V^*) (\begin{pmatrix} \alpha \\ \beta \end{pmatrix}) + \sigma u
\]

where \( W \) and \( V \) are \( T \times 1 \) column vectors:

\[
W = (w_0, w_1, \ldots, w_{T-1})' \quad \text{and} \quad V = (v_0, v_1, \ldots, v_{T-1})'
\]

\(^1\)In the dynamic model, the error variance relative to the dispersion of \( X \) is relevant, i.e., the noise-signal ratio. This method cannot be used to study the case where the equation contains no exogenous variables (an infinite ratio).
with elements

\[
\begin{align*}
\mathbf{w}_t &= \begin{cases} 
\sum_{j=1}^{t} \alpha^{t-j} x_j + \alpha^t y_0, & t = 1, 2, \ldots, T-1 \\
y_0, & t = 0 
\end{cases} \\
\mathbf{v}_t &= \begin{cases} 
\sum_{j=1}^{t} \alpha^{t-j} u_j, & t = 1, 2, \ldots, T-1 \\
0, & t = 0 
\end{cases}
\end{align*}
\]

(3) \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} (4)

and

\[ Z \text{ and } V^* \text{ are } T \times (k+1) \text{ matrices. } Z = (W X) \text{ and } V^* = (V 0).^2 \]

\[ T \times k \hspace{1cm} T \times k \]

\[ T \times 1 \hspace{1cm} T \times k \]

Z is the nonstochastic part of the regressors including the systematic component of \( y_{-1} \) and \( V^* \) is the remaining stochastic component. It should be noted that \( W \) and \( V \) correspond to the fixed and random parts of \( y_{-1} \); not to \( y \) itself.

The following definitions of expectations are used below. \( \text{E}uu' = \Omega, \) \( \text{E}uV' = C, \) and \( \text{E}VV' = G. \) \( \Omega \) is given by assumption. \( C \) and \( G \) depend in turn on \( \Omega. \) The theorems to be presented are valid for any nonsingular \( \Omega \) provided that \( C \) and \( G \) are appropriately computed. From expression (4) above it is clear that \( V = Au \) where \( A \) is the lower triangular matrix

\[
A = \begin{bmatrix}
0 & 0 & \ldots & 0 \\
\alpha^0 & 0 & \ldots & 0 \\
\alpha^1 & \alpha^0 & 0 & \ldots \\
\vdots \\
\alpha^{T-2} & \alpha^{T-3} & \ldots & \alpha^0 & 0
\end{bmatrix}
\]

\[^2 \text{A derivation of this form is given in Peck [6].}\]
Therefore, for any \( \Omega \), \( C = EuV' = Euu'A' = \Omega A' \) and 
\[ C = EVV' = Euu'A = A\Omega A' = AC. \]

The difficulty in the estimation of equation (1) is that the lagged dependent variable is correlated with \( u \) unless \( \Omega = I \). Thus ordinary least squares is an inconsistent estimator. IV is consistent in this case for appropriately chosen instruments although biased but is asymptotically inefficient compared with estimators which take the covariance structure into account (See Sargan [10]). IV is the simplest consistent estimator to compute and can be used even when \( \Omega = I \), therefore, a preliminary test for autocorrelation or other nonrandomness in the disturbances is not required although it may be beneficial (see Peck [7]). We compute first the small disturbance moments of IV and then compute the small disturbance moments of (inconsistent) OLS for comparison.

It is assumed that a \( T \times r \) matrix, \( N \), of \( r \) nonstochastic instruments for \( y_{-1} \) is available. Let \( D = (N'; X) \). The instruments \( N \) are assumed to satisfy \( E(u|N) = 0 \) and \( D'Z \) is of full rank \( k+1 \). \( N \) is thus correlated with \( W \), the nonstochastic part of \( y_{-1} \), and uncorrelated with \( V \), the stochastic part. Lagged values of the exogenous variables \( X \) will generally satisfy these assumptions (except for the constant term and a time trend). The exogenous variables are their own instruments.\(^3\)

Define \( P_R \) as the orthogonal projector into the space spanned by the columns of \( R \), \( P_R = R(R'R)^{-1}R' \) and \( \bar{P}_R = I - P_R \). Then the instrumental variables estimator is

\[^3\text{This formulation rules out IV procedures not using all variables in } X \text{ as instruments.}\]
\begin{equation}
\left( \begin{array}{c}
\alpha \\
\beta 
\end{array} \right) = \left( y_{-1}X \right)' P_D \left( y_{-1}X \right)^{-1} \left( y_{-1}X \right)' P_D y .
\end{equation}

Since \( y_{-1}X = Z + \sigma V^* \), the error of the estimator \( e = \left( \begin{array}{c}
e_1 \\
e_2
\end{array} \right) \) can be written as
\begin{equation}
e = \sigma \left( Z + \sigma V^* \right)' P_D \left( Z + \sigma V^* \right)^{-1} \left( Z + \sigma V^* \right)' P_D u .
\end{equation}

It is assumed that the matrix inverted in expressions (5) and (6) is non-singular. The presence of the random quantity \( y_{-1} \) or \( V^* \) in these expressions means that this is not always true. However, \( Z' P_D Z \) is non-singular by assumption. The bracketed matrix can be written as
\( Z' P_D Z + \sigma M \), where \( M \) contains the random components. This is a continuous function of \( \sigma \), and, therefore, there exists a neighborhood for \( \sigma \) around zero in which the bracketed matrix is nonsingular. Then for \( \sigma \) sufficiently small the necessary inverse matrix will exist.

