A NOTE ON THE USE OF TWO-STEP AITKEN METHOD IN INAPPROPRIATE SITUATIONS*

By TAKU YAMAMOTO

1. Introduction and Summary

Consider the estimation problem of the parameter $\beta$ in the model

$$y = X\beta + u, \quad E(uu') = \Omega.$$ 

If $\Omega$ is known, the Aitken estimator $b_A$ is given by

$$b_A = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$ 

If $\Omega$ is not known, the common practice is to obtain a consistent estimator of $\Omega$, say $\hat{\Omega}$, and then obtain the two-step Aitken (TA) estimator $b_{TA}$ based on $\hat{\Omega}$, i.e.,

$$b_{TA} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y.$$ 

Maddala [11] has shown that the TA estimator does not have the same asymptotic distribution as the Aitken estimator, when there are lagged dependent variables among regressors.

Further, he showed that, with a special reference to the distributed lag model examined by Amemiya and Fuller [1], the TA estimator is not necessarily more efficient than the (first-step) instrument variable (IV) estimator even in the large sample, when (i) the same parameter appears both in coefficients and in the error covariance matrix and (ii) there are lagged dependent variables among regressors. Actually, he showed that relative efficiency between these estimators is indefinite; that is, the relative efficiency is depending upon the true parameter values. This result itself is quite interesting, since it shows that taking into account more information of the model does not necessarily pay off, quite contrary to the general belief.1) The interesting question now is whether we would have a stronger situation in which the TA estimator is definitely less efficient than the first-step estimator.

In this note we answer the above question by a particular example, that is, we present an estimation problem in which the TA estimator is less efficient than the first-step (in our case) ordinary least squares (OLS) estimator. Certainly, our estimation problem serves as a stronger and more clear-cut counter-example to the general belief on the TA estimator; even more so, since Maddala's result depends on the arbitrarily chosen numerical values. To my knowledge, even in other estimation problems which employ various two-step or multi-step estimation procedures,

---

* I am indebted to the referee for his valuable comments. The earlier version of this paper was written during my visit at C.O.R.E., Louvain, whose hospitality is gratefully acknowledged.

1) Grether and Maddala [5] showed that various other two-step procedures for distributed lag models are also not necessarily more efficient than their first-step estimators.
there has not been an example which gives a definitely worse result in the large sample sample for method that incorporates more information.\(^2\) Thus, our result seems to be a unique one even in a wider class of estimation problems, although ours in the present study is pedagogical rather than practical.

Our problem will be formulated as an estimation problem of an autoregressive model from the prediction point of view, and the OLS and TA estimators are introduced as ad-hoc estimation methods. In the next section, the asymptotic distributions of the OLS and TA estimators for our model are obtained, and it is shown that the OLS estimator is more efficient than the TA estimator. Further, it turns out that the above mentioned two conditions of the model are essential also in the present case.

2. Relative Efficiency of OLS and TA Estimators: A Dynamic Estimator Case

Let us consider the first order autoregressive model

\[
y_t = \alpha y_{t-1} + u_t, \quad |\alpha| < 1, \quad t=1, 2, \ldots, T, \tag{2.1}
\]

where \(y_0\) is assumed to be given, and \(u_t\) is independently identically distributed (i.i.d.) with zero mean, finite variance \(\sigma^2\) and bounded fourth absolute moment.

Klein [9, pp. 55-62] proposed several dynamic estimators of \(\alpha\) for multiperiod predictions. One of them was later called the DYN estimator by Fair [4], which is obtained by minimizing

\[
\sum_{t=0}^{T-1} (y_{t+h} - \alpha^h y_t)^2,
\]

for the sole purpose of \(h\)-period \((h \geq 2)\) ahead prediction.\(^3\) Specializing to the case of \(h=2\) for illustrative purpose, the underlying model of the above minimization is given as

\[
y_t = \alpha^2 y_{t-2} + \epsilon_t, \quad \epsilon_t = u_t + \alpha u_{t-1}, \quad t=2, 3, \ldots, T. \tag{2.2}
\]

Compactly, we express the above model as

\[
y = \alpha^2 y_{-2} + w,
\]

where \(y = [y_2, y_3, \ldots, y_T]'\), \(w = [w_2, w_3, \ldots, w_T]'\), and

\[
\text{cov}(w) = \sigma^2 \Omega = \sigma^2 \begin{bmatrix} 1 + \alpha^2 & \alpha & 0 & \cdots & 0 \\ \alpha & 1 + \alpha^2 & \alpha & \cdots & 0 \\ 0 & \alpha & 1 + \alpha^2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \alpha \\ 0 & \cdots & \cdots & \cdots & 1 + \alpha^2 \end{bmatrix}
\]

---

\(^2\) We might note that the last procedure discussed in Grether and Maddala’s paper [5] is an interesting partial example, since in that case the second-step OLS (not Aitken) estimator is definitely less efficient than the first-step IV estimator for one of the parameters. The procedure is considered partial, because the overall relative efficiency for all parameters is known to be always indefinite.

