A jackknife interpretation of the continuous updating estimator

Stephen G. Donald\textsuperscript{a,}\textsuperscript{*}, Whitney K. Newey\textsuperscript{b}

\textsuperscript{a}Department of Economics, 270 Bay State Rd., Boston University, Boston, MA 02215, USA
\textsuperscript{b}Department of Economics, M.I.T., Cambridge, MA, USA

Received 16 August 1999; accepted 21 October 1999

Abstract

Recent work on Generalized Method of Moments (GMM) estimators has suggested that the continuous updating estimator is less biased than the commonly used two-step estimator. We show that the continuous updating estimator can be interpreted as jackknife estimator. The interpretation gives some insight into why there is less bias associated with this estimator. © 2000 Elsevier Science S.A. All rights reserved.

Keywords: Generalized Method of Moments, Jackknife, Continuous updating estimator, Bias, Instrumental variables

JEL classification: C1; C2; C3; C4

1. Introduction

Recent work by Hansen et al. (1996) suggests that the continuous updating GMM (generalized method of moments) estimator has smaller bias than the more common two-step efficient GMM estimator. The purpose of this paper is to give an interpretation of the continuous updating estimator (CUE) that explains why it should have relatively small bias. We show that without autocorrelation the first-order conditions have a jackknife interpretation, where ‘own observation’ terms are deleted. These terms are known to be an important source of bias in some GMM estimators. Angrist et al. (1999) had developed a jackknife instrumental variables estimator (JIVE) that eliminates the ‘own observation’ terms to reduce bias. Here we show that this elimination is automatic for the CUE.

When an autocorrelation consistent variance matrix is used for the CUE, the first-order conditions no longer have a jackknife form, but can still be shown to be conducive to small bias. In that case the Jacobian term is orthogonal to the moment term, leading to a first-order condition with population expectation close to zero. In Donald and Newey (1999), a conditional version of this orthogonality

\textsuperscript{*}Corresponding author. Tel.: +1-617-353-4824; fax: +1-617-353-4449.
E-mail address: ocker@bu.edu (S.G. Donald)
condition was shown to hold for the Limited Information Maximum Likelihood (LIML) estimator in linear models — an estimator that is also known to have small bias relative to other estimators (such as two-stage least squares) in instrumental variable models (see Morimune, 1983; Rothenberg, 1983). These results motivate the use of these estimators as ones with relatively small bias, leading to confidence intervals that are better centered and hence more accurate.

To describe the continuous updating estimator let \( g(z, \beta) \) be an \( r \times 1 \) vector of functions of a data observation \( z \) and \( q \times 1 \) parameter vector \( \beta \) satisfying the moment condition:

\[
E[g(z, \beta)] = 0
\]

Let \( z_i, (i = 1, \ldots, n) \) denote the observations and \( g_i(\beta) = g(z_i, \beta) \). The sample first and second moments of the \( g \) are given by:

\[
\hat{g}(\beta) = \frac{1}{n} \sum_{i=1}^{n} g_i(\beta)
\]

\[
\hat{\Omega}(\beta) = \frac{1}{n} \sum_{i=1}^{n} g_i(\beta)g_i(\beta)'
\]

Here we begin with the case of no autocorrelation, where \( \hat{\Omega}(\beta) \) estimates \( \text{Var}(\sqrt{n}\hat{g}(\beta_0)) \). The CUE is obtained as the solution to a minimization problem as follows:

\[
\hat{\beta} = \arg \min_{\beta \in \beta} \hat{g}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{g}(\beta)
\]

where the minimum is over some compact parameter set \( \beta \). Assuming that the minimum occurs at an interior point, \( \hat{\beta} \) will have a first-order condition obtained by differentiating the objective function and setting it equal to zero. Assuming that \( \beta \) is a scalar, and using well known results on derivatives of inverse matrices, this first-order condition is:

\[
0 = \hat{G}'\hat{\Omega}^{-1}\hat{g} - \hat{g}'\hat{\Omega}^{-1}\hat{\Lambda}\hat{\Omega}^{-1}\hat{g}
\]