Writing the inverse matrix in (6) as
\[ Q [ I + \sigma ( S + \sigma V^* P_D V^* ) ] Q^{-1} \]
where \( Q = ( Z' P_D Z )^{-1} \) and \( S = Z' P_D V^* + V^* P_D Z \), and expanding the inverse as a power series in \( \sigma \) gives an approximate expression for the error of IV as

**Lemma 1.**

\[ e = \sigma Q \left[ \hat{z} u + \sigma ( -SQ\hat{z} u + V^* P_D u ) + \sigma^2 ( V^* P_D V^* \hat{z} u + SQV^* P_D u ) + \sigma^3 ( SQV^* P_D V^* Q\hat{z} u + V^* P_D V^* SQ\hat{z} u ) - SQSQ SQ\hat{z} u - V^* P_D V^* QV^* P_D u + SQSQV^* P_D u ) + O_{ P } ( \sigma^5 ) \]
where \( \hat{Z} = P_D Z = (\mathcal{R} X) \) since \( X \subset D \).

Lemma 1 will be used to find the small disturbance bias and mean squared error of IV. The bias is found in the Appendix as

\[
\text{Theorem 1}.
\]

\[
\text{Ee}_{IV} = \frac{\sigma^2}{W'P_1W} \left\{ \left[ \begin{array}{c}
1
\end{array} \right] - \text{tr}P_2C - 2Z'C'P_1W \right\}
\]

\[
+ \frac{\sigma^4}{(W'P_1W)^2} Q \left\{ \left[ \begin{array}{c}
\text{tr}P_2C \text{tr}(P_1 - P_2)G + 2 \text{tr}P_2C(P_1 - P_2)G + 2 \text{tr}P_1C \text{tr}P_2G
\end{array} \right] \right\}
\]

\[
+ 2Z'[G \text{tr}P_2C + G(P_2 - P_1)C' - \frac{1}{2} C' \text{tr}(P_1 - P_2)G + CP_2C + C'P_2C]P_1W + O(\sigma^6)
\]

where \( P_1 = \overline{P_N X W} \) and \( P_2 = \overline{P_N Y W} \). Roughly, the \( P_2 \) space contains the useful contribution of the instruments and the \( P_2 \) space the instrumental variation uncorrelated with the systematic part of \( y_{-1} \).

Considering only terms of order \( \sigma^2 \), the bias can be written as

\[
(7) \quad \frac{\sigma^2}{(W'P_1W)^2} \left[ \text{tr}[(W'P_1W)P_2 - P_1WW'P_1]C \right]
\]

\[
\quad \text{tr}(P_1WW'P_1 - W'P_1W)C - W'P_1W(X'\overline{W}X)^{-1}X'C'P_1W \right) + O(\sigma^4)
\]

where the first line gives the bias for \( \hat{\alpha} \) and the second line is the bias for \( \hat{\beta} \). The size of the bias depends not only on the number of instrumental variables, but on their statistical characteristics and the parameters of the model as well.

An immediate corollary of the theorem is
**Corollary 1.** The IV estimator will be unbiased to order $\sigma^4$ if $P_N C = 0$, i.e., the instruments are orthogonal to the columns of $C = \Omega A'$. This condition can be met in principle if only one instrument is used since the rank of $C$ is at most $T-1$ and the null space of $P_N$ is then of dimension $T-1$. For a larger number of instruments this condition would be satisfied only by chance.

The sign of the bias generally depends on several factors. Considering only the bias in $\alpha$ and making the simplifying assumptions that $N$ and $X$ are orthogonal, expression (7) becomes

$$E(\hat{\alpha} - \alpha) = \sigma^2 \frac{1}{(W'P_N W)^2} \text{tr}[(W'P_N W)^2 - P_N WW']P_N C + O(\sigma^4).$$

Since $C$ is indefinite, the bias can have either sign. If $W$ were used as an instrument, expression (8) becomes

$$E(\hat{\alpha} - \alpha) = -\sigma^2 \frac{W'CW}{(W'W)^2} + O(\sigma^4).$$

Ignoring the initial value of $y$, $y_0$, expression (9) is $-\sigma^2 \beta' X' A' \Omega A X \beta$. Thus the bias can be either positive or negative even if $\Omega = I$, since $A$ is indefinite.

From formula (8), the effect of adding an instrument uncorrelated with $W$, $X$ and $N$ can be found. The change in bias from adding such an instrument, $n$, to $N$ is

$$\sigma^2 \frac{\text{tr} P_n C}{(W'P_N W)^2} + O(\sigma^4).$$

---

Note: This is true if $X$ is serially uncorrelated and $X_{-1}$ is used as instruments.
In a case where the sign of the bias is known, it may be possible to add a random instrument which will reduce the finite sample bias of the IV estimator. Of course, this may not improve the overall performance of the estimator.

The next corollary gives the bias of \( \hat{\alpha} \) when the most obvious set of instruments, \( X_{-1} \), is used.

**Corollary 2.** If \( X \) is not autocorrelated and \( N = X_{-1} \), the bias of \( \hat{\alpha} \) is

\[
E(\hat{\alpha} - \alpha) = \frac{\sigma^2}{(\beta' X_{-1} X_{-1} \beta) \text{tr}(P_{X_{-1} W} - P_{W_{-1}})C + O(\sigma^4)}
\]

assuming \( X_0 \) is known and \( y_0 \) is zero.

**Proof:** This corollary follows from the observation that \( P_{X_{-1} W} = X_{-1} \beta \), if \( X \) is serially independent.

Using the same procedure the bias of \( \hat{\alpha} \) when a subset of the lagged exogenous variables is used is found. Assume for simplicity that \( X \) is serially independent and \( X_{-1} = (X_{11}; X_{12}) \) where \( X_{11} \) is the set of instruments for \( y_{-1} \). Correspondingly \( W = (W_1; W_2) = A(X_{11}; X_2)^{(\beta_1)}(\beta_2) \).