\(^3\) For discussions of DYN estimator’s theoretical and experimental properties, see Johnston [8] and Lahiri [10].

Another dynamic estimator, which was the main concern of Klein’s book, can be called the simulation path estimator. Its properties are further discussed in Klein [9, pp. 115-133] and Hartley [6].
It is obvious that the DYN estimator does not take into account of serial correlation of error terms. The interesting question is whether we gain in the efficiency of estimate by taking into account of serial correlation through the well-known TA procedure.

In the subsequent discussions, we compare the estimators of $\alpha^2$ in (2.2) by the OLS and the TA methods, which can be regarded as the DYN and the generalized DYN estimators, respectively.\(^4\) We note that the model (2.2) has quite peculiar features: (i) the coefficient parameter $\alpha^2$ is a function of the parameter $\alpha$ in error term, and (ii) it has a lagged endogenous variable as an explanatory variable, while (iii) $y_{t-2}$ and $w_t$ are uncorrelated. The first two of these features are the same as those in the distributed lag model considered by Amemiya and Fuller, when their model is written out in the reduced form.

In order to prepare for later discussions, we first recall that the maximum likelihood (ML) estimator $a_{ML}$ of $\alpha$ in (2.1) has the following asymptotic distribution (e.g. Mann and Wald [12]):

$$\sqrt{T}(a_{ML} - \alpha) \sim N(0, (1 - \alpha^2)).$$

(2.3)

Further, the asymptotically equivalent estimator of $a_{ML}$ can be obtained by performing the OLS estimation to (2.1).

Now, we consider the OLS estimator $a_{OLS}^2$ of $\alpha^2$ in (2.2), which is given as

$$a_{OLS}^2 = (y_{-2}'y_{-2})^{-1}y_{-2}'w.$$  

(2.4)

The asymptotic distribution of $\sqrt{T}(a_{OLS}^2 - \alpha^2)$ can be derived from

$$\sqrt{T}(a_{OLS}^2 - \alpha^2) = \left(\frac{1}{T}y_{-2}'y_{-2}\right)^{-1}y_{-2}'w.$$  

(2.5)

It is easily verified (e.g. Theil [13, pp. 409–411]) that

$$\lim_{T \to \infty} \frac{1}{T} y_{-2}'y_{-2} = \frac{\sigma^2}{1 - \alpha^2}.$$  

(2.6)

Further, following the central limit theorem of $m$-dependent variables by Hoeffding and Robbins [7]), we have, for large $T$,

$$\frac{1}{\sqrt{T}} y_{-2}'w \sim N\left(0, \frac{\sigma^2(1 + 3\alpha^2)}{1 - \alpha^2}\right).$$  

(2.7)

Thus, from (2.5), (2.6) and (2.7), noting the well-known convergence theorem (Crâmer [2, pp. 254–255]), we immediately find the asymptotic distribution of $\sqrt{T}(a_{OLS}^2 - \alpha^2)$ as follows:

$$\sqrt{T}(a_{OLS}^2 - \alpha^2) \sim N(0, (1 - \alpha^2)(1 + 3\alpha^2)).$$  

(2.8)

Next, we will consider the TA estimator $a_{TA}^2$ of $\alpha^2$ in (2.2). The first step is to estimate $a_{ML}$ of $\alpha$ in (2.1). The second step is to obtain the TA estimator $a_{TA}^2$ as follows:

$$a_{TA}^2 = (y_{-2}'\hat{Q}^{-1}y_{-2})^{-1}y_{-2}'\hat{Q}^{-1}w,$$  

(2.9)
where $\tilde{Q}$ is $Q$ with estimated parameter $a_{ML}$. The asymptotic distribution of $\sqrt{T}(a_{TA}-\alpha^2)$ can be derived from

$$\sqrt{T}(a_{TA}-\alpha^2) = \left(\frac{1}{T} y_{-1}' \tilde{Q}^{-1} y_{-1}\right)^{-1} \frac{1}{\sqrt{T}} y_{-1}' \tilde{Q}^{-1} w. \quad (2.10)$$

It is shown in Appendix A that

$$\lim_{T \to \infty} \frac{1}{T} y_{-1}' \tilde{Q}^{-1} y_{-1} = \lim_{T \to \infty} \frac{1}{T} y_{-1}' Q^{-1} y_{-1} = \frac{\sigma^2}{1-\alpha^2}. \quad (2.11)$$

The Taylor expansion of $(1/\sqrt{T})y_{-1}' \tilde{Q}^{-1} w$ gives us

$$\frac{1}{\sqrt{T}} y_{-1}' \tilde{Q}^{-1} w = \frac{1}{\sqrt{T}} y_{-1}' Q^{-1} w + \frac{1}{T} y_{-1}' \frac{\partial Q^{-1}}{\partial \alpha} w \cdot \sqrt{T}(a_{ML}-\alpha) + \frac{1}{\sqrt{T}} \Delta. \quad (2.12)$$

