ELSEVIER

Economics Letters 75 (2002) 415–418

economics letters

www.elsevier.com/locate/econbase

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.

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

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t \equiv X_t' \phi + \varepsilon_t. \tag{1}$$

Here, X_t is the $(p \times 1)$ vector $(y_{t-1}, \ldots, y_{t-p})'$; $\phi \equiv (\phi_1, \ldots, \phi_p)'$ 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 ε_t, the zero mean innovation in y_t, is a martingale difference satisfying

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

The innovation ε_t may be heteroskedastic conditional on lagged y_t's.

I consider GMM estimators of ϕ in which the moment conditions rely on the zero correlation between ε_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
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 $EX_t\varepsilon_t = 0$. Another possibility is that one uses not just p but $m > p$ lags of y_t as instruments. In this case, let

$$y_t^{(m)} \equiv (y_{t-1}\ y_{t-2} \ldots\ y_{t-m})'. \tag{3}$$

Then $\hat{Z}_t = \hat{A}y_t^{(m)\prime}$ for a $(p \times m)$ matrix $\hat{A}$. If $\hat{A}$ is chosen optimally, then it follows from Hansen (1982) that $\hat{A} \to_p A \equiv (EX_t y_t^{(m)\prime})\Omega^{-1}$, $\Omega \equiv Ey_t^{(m)} y_t^{(m)\prime} \varepsilon_t^2$. The underlying set of GMM moment conditions is $Ey_t^{(m)}\varepsilon_t = 0$. A final possibility is that $\hat{Z}_t$ asymptotically uses information in not just m lags but all lags of y_t; see Theorems 1 and 2 below.

Let the sample size be T. The estimator of ϕ is

$$\left(\sum_{t=1}^{T} \hat{Z}_t X_t'\right)^{-1} \sum_{t=1}^{T} \hat{Z}_t y_t. \tag{4}$$

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 $Ey_t^{(m)}\varepsilon_t = 0$, then $Z_t = Ay_t^{(m)\prime}$, 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 $\hat{Z}_t$: $\left(\Sigma_{t=1}^{T} \hat{Z}_t X_t'\right)^{-1} \Sigma_{t=1}^{T} Z_t y_t$. 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 $T^{-1} \Sigma_{t=1}^{T} \hat{Z}_t X_t'$ converges in probability to a constant matrix of rank p and that $T^{-\frac{1}{2}} \Sigma_{t=1}^{T} Z_t \epsilon_t$ converges in distribution to a N(0, S) random variable, $S \equiv EZ_t Z_t' \varepsilon_t^2$.

Hansen (1985) establishes the following result: If Z^*_t satisfies the condition

$$EZ_t X_t' = EZ_t Z^{*\prime}_t \varepsilon_t^2 \tag{5}$$

for all possible Z_t, then Z^*_t is the optimal instrument. Clearly, if Z^*_t is the optimal instrument, then so, too, is BZ^*_t 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 ε_t is conditionally homoskedastic: $E(\varepsilon_t^2 | y_{t-1}, y_{t-2}, \ldots) = E\varepsilon_t^2$. Then the least squares estimator is optimal.*

Proof. Under conditional homoskedasticity, $EZ_t Z^{*\prime}_t \varepsilon_t^2 = (EZ_t Z^{*\prime}_t)(E\varepsilon_t^2)$. The optimality condition (5) becomes $EZ_t X_t' = EZ_t Z^{*\prime}_t E\varepsilon_t^2$ for all Z_t. This is satisfied by $Z^*_t = X_t / E\varepsilon_t^2$. But instrumental variables estimation with $Z^*_t = X_t / E\varepsilon_t^2$ is identical to estimation with $Z^*_t = X_t$. □

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

$$E\varepsilon_t^2 \varepsilon_{t-j} \varepsilon_{t-k} = 0 \quad \text{when } j \neq k. \tag{6}$$

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 ε_t, and then transforming to y_t. So write Z^*_t as

$$Z^*_t = \sum_{j=1}^{\infty} G^*_j \varepsilon_{t-j} \tag{7}$$

for a sequence of $(p \times 1)$ vectors $\{G^*_j\}$ that are to be determined. Since the instruments are linear combinations of lagged y_t's, they can also be expressed as linear combinations of lagged ε_t's.