Suppose \( X_1 \) and \( X_2 \) are orthogonal and \( W_{-1} = (W_{11}; W_{12}) \). Then for these instruments

\[
E(\hat{\alpha} - \alpha) = \frac{1}{W_{11}^t W_{11}} \text{tr}(P_{X_{11} W} - P_{W_{11}})C + O(\sigma^4)
\]

The difference between the expectations in (11) and (12) is the change in bias due to reducing the set of instruments. This difference is
(13) \[
\frac{1}{(W_{12}^{-1}W_{11}^{-1})} \{ \text{tr}(P_{X_{12}} \bar{P}_W - P_{W_{12}})C - \frac{W_{12}^{-1}W_{11}^{-1}}{W_{12}^{-1}W_{11}^{-1}} \text{tr}(P_{X_{11}} \bar{P}_W - P_{W_{11}})C \}.
\]

We consider now the mean squared error of the IV estimator. The mean squared error is found in the appendix as

**Theorem 2**

\[
E \text{ee}' = \sigma^2 q_1 q_2 (\tilde{Z}Q + \sigma^4 q_1 q_2 [B_{11} - b - b' + \hat{Z}'(B_2 + B_2') \hat{Z}]Q + O(\sigma^6)
\]

where \( B_{11} = \begin{pmatrix} b_{11} & 0 \\ 0 & 0 \end{pmatrix} \), \( b_{11} = q_{11}^{-1}(\text{tr} P_2 C \text{tr} P_2 C + \text{tr} P_2 C P_2 C + \text{tr} P_2 G P_2) \),

\[
b = \hat{Z}'(b_1 + b_1') [P_1 W : 0],
\]

\[
b_1 = CP_2 C + CP_2 C' + C \text{tr} P_2 C + GP_2 \Omega + \frac{1}{2} \text{tr} P_2 G,
\]

\[
B_2 = C(P_1 - P_2)C + C'P_1 C - C \text{tr} P_2 C + \Omega(P_1 - P_2)G + \frac{1}{2} G \text{tr} P_1 \Omega,
\]

and \( q_{11} = (w' P_{11} w)^{-1} \).

Choosing different instruments affects the MSE through \( \hat{Z} \), \( P_1 \) and \( P_2 \). The term of order \( \sigma^2 \) in this formula agrees with the large sample result for fixed regressors except for the use of only \( Z \), the nonstochastic part of the regressors \((y_{-1}, X)\) a difference which is \( T \)-asymptotically zero. The first three elements of the \( \sigma^4 \) term contribute only to the variance and covariances of \( \hat{e} \) while the remaining terms affect the entire covariance matrix. Using the same method employed for theorems 1 and 2, the bias and MSE for OLS are found.
Theorem 3.

\[ e_{OLS} = \frac{\sigma^2}{w^T_w} \left[ \left( (X'X)^{-1} X' \right) X_w^T X_w \right] \text{tr} \left( \bar{P}_w \bar{P}_w C + \frac{1}{w^T_w} \left( (X'X)^{-1} X' \bar{W} \bar{P}_w (X' \bar{P}_w X)^{-1} X' \right) C^T \bar{P}_w \right] \]

\[ + 0(\sigma^4) \]

where \( e_{OLS} \) is \( \left( \hat{\alpha}, \hat{\beta} \right) - \left( \alpha, \beta \right) \).

Proof: \( e_{OLS} = \sigma \left( (Z + \sigma V)^{-1} (Z + \sigma V) \right)^{1} (Z + \sigma V)^{\prime} u \)

\[ = \sigma Q_0 \left( I - \sigma (V^T Z + Z^T V) Q_0 \right) (Z + \sigma V)^{\prime} u + 0_p (\sigma^3) \]

\[ = \sigma Q_0 Z^\prime u + \sigma^2 Q_0 \left( V^T u - (V^T Z + Z^T V) Q_0 Z^\prime u \right) + 0_p (\sigma^3), \text{ where } Q_0 = (Z'Z)^{-1} \]

\[ E Q_0 Z^\prime u = 0 \text{ and } EV^T u = E \begin{pmatrix} V^T u \\ 0 \end{pmatrix} \begin{pmatrix} \text{tr} C \\ 0 \end{pmatrix}, \]

\[ EV^T Q_0 Z^\prime u = \begin{pmatrix} \text{tr} P_2 C \\ 0 \end{pmatrix} \text{ and } EV^T Q_0 Z^\prime u = EV^T Q_0 Z^\prime u = C^T q_0^c, \]

where \( q_0^c \) is the first row of \( Q_0 \) and \( q_0^c \) the first column. Collecting terms gives the result:

Theorem 4: \( E e_{OLS} e_{OLS}' = \sigma^2 Q_0 Z^\prime Z Q_0 + 0(\sigma^4). \)

Proof: \( E e_{OLS} e_{OLS}' = E \sigma^2 Q_0 Z^\prime u u' Z Q_0 = \sigma^2 Q_0 Z^\prime Z Q_0 + 0(\sigma^4). \)

The terms of order \( \sigma^4 \) in these two expressions are exceedingly lengthy and are, therefore, omitted.

From Theorems 2 and 4 we find:
**Corollary 3:** If \( \Omega = I \) the difference of the covariance matrices of OLS and IV is

\[
\text{E}e_{\text{OLS}}^e' - \text{E}e_{\text{IV}}e_{\text{IV}}' = -\sigma^2 \text{K}f' + O(\sigma^4)
\]

where

\[
f = \begin{bmatrix} 1 \\ -(X'X)^{-1}X'W \end{bmatrix}
\]

and

\[
k = W_P^T \bar{W}(W_P P_N X W)^{-1}(W_P P_N X)^{-1} \left( \frac{W_X^T W_X}{W_X^T P_N W_X} - 1 \right) (W_X^T W_X)^{-1},
\]

where \( W_X = \bar{W} P_N W_X \).