where \( \hat{g} = \hat{g}(\hat{\beta}) \) whose \( i \)th element is denoted by \( \hat{g}_i = g(z, \hat{\beta}) \), \( \hat{\Omega} = \hat{\Omega}(\hat{\beta}) \), \( \hat{G} = \frac{\partial \hat{g}(\hat{\beta})}{\partial \beta} \) whose \( i \)th element is denoted by \( \hat{g}_i = \frac{\partial \hat{g}(\hat{\beta})}{\partial \beta} \) and \( \hat{\Lambda} = \frac{\sum_{i=1}^{n} \hat{g}_i^2}{\sqrt{n}} \). In deriving this expression we have used the fact that \( \frac{\partial \hat{\Omega}(\hat{\beta})}{\partial \beta} = \hat{\Lambda} + \hat{\Lambda}' \). When \( \beta \) is a vector, then the expression in (2) is a valid expression of the first-order condition for a single element of \( \beta \) where the terms \( \hat{G} \) and \( \hat{\Lambda} \) would involve derivatives for a particular element of the \( \beta \) vector. Consequently the above expression and all of the results to follow can easily be obtained in the case where \( \beta \) is a vector. For simplicity of notation we derive all results for the case of a scalar \( \beta \).

Let \( \hat{B} = \hat{\Omega}^{-1}\hat{\Lambda} \) denote the matrix of coefficients from the multivariate regression of \( \hat{g}_i \) on \( \hat{g}_i \) and \( \hat{U}_i = \hat{g}_i - \hat{B}'\hat{g}_i \) the vector of associated residuals. By the usual orthogonality property of least-squares residuals, \( \sum_{i=1}^{n} \hat{U}_i \hat{g}_i' / n = 0 \). Collecting terms, we can rewrite the first-order condition as:

\[
0 = (\hat{G} - \hat{B}'\hat{g})'\hat{\Omega}^{-1}\hat{g} = \frac{1}{n} \sum_{i=1}^{n} \hat{U}_i \hat{\Omega}^{-1}\hat{g} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{U}_j \hat{\Omega}^{-1} \hat{g}_i = \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i} \hat{U}_j \hat{\Omega}^{-1} \hat{g}_i = \frac{1}{n} \sum_{i=1}^{n} \hat{A}_i g(z_i, \hat{\beta})
\]
where \( \hat{A}_i = (\Sigma_i^{-1})^{1/2} \hat{\Omega}^{-1}_i / n \). This expression shows the jackknife interpretation of the continuous updating estimator. We see that the usual interpretation of a GMM estimator, where a linear combination of the moments is set equal to zero, is modified to allow a linear combination coefficient for each observation, that excludes the own observation from the Jacobian of the moments. This does not effect the asymptotic distribution of the estimator, because \( \hat{g} = \alpha_p(1) \) means that each \( \hat{A}_i \) will have the same limit as \( n \to \infty \). It does, however, affect the small sample distribution of the estimator.

It is well known in the literature on finite sample properties of instrumental variables estimators that a non-zero expectation of the first-order conditions, evaluated at true parameters, leads to bias (see Nagar, 1959; Rothenberg, 1983; Buse, 1992; Angrist et al., 1999). The lack of centering of the first-order conditions will induce a lack of centering of the estimator. The usual two-step optimal GMM estimator suffers from this problem. Such an estimator is obtained from minimizing first-order conditions at \( B \) and defining

\[
\bar{g} = \left[ \sum_{i=1}^{n} g_i \right] / n,
\]

where the equality follows by the least-squares interpretation of \( \bar{g} \). Let,

\[
B = \left[ \sum_{i=1}^{n} g_i g_i' \right] = \left[ \sum_{i=1}^{n} g_i \right] \left[ \sum_{i=1}^{n} g_i' \right]
\]

This interpretation of the CUE extends to the autocorrelated case, where an autocorrelation

\[
E[\hat{g}(\beta_0) / \partial \beta | \hat{g}(\beta_0)] = \left[ \sum_{i=1}^{n} g_i \right] / n
\]

and define \( B_n = \hat{\Omega}^{-1} \Lambda_n \) to be the coefficients from a population least-squares regression of \( \hat{g}(\beta_0) / \partial \beta \) on \( \hat{g}(\beta_0) \). Then \( \hat{B} \) estimates \( B_n \) in the case where \( g_i \) is uncorrelated with \( g_j \) and \( g_{\beta_j} \) for all \( j \neq i \). Replacing \( \beta \), \( \hat{B} \), and \( \hat{\Omega} \) by \( \beta_0 \), \( B_n \) and \( \Omega_n \), respectively, in the first-order conditions for the CUE, and taking expectations gives:

\[
E[\partial \hat{g}(\beta_0) / \partial \beta - B_n \hat{g}(\beta_0)' \hat{\Omega}_n^{-1} \hat{g}(\beta_0)] = 0
\]  

where the equality follows by the least-squares interpretation of \( B_n \) and the usual orthogonality property of least-squares residuals. Thus, the centered property of the first-order conditions arises directly from a population regression residual interpretation of the first-order conditions.
consistent \( \hat{\Omega}(\beta) \) is used to construct the estimator. Consider an autocorrelation consistent estimator of the usual time domain form:

\[
\hat{\Omega}(\beta) = \hat{C}_0(\beta) + \sum_{k=1}^{K} w_{kk} \{ \hat{C}_k(\beta) + \hat{C}_k(\beta)' \}
\]

where \( \hat{C}_k(\beta) = \sum_{i=1}^{n-k} g_i(\beta)g_{i+k}(\beta)'/n \) and \( w_{kk} \) are weights that can be used to make \( \hat{\Omega}(\beta) \) positive semi-definite (for example, see Newey and West, 1987; Andrews, 1991). Here \( \hat{\Omega}(\hat{\beta}) \) is an estimator of \( \Omega_{\eta} = E[\hat{g}(\beta_0)\hat{g}(\beta_0)'] \) that allows for \( g_i \) to be autocorrelated. Also, differentiating gives, as before, \( \partial \hat{\Omega}(\hat{\beta})/\partial \beta = \hat{A} + \hat{A}' \) with:

\[
\hat{A} = \sum_{i=1}^{n} \hat{g}_i \hat{g}_i'/n + \sum_{k=1}^{K} w_{kk} \left[ \sum_{i=1}^{n-k} (\hat{g}_i \hat{g}_{i+k} + \hat{g}_{i+k} \hat{g}_i')/n \right]
\]

being an estimator of \( A_n = E[\hat{g}(\beta_0)\partial \hat{g}(\beta_0)/\partial \beta'] \) that allows for \( g_i \) to be correlated with \( g_{\beta_i} \) for all \( i \) and \( j \). Consequently, \( \hat{\beta}_n = \hat{\Omega}(\hat{\beta})^{-1} \hat{A} \) is an estimator of the population regression coefficients from a regression of \( \partial \hat{g}(\beta_0)/\partial \beta \) on \( \hat{g}(\beta_0) \), that allows for autocorrelation in the moment functions and for cross-autocorrelation between them and their derivatives. Then, the first-order conditions for the continuous updating estimator with an autocorrelation consistent estimator of the variance of the average moments will be:

\[
0 = \left[ \partial \hat{g}(\hat{\beta})/\partial \beta - \hat{\beta}_n \hat{g}(\beta_0) \right] \hat{\Omega}(\hat{\beta})^{-1} \hat{g}(\hat{\beta})
\]

It follows as before that this first-order condition will be centered at zero. Replacing \( \hat{\beta}_n \) with \( B_n = \Omega_n^{-1} A_n \), \( \hat{\Omega}(\hat{\beta}) \) with \( \Omega_{\eta} \), and \( \hat{\beta} \) with \( \beta_\eta \) in the expression to the right of the equality, and taking expectations gives zero, exactly as in Eq. (3).

It is interesting to note that if an autocorrelation consistent estimator is used for the CUE then the first-order conditions retain their centered property even if the moments are autocorrelated with their derivatives. The CUE estimator without an autocorrelation consistent weighting matrix does not share this property. Thus, even when the moments are not autocorrelated, the first-order conditions may not have zero expectation at the truth due to autocorrelation of moments and derivatives. There are many cases where such autocorrelation would be expected to occur. In rational expectations models the derivatives of the moments with respect to parameters will not generally have conditional expectation zero given the information set, and so should be correlated with past values of the moments.

This autocorrelation robust centering of the first-order conditions for the CUE should lead to an estimator with relatively small bias. Of course, bias is not the only concern in GMM estimation, and use of an autocorrelation consistent estimator could also inflate the variance of the estimator. Also, Hansen et al. (1996) show that the continuous updating estimator has relatively small bias in some models where there is autocorrelation between the moments and their derivatives, so that the potential reduction in bias from use of an autocorrelation consistent weighting matrix may not be important in practice.

**Acknowledgements**

Donald would like to thank the Alfred P. Sloan foundation for support. Newey thanks the National Science Foundation for financial support.
References

Morimune, K., 1983. Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size. Econometrica 51, 821–841.