Efficient GMM estimation of weak AR processes

Kenneth D. West*

Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706, USA

Received 16 August 2001; accepted 28 November 2001

Abstract

A simple argument is used to derive the optimal GMM estimator of a finite order autoregressive process whose innovation may be conditionally heteroskedastic. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: GMM; IV; Instrumental variables; Efficiency; Moment conditions

JEL classification: C22; C32

Recent papers by Kim et al. (1999) and Broze et al. (2001) have considered GMM estimation of autoregressive processes. In this note I exploit Hansen (1985) to generalize some of their results with an alternative, and in my view simpler, proof.

Let $y_t$ be a zero mean covariance stationary variable. I consider the univariate AR($p$) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t = X'_t \phi + \epsilon_t. \quad (1)$$

Here, $X_t$ is the $(p \times 1)$ vector $(y_{t-1}, \ldots, y_{t-p})'$; $\phi = (\phi_1, \ldots, \phi_p)'$ is the $(p \times 1)$ parameter vector that is to be estimated; the parameters $\phi_1, \ldots, \phi_p$ satisfy the usual condition that the roots of the associated lag polynomial lie outside the unit circle; and $\epsilon_t$, the zero mean innovation in $y_t$, is a martingale difference satisfying

$$E(\epsilon_t | y_{t-1}, y_{t-2}, \ldots) = 0. \quad (2)$$

The innovation $\epsilon_t$ may be heteroskedastic conditional on lagged $y_t$'s.

I consider GMM estimators of $\phi$ in which the moment conditions rely on the zero correlation between $\epsilon_t$ and lags of $y_t$. It will be expositionally convenient to restate this class as: instrumental variables estimators in which the instruments are lags of $y_t$. Specifically, let $\hat{Z}_t$ be the $(p \times 1)$ vector that results after one takes an appropriate (presumably, optimal) linear combination of instruments.

*Tel.: +1-608-262-0033; fax: +1-608-262-2033.
E-mail address: kdwest@facstaff.wisc.edu (K.D. West).

0165-1765/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0165-1765(02)00015-0
One possibility is of course least squares, in which case \( \hat{Z}_t = X_t \), and the underlying set of GMM moment conditions is \( E X_t \varepsilon_t = 0 \). Another possibility is that one uses not just \( p \) but \( m > p \) lags of \( y \) as instruments. In this case, let
\[
y_t^{(m)} = (y_{t-1} \ y_{t-2} \ldots \ y_{t-m})'.
\]
Then \( \hat{Z}_t = \hat{A} y_t^{(m)r} \) for a \((p \times m)\) matrix \( \hat{A} \). If \( \hat{A} \) is chosen optimally, then it follows from Hansen (1982) that \( \hat{A} \rightarrow \rho A = (E X_t y_t^{(m)})\Omega^{-1} \), \( \Omega = E y_t^{(m)} y_t^{(m)r} \varepsilon_t^2 \). The underlying set of GMM moment conditions is \( E y_t^{(m)} \varepsilon_t = 0 \). A final possibility is that \( \hat{Z}_t \) asymptotically uses information in not just \( m \) lags but all lags of \( y_\cdot \); see Theorems 1 and 2 below.

Let the sample size be \( T \). The estimator of \( \phi \) is
\[
\left( \sum_{i=1}^{T} \hat{Z}_i X_i' \right)^{-1} \sum_{i=1}^{T} \hat{Z}_i y_i.
\]

I assume that \( \hat{Z}_t \) is such that estimation with \( \hat{Z}_t \) is asymptotically equivalent to estimation with a \((p \times 1)\) stationary vector \( Z_t \). For example, when the set of moment conditions is \( E y_t^{(m)} \varepsilon_t = 0 \), then \( Z_t = A y_t^{(m)r} \), for \( y_t^{(m)} \) and \( A \) defined in the previous paragraph. This is a standard assumption in GMM estimation. It allows us to henceforth conveniently analyze the analogue to (4) that uses \( Z_t \) rather than \( \hat{Z}_t \): \( (\Sigma_t^{1/2} \hat{Z}_t X_t')^{-1} \Sigma_t^{1/2} Z_t y_t \). Such an estimator is typically not feasible, since the optimal combination of instruments is typically not known a priori. I further make the standard assumptions that \( T^{-1} \Sigma_t^{1/2} Z_t X_t \) converges in probability to a constant matrix of rank \( p \) and that \( T^{-1/2} \Sigma_t^{1/2} Z_t \varepsilon_t \) converges in distribution to a \( N(0, S) \) random variable, \( S = EZ_i Z_i' \varepsilon_i^2 \).