**Proof:**

\[
\text{E}e_{\text{OLS}}^e' - \text{E}e_{\text{IV}}e_{\text{IV}}' = \sigma^2 (Z'Z)^{-1} [I - (Z'Z)(Z'P_D Z)^{-1}] + O(\sigma^4)
\]

\[
= -\sigma^2 (Z'Z)^{-1} Z_P Z (Z'P_D Z)^{-1} + O(\sigma^4)
\]

\[
= W_P^T \bar{W}(W_P P_N X W)^{-1}(W_P P_N X)^{-1} \begin{bmatrix} 1 & -W_X(X'X)^{-1} \\ -(X'X)^{-1}X'W & 0 \end{bmatrix} \begin{bmatrix} 1 & -\hat{W}_X(X'X)^{-1} \\ 0 & 0 \end{bmatrix}
\]

using the formula for a partitioned inverse. Since \( \hat{W}_X = W_P P_D X = W_X \), the result follows.

Since the difference of MSE matrices is negative definite unless \( D \) contains \( W \), e.g. \( W \) is its own instrument, OLS is always superior in the dynamic model as long as the errors satisfy classical assumptions. Thus this result holds not only asymptotically in \( T \) (see e.g. Theil [12, p. 412]) but in finite samples for \( \sigma \) small. Since \( k \) is the only element in this difference which is affected by the choice of instruments, more valid instruments correlated with \( W \) always lead to an improvement.
in the IV estimator, or at least no deterioration, for \( \sigma \) sufficiently small.

Comparing the bias of OLS and IV for the estimate of \( \alpha \) we find

\[
\frac{E_{\text{OLS}}}{E_{\text{IV}}} = \left( \frac{\text{tr} \bar{P}_X \bar{P}_W}{\text{tr} \bar{P}_X \bar{P}_W} \right)^2 - \frac{\text{tr}(W'\bar{P}_X \bar{P}_W' - \bar{P}_X \bar{P}_W)^2}{\text{tr}(W'\bar{P}_X \bar{P}_W' - \bar{P}_X \bar{P}_W)^2} + O(\sigma^4)
\]

assuming that the denominator is nonzero. Thus the relative bias of the two estimators depends on the relationship of the instruments to \( W \) and to \( C = EuV' \).

Finally, we illustrate some bias calculations in detail for a special case. Consider the equation

\[
y = \alpha y_{-1} + x + \gamma + u
\]

containing one exogenous variable \( x \) with unit coefficient and a constant term. Assume \( x_t = \lambda x_{t-1} + v \), and \( u_t = \alpha u_{t-1} + \eta_t \) with \( Ev = E\eta = 0 \), \( E(v|x, u, v_{-1}) = E(\eta|x, v, \eta_{-1}) = 0 \), \( Ev^2 = \sigma_v^2 \), \( E\eta^2 = \sigma^2 \). Using \( x_{-1} \) as the instrument for \( y_{-1} \), the bias expressions for the OLS and IV estimates of \( \alpha \) can be written

\[
E(\hat{\alpha} - \alpha)_{\text{OLS}} = \frac{\sigma^2}{\bar{W}'\bar{W}} \left[ \text{tr} C - \text{tr} (X'X)^{-1}X'CX - 2 \frac{\bar{W}^*\bar{W}^*}{\bar{W}'\bar{W}} \right] + O(\sigma^4)
\]

and

\[
E(\hat{\alpha} - \alpha)_{\text{IV}} = \frac{\sigma^2}{\bar{W}'\bar{W}} \left[ \text{tr}(R'R)^{-1}R'CR - 2 \frac{\bar{V}'\bar{C}^*}{\bar{W}'\bar{W}} \right] + O(\sigma^4)
\]

where \( \bar{W}^* = \bar{P}_W \), the residuals from the regression of \( W \) on \( x \), \( \bar{W} = P_N \bar{P}_X W \), the fitted values of the regression of \( W^* \) on the instruments, and \( R = P_X N \), the fitted values of the instrument (here \( x_{-1} \)) on \( x \).
With autoregressive errors it is straightforward to show that

$$\text{tr } C = \frac{\rho}{1 - \alpha \rho} \left[ T - 1 - \frac{\alpha \rho - (\alpha \rho)^T}{1 - \alpha \rho} \right],$$

which is $O(T)$.

All of the terms in square brackets in (16) and (17) are $O(1)$ except $\text{tr } C$, and $\tilde{W}^T \hat{W}$ and $\hat{W}^T \tilde{W}$ are $O(1)$, assuming $(X'X)^{-1}$ tends to a finite limit, $|\alpha| < 1$, and $|\rho| < 1$. Therefore, the large sample value of the small-sigma IV bias is zero and (16) tends to

$$\frac{\sigma^2 \text{tr } C}{\lim \tilde{W}^T \hat{W}},$$

which is nonzero unless $\rho = 0$. If further, $\lambda = 0$, (17) is $T$-asymptotically

$$\frac{\sigma (1 - \alpha^2) \sigma^2}{(1 - \alpha \rho) \sigma_X^2 (1 - \rho^2)}.$$

This differ from the $T$-asymptotic formula (Malinvaud [5], p. 560) which is, in our notation,

$$\text{plim} (\hat{\alpha} - \alpha)_{OLS} = \rho \left[ \frac{1 + \alpha \rho}{1 - \alpha^2} + \frac{(1 - \alpha \rho)(1 - \rho^2)}{(1 - \alpha^2) \sigma_X^2} \right]^{-1}$$

which includes the additional term $(1 + \alpha \rho)/(1 - \alpha^2)$. As $\sigma^2$ becomes small in (20) the second term predominates and (20) tends to (19). The additional term in (20) is of higher order in $\sigma^2$ and appears in the next term in the small-sigma bias approximation, which was not presented. No such asymptotic disagreements occur with the IV expressions.