This further implies that it suffices to find a Z^*_t that satisfies (5) when the vector Z_t is replaced by the scalar ε_{t-k} for arbitrary $k \geq 1$. Thus we seek G^*'s that satisfy

$$E\varepsilon_{t-k}X_t' = E\varepsilon_{t-k}\left(\sum_{j=1}^{\infty} G^*_j \varepsilon_{t-j}\right)'\varepsilon_t^2, \quad k = 1, 2, \ldots. \tag{8}$$

From (6), $E\varepsilon_{t-k}(\sum_{j=1}^{\infty} G^*_j \varepsilon_{t-j})'\varepsilon_t^2 = E\varepsilon_{t-k}^2 G^{*\prime}_k \varepsilon_t^2$. Upon combining this with (8) and rearranging, we have

$$G^*_k = E\varepsilon_{t-k}X_t / E\varepsilon_{t-k}^2\varepsilon_t^2 \Rightarrow Z^*_t = \sum_{j=1}^{\infty}(E\varepsilon_{t-j}X_t / E\varepsilon_{t-j}^2\varepsilon_t^2)\varepsilon_{t-j} \tag{9}$$

From this Theorem 2 follows:

Theorem 2. *The instrument is optimal if it is a non-singular linear transformation of Z^*_t satisfying*

$$\begin{aligned} Z^*_t &= \sum_{j=1}^{\infty}(E\varepsilon_{t-j}X_t / E\varepsilon_{t-j}^2\varepsilon_t^2)(y_{t-j} - \phi_1 y_{t-j-1} - \phi_2 y_{t-j-2} - \ldots - \phi_p y_{t-j-p}) = (\text{say}) \sum_{j=1}^{\infty} H^*_j y_{t-j}; \\ H_1^* &= E\varepsilon_{t-1}X_t / E\varepsilon_{t-1}^2\varepsilon_t^2, H_2^* = -\phi_1 E\varepsilon_{t-1}X_t / E\varepsilon_{t-1}^2\varepsilon_t^2 + E\varepsilon_{t-2}X_t / E\varepsilon_{t-2}^2\varepsilon_t^2, \ldots. \end{aligned} \tag{10}$$

We see in (10) that solving explicitly for the optimal instrument in terms of lagged y's requires solving for the sequences $\{E\varepsilon_{t-j}X_t\}$ and $\{E\varepsilon_{t-j}^2\varepsilon_t^2\}$, $j = 1, 2, \ldots$. The first sequence, $\{E\varepsilon_{t-j}X_t\}$, is proportional to the weights in y_t's moving average representation, with a factor of proportionality of $E\varepsilon_t^2$. So this sequence may be computed in familiar fashion from $\phi_1, \ldots, \phi_p$ and $E\varepsilon_t^2$. For example, $E\varepsilon_{t-1}X_t = E\varepsilon_t^2$, $E\varepsilon_{t-2}X_t = \phi_1 E\varepsilon_t^2$. The second sequence, $E\varepsilon_{t-j}^2\varepsilon_t^2$, reflects conditional heteroskedasticity in ε_t. 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., $E\varepsilon_{t-j}^2\varepsilon_t^2 = E\varepsilon_{t-j}^2 E\varepsilon_t^2$ 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 ($p = 1$, $X_t = y_{t-1}$, $\phi = \phi_1$), with $\varepsilon_t = \eta_t\eta_{t-1}$ for an i.i.d. (0, 1) sequence $\{\eta_t\}$. Then (6) is satisfied. Further, $E\varepsilon_{t-k}^2\varepsilon_t^2 = E\varepsilon_{t-k}^2 E\varepsilon_t^2$ for $k \geq 2$. Let $\kappa = E\varepsilon_{t-1}^2 E\varepsilon_t^2 > 1$. Observe that $E\varepsilon_{t-k}X_t = \phi^{k-1}$, $y_{t-2} = \varepsilon_{t-2} + \phi\varepsilon_{t-3} + \cdots$, $E\varepsilon_t^2 = 1$. Then according to (9),

$$\begin{aligned} G^*_1 &= 1/\kappa; G^*_k = \phi^{k-1} \text{ for } k \geq 2 \Rightarrow \\ Z^*_t &= (1/\kappa)\varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + \cdots \\ &= (1/\kappa)(y_{t-1} - \phi y_{t-2}) + \phi y_{t-2}. \end{aligned} \tag{11}$$

Thus, for H^*_j 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^*_t 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 $Ey_t^{(m+1)}\varepsilon_t = 0$ is strictly more efficient asymptotically than that with estimation using $Ey_t^{(m)}\varepsilon_t = 0$, 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

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 302–327.

Breusch, T., Qian, H., Schmidt, P., Wyhowski, D., 1998. Redundancy of moment conditions. Journal of Econometrics 91, 89–111.

Broze, L., Francq, C., Zakoïan, J.-M., 2001. Non-redundancy of high order moment conditions for efficient GMM estimation of weak AR processes. Economics Letters 71, 317–322.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hansen, L.P., 1985. A method for calculating bounds on the asymptotic variance–covariance matrices of generalized method of moments estimators. Journal of Econometrics 30, 203–228.

Kim, Y., Hailong, Q., Schmidt, P., 1999. Efficient GMM and MD estimation of autoregressive models. Economics Letters 62, 265–270.

West, K.D., in press. On optimal instrumental variables estimation of stationary time series models. NBER Technical Working Paper No. 243; forthcoming, International Economic Review.

West, K.D., Wong, K.-f., Anatolyev, S., 2001. Instrumental variables estimation of heteroskedastic linear models using all lags of instruments. Manuscript, University of Wisconsin.