Hansen (1985) establishes the following result: If \( Z^* \) satisfies the condition
\[
E Z_t X_t' = EZ_t Z_t^* \varepsilon_t^2
\]
for all possible \( Z_t \), then \( Z^* \) is the optimal instrument. Clearly, if \( Z^* \) is the optimal instrument, then so, too, is \( BZ^* \), for any nonsingular \( B \). A non-technical proof of (5) may be found in West (in press).

From (5), we immediately have:

**Theorem 1.** Suppose that \( \varepsilon_t \) is conditionally homoskedastic: \( E(\varepsilon_t^2 | y_{t-1}, y_{t-2}, \ldots) = \varepsilon_t^2 \). Then the least squares estimator is optimal.

**Proof.** Under conditional homoskedasticity, \( E (Z_t X_t' \varepsilon_t^2 | y_{t-1}, y_{t-2}, \ldots) = (EZ_t Z_t^* \varepsilon_t^2) (E \varepsilon_t^2) \). The optimality condition (5) becomes \( E (Z_t X_t' \varepsilon_t^2 | y_{t-1}, y_{t-2}, \ldots) = (EZ_t Z_t^* \varepsilon_t^2) (E \varepsilon_t^2) \) for all \( Z_t \). This is satisfied by \( Z^* = X_t' \varepsilon_t^2 \). But instrumental variables estimation with \( Z^* = X_t / E \varepsilon_t^2 \) is identical to estimation with \( Z^* = X_t \). \( \square \)

Thus, under conditional homoskedasticity there are no asymptotic benefits to using additional moments. Kim et al. (1999) establish this with results in Breusch et al. (1998), in the specific case in which the instrument vector is \( y_t^{(m)} \). The more general result (that there is no efficiency gain from use of an arbitrary, and possible infinite, set of lags beyond \( p \)) follows from (5).

Let us now allow conditional heteroskedasticity, under the assumption that
\[
E \varepsilon_t^2 \varepsilon_t | y_{t-j}, \varepsilon_{t-k} = 0 \quad \text{when } j \neq k.
\]
This assumption is satisfied by GARCH models (Bollerslev, 1986) with conditionally symmetric innovations. West (in press) and West et al. (2001) suggest solving for the optimal instrument vector in terms of \( e_i \), and then transforming to \( y_i \). So write \( Z_i^* \) as

\[
Z_i^* = \sum_{j=1}^{\infty} G_j^* e_{i-j}
\]

for a sequence of \((p \times 1)\) vectors \( \{G_j^*\} \) that are to be determined. Since the instruments are linear combinations of lagged \( y_i \)’s, they can also be expressed as linear combinations of lagged \( e_i \)’s.

This further implies that it suffices to find a \( Z_i^* \) that satisfies (5) when the vector \( Z_i \) is replaced by the scalar \( e_{i-k} \) for arbitrary \( k \geq 1 \). Thus we seek \( G_j^* \)’s that satisfy

\[
E e_{i-k} X'_t = E e_{i-k} (\sum_{j=1}^{\infty} G_j^* e_{i-j})' e_i^2, \quad k = 1, 2, \ldots
\]

From (6), \( E e_{i-k} (\sum_{j=1}^{\infty} G_j^* e_{i-j})' e_i^2 = E e_{i-k} G_k^* e_i^2 \). Upon combining this with (8) and rearranging, we have

\[
G_k^* = E e_{i-k} X_t / E e_{i-k}^2 e_i^2 \Rightarrow Z_i^* = \sum_{j=1}^{\infty} (E e_{i-j} X_t / E e_{i-j}^2 e_i^2) e_{i-j}
\]

From this Theorem 2 follows:

**Theorem 2.** The instrument is optimal if it is a non-singular linear transformation of \( Z_i^* \), satisfying

\[
Z_i^* = \sum_{j=1}^{\infty} (E e_{i-j} X_t / E e_{i-j}^2 e_i^2)(y_{i-j} - \phi_1 y_{i-j-1} - \phi_2 y_{i-j-2} - \ldots - \phi_p y_{i-j-p} = (say) \sum_{j=1}^{\infty} H_j^* y_{i-j};
\]

\[
H_1^* = E e_{i-1} X_t / E e_{i-1}^2 e_i^2, H_2^* = -\phi_1 E e_{i-1} X_t / E e_{i-2}^2 e_i^2 + E e_{i-2} X_t / E e_{i-2}^2 e_i^2, \ldots
\]

We see in (10) that solving explicitly for the optimal instrument in terms of lagged \( y_i \)’s requires solving for the sequences \( \{E e_{i-j} X_t\} \) and \( \{E e_{i-j}^2 e_i^2\}, j = 1, 2, \ldots \). The first sequence, \( \{E e_{i-j} X_t\} \), is proportional to the weights in \( y_i \)’s moving average representation, with a factor of proportionality of \( E e_i^2 \). So this sequence may be computed in familiar fashion from \( \phi_1, \ldots, \phi_p \) and \( E e_i^2 \). For example, \( E e_{i-1} X_t = E e_i^2, E e_{i-2} X_t = \phi_1 E e_i^2 \). The second sequence, \( E e_{i-j}^2 e_i^2 \), reflects conditional heteroskedasticity in \( e_i \). Given a model for conditional heteroskedasticity, one can also compute this sequence in terms of parameters and observable data. Of course, Theorem 1 is a special case of Theorem 2; under conditional homoskedasticity (i.e., \( E e_{i-j}^2 e_i^2 = E e_{i-j}^2 e_i^2 \) for all \( j \neq 0 \)), it follows from (10) that the optimal instrument is proportional to \( X_t \).

Let us apply Theorem 2 to the specific case analyzed by Broze et al. (2001). They consider an AR(1) model \((p = 1, X_t = y_{i-1}, \phi = \phi_1)\), with \( e_i = \eta_i \eta_{i-1} \) for an i.i.d. \((0, 1)\) sequence \( \{\eta_i\} \). Then (6) is satisfied. Further, \( E e_{i-k}^2 e_i^2 = E e_{i-1}^2 e_i^2 \) for \( k \geq 2 \). Let \( k = E e_{i-1}^2 e_i^2 > 1 \). Observe that \( E e_{i-k} X_t = \phi_1^{k-1} \), \( y_{i-2} = e_{i-2} + \phi e_{i-3} + \cdots, E e_i^2 = 1 \). Then according to (9),

\[
G_1^* = 1/\kappa; \quad G_k^* = \phi^{k-1} \text{ for } k \geq 2 \Rightarrow
\]

\[
Z_i^* = (1/\kappa) e_{i-1} + \phi e_{i-2} + \phi^2 e_{i-3} + \cdots
\]

\[
= (1/\kappa)(y_{i-1} - \phi y_{i-2}) + \phi y_{i-2}.
\]

Thus, for \( H_i^* \) defined in (10), we have \( H_1^* = (1/\kappa), H_2^* = \phi[1 - (1/\kappa)], H_j^* = 0 \) for \( j \geq 3 \). Using the
results in Hansen (1982), Broze et al. solve for (in my notation) $H^{*}_1$ and $H^{*}_2$, and, importantly, present numerical results showing that use of $Z^*$, as an instrument can yield great efficiency gains over least squares.

Broze et al. (2001) state that they limit themselves to consideration of two lags of $y_t$ ‘for simplicity.’ The results here establish that this restriction is without loss of generality. There are no asymptotic gains to use of any lags beyond $y_{t-2}$; such lags are redundant in the sense of Kim et al. (1999). For the general AR($p$) model, it may be shown that with the assumed form of heteroskedasticity there are no efficiency gains from use of lags beyond the $(p+1)$th. But this result is tied to the assumed form of heteroskedasticity. If one assumes instead (say) GARCH errors, then each additional lag results in strictly increased efficiency. That is, with GARCH errors, GMM estimation based on the moment condition $E y_{t}^{(m+1)} e_t = 0$ is strictly more efficient asymptotically than that with estimation using $E y_{t}^{(m)} e_t = 0$, though of course there are diminishing returns to use of distant lags.

Acknowledgements

I thank the National Science Foundation for financial support, and Shane Sherlund for helpful comments.

References