APPENDIX

The derivation of the bias and mean squared error of the estimators requires the following lemmas.

We record some relationships among projections as

Lemma A1.

a) \( M = q_{11}(P_Z - P_X) = q_{11}P_1 \), where \( M = 2Q^CQ^C'\hat{Z}' \) and \( Q^C \) is the first column of \( Q \),

b) \( P_Z = P_X + P_1 \),

c) \( P_D = P_X + P_XP_N \),

d) \( P_D - P_Z = P_2 \),

e) \( P_XP_N = P_1 + P_2 \).

Proof:

a) The matrix \( M \) is equal to \( q_{11}(P_Z - P_X) \). This follows from expressing \( Q^CQ^C' \) as

\[
(A-14) \quad q_{11}\begin{bmatrix}
q_{11} & q_{12} \\
q_{21} & q_{22}
\end{bmatrix} - \begin{bmatrix}
0 & 0 \\
0 & q_{22} - \frac{q_{21}q_{12}}{q_{11}}
\end{bmatrix}
\]

Then, using the formula for the partitioned inverse, \( q_{22} - \frac{q_{21}q_{12}}{q_{11}} = (X'X)^{-1} \).

Hence \( M = q_{11}(2Q^C - X(X'X)^{-1}X') = q_{11}(P_Z - P_X) \). Since \( \hat{Z} = (\hat{W} X) \), \( P_Z - P_X \) is the projection operator \( P_{\hat{W}} \) where \( \hat{W} \) is obtained by regressing the residuals of the regression of \( W \) on \( X \) in turn on \( N \), i.e. \( \hat{W} = P_{N'|X'} \). Therefore \( M = q_{11}P_{N'|X'}P_{\hat{W}} = q_{11}P_1 \).
b) \[ P_Z^* = \hat{Z}'(\hat{Z}'\hat{Z})^{-1}\hat{Z}' = P_D Z'(Z'Z - Z'P_D Z)^{-1}Z'P_D. \]

Using an inversion formula from Rao ([9], p. 33), this expression is

\[ P_D Z'(Z'Z)^{-1} - (Z'Z)^{-1}Z'P_D (P_D Z'P_D - I)^{-1}P_D Z(Z'Z)^{-1}Z'P_D \]

\[ = P_D P_D P_D - P_D P_D (P_D Z'P_D - I)^{-1}P_D P_D Z'P_D = P_D P_Z, \]

interpreting the inverse matrix as a generalized inverse. Using (c),

\[ P_D P_Z = (P_X + \overline{P}_X P_N)(P_X + \overline{P}_X P_W) \]

\[ = P_X + \overline{P}_X P_N P_W = P_X + P_1. \]

c) \[ P_D = P_X + P_N - P_X P_N = P_X + \overline{P}_X P_N. \]

d) \[ P_D - P_Z = P_X + \overline{P}_X P_N - (P_X + P_1) = \overline{P}_X P_N - \overline{P}_X P_N P_W = P_2. \]

e) \[ \overline{P}_X P_N = \overline{P}_X P_N (P_W + \overline{P}_W) = P_1 + P_2. \]

\[ \text{Q.E.D.} \]

The group of lemmas below are derived from the following result
(see Anderson [1], p. 39). Let \( X_i, \ i = 1, \ldots, 4 \) be random variables with a joint normal distribution \( N(0, \Sigma) \). Then

\[ E X_i X_j X_k X_l = \sigma_{ij} \sigma_{kl} + \sigma_{ik} \sigma_{jl} + \sigma_{il} \sigma_{jk} \text{ where } \Sigma = (\sigma_{ij}). \]

Recall that \( \Psi = \begin{pmatrix} V & 0 \end{pmatrix} \), and let \( D, F, \) and \( L \) be arbitrary conformable \( T \times I \) \( T \times K \) constant matrices, the elements of \( D \) are \( d_{ij} \). Let \( d^C \) be the first column and \( d^T \) the first row of \( D \) with similar definitions for \( F \) and \( L \).
Lemma A2: \( EV^*DV^*V^*Fu = \begin{bmatrix} \text{tr } GD \& \text{tr } CF + \text{tr } GFCD' + \text{tr } GFC \\ 0 \end{bmatrix} \)

Proof: \( V^*DV^*V^*Fu = \begin{bmatrix} V^*DVV^*Fu \\ 0 \end{bmatrix} \).

The expectation of \( V^*DVV^*Fu \) is

\[
= \sum_{ijkl} E(v_1 \cdot v_j)(E(u_2 \cdot v_k)_{ij} f_{kl} + \sum_{ijkl} E(v_1 \cdot v_j)E(u_2 \cdot v_k)_{ij} f_{kl}
\]

\[
= \sum_{ijkl} E(u_2 \cdot v_1)E(v_j \cdot v_k)_{ij} f_{kl}
\]

which is the element-by-element expression for the stated expectation.

The proofs of the remaining lemmas are similar to the proof of Lemma A2 and are omitted but can be found in Peck [6].

Lemma A3: \( EV^*V^*DV^*Fu = (GD + GD' + I \text{ tr } GD)C'F^T \).

Lemma A4: \( EV^*V^*DV^*Fu = (G \text{ tr } CF + CFC + C'F'G)dC \).

Lemma A5: \( EV^*Duu'FV^* = \begin{bmatrix} a \\ 0 \end{bmatrix} \) where \( a = \text{tr } CD \& \text{tr } C'F + \text{tr } CDGF' + \text{tr } GDCF \).

Lemma A6: \( EV^*Duu'FV^* = (C' \text{ tr } CF' + C'FC' + GF')\frac{G}{dF} \).

Lemma A7: \( EV^*Duu'FV^* = C'FD'C + C'dF'C + Gtr f^CT \).
Lemma A8: \[ \text{Euu}'DV*FV* = C f^c d^c t^c C + C t r f^c d^c C + \Omega d^c f^c G . \]

Lemma A9: \[ \text{Euu}'DV*FV* = (\Omega t r G^c F + C F C' + C F C') [d^c U] . \]

Lemma A10: \[ \text{Euu}'DV*V* = CDG + CD'C + C t r C D^t . \]

Lemma A11: \[ \text{Euu}'DV*FV* = (C D G + C D'C + C t r C D^t) [I^t 0] . \]

With the aid of these lemmas Theorem 1 is now proved.

Proof of Theorem 1

All odd power terms in the sigma power series expansion of \( e \) are products of odd numbers of zero-mean normally distributed random variables and, therefore, have expectation zero. The terms of order \( \sigma^2 \) are

\[ (A-1) \quad -QSQ^t u + QV*P_D u . \]

The second of these is \( \begin{pmatrix} V^t P_D u \\ 0 \end{pmatrix} \) which has expectation \( \begin{pmatrix} \text{tr} \ P_D C' \\ 0 \end{pmatrix} . \)

The first term can be written

\[ (A-2) \quad -Q \left\{ \hat{Z}^t Vu^t ZQ^c + \begin{pmatrix} \text{tr} \ V^t P_Z u \\ 0 \end{pmatrix} \right\} \]

which gives for the expectation of \( A-1 \)

\[ (A-3) \quad Q[I t r (P_D - P_Z ) C - \hat{Z}^t C' \hat{Z}] u^c . \]

Omitting the common leading factor of \( \sigma^4 Q \), the terms of order \( \sigma^4 \) are
(A-4) \[-\psi^* P_D \phi^* \psi Q \psi^* \psi^* P_D u + SQSQ \psi^* P_D u - SQSQ SQ + SQ \psi^* P_D \phi^* Q \phi^* u + \psi^* P_D \phi^* SQ \phi^* u\]

\[= -B_1 + B_2 - B_3 + B_4 + B_5 .\]

Applying Lemma A2,

(A-5) \[EB_1 = q_{11} \left( \begin{array}{c} \text{tr} P_D G \text{tr} P_D C + 2 \text{tr} P_D C P_D G \text{tr} P_D C \\ 0 \end{array} \right) .\]

Expanding \(B_2\) gives

\[B_2 = q_{11} \hat{Z}^t \psi^* \phi^* Q \psi^* \phi^* P_D u + q_{11} \hat{Z}^t \psi^* \phi^* Q \psi^* \phi^* P_D u + q_{11} \psi^* \phi^* \hat{Z}^t \psi^* \phi^* P_D u + \psi^* \phi^* \hat{Z}^t \psi^* \phi^* P_D u \]

(A-6) \[= 2q_{11} \hat{Z}^t \psi^* \phi^* Q \psi^* \phi^* P_D u + \psi^* \phi^* (M + q_{11} P_D) \psi^* \phi^* P_D u .\]

Applying Lemma A2 and A4 gives

(A-8) \[EB_2 = 2q_{11} \hat{Z}^t \left[ G \text{tr} P_D C + C P_D G + C^t P_D G \right] \hat{Q}^C \]

\[+ \left[ \begin{array}{c} \text{tr} (M + q_{11} P_D) G \text{tr} P_D C + 2 \text{tr} (M + q_{11} P_D) C P_D G C \\ 0 \end{array} \right] \]

Multiplying out the terms in \(B_3\) gives
\[(A-9)\]
\[
\hat{z}^*v^*q\hat{z}^*v^*q\hat{z}^*v^*q\hat{z}u + q_{11}\hat{z}^*v^*q\hat{z}^*v^*v^*p_{2}\hat{z}u \\
+ q_{11}\hat{z}^*v^*v^*p_{2}\hat{z}^*v^*q\hat{z}u + q_{11}\hat{z}^*v^*v^*\hat{z}q\hat{z}^*p_{2}\hat{z}u \\
+ v^*p_{2}\hat{z}^*v^*q\hat{z}^*v^*q\hat{z}u + q_{11}v^*p_{2}\hat{z}^*v^*v^*p_{2}\hat{z}u \\
+ v^*\hat{z}q\hat{z}^*v^*p_{2}\hat{z}^*v^*q\hat{z}u + v^*\hat{z}q\hat{z}^*\hat{z}q\hat{z}^*v^*p_{2}\hat{z}u \\
= \hat{z}^*v^*v^*'(M + q_{11}p_{2})v^*q\hat{z}^*u + 2q_{11}\hat{z}^*v^*v^*\hat{z}q\hat{z}^*v^*p_{2}\hat{z}u \\
+ 2v^*p_{2}\hat{z}^*v^*v^*mu + v^*'(M + q_{11}p_{2})v^*v^*p_{2}\hat{z}u .
\]

The expectation of \( B_3 \) is found by applying Lemmas A2, A3, and A4 to be

\[(A-10)\]
\[
E_{B3} = \hat{z}^*[2G(M + q_{11}p_{2}) + I \text{ tr } (M + q_{11}p_{2})G]c^*q^c \\
+ 2q_{11}\hat{z}^*[G \text{ tr } p_{2}c + cp_{2}c + c^*p_{2}g]zq^c \\
+ \left[ \begin{array}{c} 2 \text{ tr } p_{2}g \text{ tr } mc + 4 \text{ tr } p_{2}gmc + \text{ tr } (M + q_{11}p_{2})g \text{ tr } p_{2}c \\ 2 \text{ tr } (M + q_{11}p_{2})gpe \end{array} \right] \\
0
\]

Expanding \( B_4 \) gives

\[
B_4 = q_{11}\hat{z}^*v^*v^*p_{d}v^*q\hat{z}^*u + v^*\hat{z}q\hat{z}^*v^*p_{d}v^*q\hat{z}^*u ,
\]

and \( E_{B4} \) is found from Lemmas A2 and A3 as

\[(A-11)\]
\[
E_{B4} = q_{11}\hat{z}^*[2gp_{d}c^* + c^* \text{ tr } p_{d}g]zq^c + \left[ \begin{array}{c} \text{ tr } p_{d}g \text{ tr } mc + 2 \text{ tr } p_{d}gmc \\ 0 \end{array} \right] .
\]

Finally, expanding \( B_5 \) gives

\[
B_5 = v^*p_{d}v^*v^*'(M + q_{11}p_{2})u
\]

which has expectation.
\begin{align}
(A-12) \quad E_{B_5} &= \begin{bmatrix}
\text{tr} \ D \text{tr} ( M + q_{11} P_2 ) C + 2 \text{tr} \ D \text{tr} ( M + q_{11} P_2 ) C \\
0
\end{bmatrix},
\end{align}

from an application of Lemma A2. Then collecting the \( \sigma^4 \) scalar terms in (A-3), (A-5), (A-8), (A-10), (A-11), and (A-12) and simplifying using Lemma A1 gives

\begin{align}
(A-13) \quad q_{11} [ &\text{tr}(P_X + \overline{P}_X P_N) C \text{tr}(P_1 - P_2) C + \text{tr} P_1 C \text{tr} P_2 G - \text{tr} P_1 C \text{tr}(P_X + P_1) G \\
&- \text{tr} P_1 G \text{tr}(P_X + P_1) C - \text{tr}(P_X + P_1) G \text{tr}(P_X + P_1) C + \text{tr} P_1 C \text{tr}(P_X + \overline{P}_X P_N) G \\
&+ \text{tr}(P_X + \overline{P}_X P_N) G \text{tr}(P_X + P_1) C \\
&+ 2 \text{tr}(P_X + \overline{P}_X P_N) C(P_1 - P_2) G + 2 \text{tr} P_1 C P_2 G - 2 \text{tr} P_1 C(P_X + P_1) G \\
&- 2 \text{tr}(P_X + P_1) C P_1 G - 2 \text{tr}(P_X + P_1) C(P_X + P_1) G + 2 \text{tr}(P_X + \overline{P}_X P_N) G P_1 G \\
&+ 2 \text{tr}(P_X + P_1) C(P_X + \overline{P}_X P_N) G]
\end{align}

\begin{align}
(A-14) \quad &= q_{11} [\text{tr}(P_1 + P_2) C \text{tr}(P_1 - P_2) G - \text{tr} P_1 C \text{tr}(P_1 - P_2) G - \text{tr} P_1 C \text{tr} P_1 G \\
&+ 2 \text{tr} P_1 C \text{tr}(P_1 + P_2) G \\
&+ 2 [\text{tr} P_1 C(P_1 - P_2) G + \text{tr} P_2 C(P_1 - P_2) G + \text{tr} P_1 C P_2 G - \text{tr} P_1 C P_1 G \\
&+ \text{tr} P_1 C P_2 G + \text{tr} P_1 C P_2 G] \\
(A-15) \quad &= q_{11} [\text{tr} P_2 C \text{tr}(P_1 - P_2) G + 2 \text{tr} P_1 C \text{tr} P_2 G + 2 [2 \text{tr} P_1 C P_2 G + \text{tr} P_2 C(P_1 - P_2) G]],
\end{align}

using Lemma A1 repeatedly.

A similar but simpler process gives the nonscalar terms. Explicit evaluation of \( Q = (\vec{2}^\dagger \vec{2})^{-1} \) shows that
\[ Q = (\hat{\epsilon}' \hat{\epsilon})^{-1} - (X'X)^{-1} \hat{\epsilon}' \hat{\epsilon} (X'X)^{-1} \]

Applying (A-16) to (A-15) gives Theorem 1.

The proof of Theorem 2 proceeds by computing \( ee' \) from Lemma 1; then taking expectations term by term in that expression. The term of orders which are odd powers of \( \sigma \) all consist of products of odd powers of normally distributed zero-mean random variables which contribute nothing to the expectation of \( ee' \) and they are, therefore, omitted throughout.

Except for the \( \sigma^3 \) terms,

\[ ee' = \sigma^2 Q \hat{\omega}'u \hat{\omega} + \sigma^4 Q [H_1 + H_2 + H_3] \hat{\omega} + O(\sigma^5) \]

where \[ H_1 = (-SQ^2'ru + V^* P_{D}u)(-u'^* QS + u'^* P_{D} V^*) \], and \[ H_2 = -Q'u(u'^* QS'P_{D} V^* - u'^* QS Q'S + u'^* P_{D} V^* Q'S) \].

\[ E\sigma^2 Q \hat{\omega}'u \hat{\omega} = \sigma^2 Q \hat{\omega}' \hat{\omega} \]

Multiplying out \( H_1 \) gives

\[ S(Q^2'uu'Q'S - SQ^2'uu'P_{D} V^* - V^* P_{D}uu'Q'S + V^* P_{D}uu'P_{D} V^*) \]

Let (A-19) be \[ H_{11} - H_{12} - H_{13} \]. Then multiplying out \( H_{11} \) gives

\[ \hat{\omega}'V^* Q \hat{\omega}'u P_{D} V^* + \hat{\omega}'V^* Q \hat{\omega}'u Q'S' \hat{\omega} + V^* P_{D}uu' P_{D} V^* + V^* P_{D}uu' Q'S' \hat{\omega} \]

The last term is the transpose of the first. Applying Lemmas A5, A6, and A7, the expectation is computed as
\[(A-21) \quad EH_{11} = Z^t \{ C \tr P_Z C + C' P_Z C' + GP_Z \} Z^c + \begin{bmatrix} 0 \\ 0 \\ \end{bmatrix} \] 

where \( a = (\tr P_Z C)^2 + \tr P_Z CP_Z C + \tr P_Z GP_Z \). \( H_{12} \) is 
\[Z^t v^* Q Z^t u u^* P_D v^* + v^* Q Z^t u u^* P_D v^* \] 
which has expectation 

\[(A-22) \quad EH_{12} = Z^t \{ C \tr P_D C + C' P_D C' + GP_D \} Z^c \] 
\[+ \begin{bmatrix} \tr P_Z \tr P_D C + \tr P_D CP_Z C + \tr P_D GP_Z \\ 0 \\ 0 \\ \end{bmatrix} \]

from Lemmas A5 and A6. Then 

\[(A-23) \quad EH_{13} = \begin{bmatrix} \tr P_D C \tr P_D C + \tr P_D CP_Z C + \tr P_D GP_Z \\ 0 \\ 0 \\ \end{bmatrix} \]

using Lemma A5. Turning next to the second term \( H_2 \),

\[(A-24) \quad H_2 = Z^t u u^* Z Q v^* P_D v^* + Z^t u u^* Z Q S Q S - Z^t u u^* P_D v^* Q S \] 
\[-Z^t u u^* P_D v^* + Z^t u u^* Z Q v^* Z \] 
\[-H_{21} + H_{22} - H_{23} \] 

Using Lemma A9 we find 

\[(A-25) \quad EH_{21} = Z^t \{ (\tr P_D C + 2CP_D C') \} Z^c \] 

\[(A-26) \quad H_{22} = Z^t u u^* Z Q v^* Z + q_{11} Z^t u u^* Z Q v^* Z^t \] 
\[+ Z^t u u^* Z Q v^* Z^t + Z^t u u^* Z Q v^* Z^t \] 

The expectation is found by applying Lemmas A8, A9, A10, and All to be
\( (A-27) \quad \text{EH}_{22} = 2'\{ [G_2C + CP_{2C} + C\text{tr}P_{2C}]^{2}Q^{C}0 + 
\quad 2'\{ [\text{tr}P_{2C} + 2CP_{2C}']^{2}Q^{C}0 + 
\quad 2'[\text{CMC} + C\text{trMC} + \Omega \text{MG}]Z \}. 
\)

Finally, \( H_{23} = 2'u'u'P_{D}V^{*}Q^{2}V^{*} + q_{11}2'u'u'P_{D}V^{*}Q^{2}V^{*} \) which has expectation

\( (A-28) \quad \text{EH}_{23} = 2'\{ [\Omega P_{D}G + CP_{D}C + C\text{tr}P_{D}C]^{2}Q^{C}0 + 
\quad q_{11}2'\{ [\Omega P_{D}G + CP_{D}C + C\text{tr}P_{D}C]^{2}Q^{C}0 + 
\)

from Lemmas A10 and A11.

Collecting the scalar terms of the expectations above simplifies to

\( (A-29) \quad -\text{tr}P_{2C}\text{tr}P_{2C} + \text{tr}P_{D}\text{tr}P_{2C} - \text{tr}P_{2}P_{2}C + \text{tr}P_{D}CP_{2C} 
\quad - \text{tr}P_{2}^{2}\Omega + \text{tr}P_{D}\Omega 
\quad (A-30) \quad = \text{tr}P_{2C}\text{tr}P_{2C} + \text{tr}P_{2}CP_{2C} + \text{tr}P_{2}\Omega 
\)

The column vector terms (and the transpose of the row vector terms) are

\( (A-31) \quad 2'\{ [\Omega \text{tr}P_{2C} + C'P_{2C}' + CP_{2}\Omega + C'\text{tr}P_{D}C - C'P_{D}C' - CP_{D} \Omega 
\quad - \text{tr}P_{D}G - 2CP_{2C}' + \text{tr}P_{2}P_{2C} 
\quad + \text{tr}P_{2}^{2}\Omega + \text{tr}P_{D}G - CP_{D}C - \text{tr}P_{D}C \Omega \}Q^{C} 
\quad (A-32) \quad = -2'\{ [\text{tr}P_{2C} + CP_{2C} + CP_{2} \Omega + CP_{2C}' + 1/2\text{tr}P_{2G} 
\quad + [C'\text{tr}P_{2C} + C'P_{2C}' + \text{tr}P_{2G} + CP_{2C}' + 1/2\text{tr}P_{2G}]Q^{C} \}. 
\)

Finally the full matrix terms are
\[(A-33) \quad \hat{\mathcal{L}}^I \left[ 2c'MG + Gtr M \Omega + q_{11} \left( \mathcal{P}_2 G + \mathcal{G}_C \mathcal{P}_2 C + C'P_2 C' + (C + C')tr P_2 C \right) \right. \\
+ CMC + C'MC' + (C + C')tr MC + \mathcal{G}MC + \mathcal{G}C^2 \right] \\
\left. + q_{11} \left( \mathcal{P}_D G + \mathcal{G}_D \mathcal{P}_D C + \mathcal{G}_D C^2 + (C + C')tr P_D C \right) \right]^{1/2} \\
\[(A-34) \quad = q_{11} \hat{\mathcal{L}}^I \left[ Gtr P_1 \Omega - \mathcal{P}_2 G - \mathcal{G}_2 \mathcal{P}_2 C - C'P_2 C' - (C + C')tr P_2 C \right. \\
+ \mathcal{P}_1 G + \mathcal{G}_1 \mathcal{P}_1 C + \mathcal{G}_1 C^2 + C'P_1 C' + 2C'P_1 C \right]^{1/2} .
\]

Theorem 2 now follows by collecting these expectations.