We shall neglect the third term $(1/\sqrt{T})\Delta$ in the sequel, since it is easily verified that $\lim_{T \to \infty} (1/\sqrt{T})\Delta = 0$. In order to obtain the asymptotic distribution of the above expression, we first have to get a tractable expression of the second member on the right-hand side. It is shown in Appendix B that

$$\lim_{T \to \infty} \frac{1}{T} y_{-1}' \frac{\partial Q^{-1}}{\partial \alpha} w = \frac{\alpha \sigma^2}{1-\alpha^2}. \quad (2.13)$$

Further, it is obvious from the construction of the ML estimator $a_{ML}$ that, for large $T$,

$$\sqrt{T}(a_{ML}-\alpha) = \left(\frac{1}{T} y_{-1}' y_{-1}\right)^{-1} \frac{1}{\sqrt{T}} y_{-1}' u = \left(\frac{\sigma^2}{1-\alpha^2}\right)^{-1} \frac{1}{\sqrt{T}} y_{-1}' u. \quad (2.14)$$

Then, we can express (2.12) as follows:

$$\frac{1}{\sqrt{T}} y_{-1}' \tilde{Q}^{-1} w = \frac{1}{\sqrt{T}} \left( y_{-1}' Q^{-1} w + \frac{\alpha(1-\alpha^2)}{(1-\alpha)} y_{-1}' u \right). \quad (2.15)$$

It is shown in Appendix C that we have asymptotically

$$\frac{1}{\sqrt{T}} y_{-1}' \tilde{Q}^{-1} w \sim N\left(0, \frac{\sigma^2}{(1-\alpha^2)^2}(1-\alpha^2)(1+4\alpha^2)\right). \quad (2.16)$$

As before, noting the convergence theorem, we find, from (2.10), (2.11) and (2.15), the asymptotic distribution of $\sqrt{T}(a_{TA}-\alpha^2)$ as

$$\sqrt{T}(a_{TA}-\alpha^2) \sim N(0, (1-\alpha^2)(1+4\alpha^2)). \quad (2.16)$$

Comparing (2.8) with (2.16), it is obvious in the present model that the TA estimator is less efficient than the OLS estimator (for all admissible parameter values).

The essential point of the inefficiency of the TA estimator comes from the fact that, as shown in (2.13), $\lim_{T \to \infty} (1/\sqrt{T})\Delta = 0$, but converges to a finite value $\alpha a^2/(1-\alpha^2)$, although $y_{-1}$ and $w$ are uncorrelated.

If it vanishes, or equivalently if $Q$ is known exactly, the Aitken estimator (with known $Q$), denoted by $a_{A}$, can be obtained and its asymptotic distribution is given by

$$\sqrt{T}(a_{A}-\alpha^2) \sim N(0, 1-\alpha^2).$$

Of course, the above Aitken estimator is rather self-contradictory, since the known $Q$ means the known $\alpha$, and then there is no use of estimating it. Still, it presents
an alternative example that the TA estimator is less efficient than the Aitken estimator when there is a lagged dependent variable among explanatory variables.\(^6\)

It is worth noting that the statement holds even if explanatory variables and error term, \(y_{t-2}\) and \(\epsilon_t\) in the present case, are contemporaneously uncorrelated.\(^7\)

Finally, we might note that, in the above argument, the TA estimator is based on the first-step ML estimator \(a_{ML}\) of \(\alpha\), rather than the OLS estimator \(a_{OLS}\) of \(\alpha^2\). Of course, we can easily construct the TA estimator based on \(a_{OLS}\) as the first step, provided that we know \textit{a priori} the sign of the true parameter value \(\alpha\). Following the above procedure, we can similarly obtain the asymptotic distribution of the TA estimator based on \(a_{OLS}\), denoted by \(a_{TAO}\), as

\[
\sqrt{T}(a_{TAO} - \alpha^2) \sim N\left(0, (1-\alpha^2)\left(\frac{9}{4} + \frac{15}{4} \alpha^2 + \alpha^4\right)\right).
\]

The simple comparison with (2.8) shows that \(a_{TAO}\) is also less efficient than \(a_{OLS}\).

Thus, regardless of the first-step estimator used, the TA estimator has been shown to be less efficient than the OLS estimator. Needless to say, when used in two-period predictions, the TA estimator (generalized DYN) yields less efficient predictions than the OLS estimator (DYN).

Appendix A

In this appendix, we show that

\[
\lim_{T \to \infty} T^{-\frac{1}{2}} \frac{y_{-2}' \hat{Q}^{-1} y_{-2}}{y_{-2}' \hat{Q}^{-1} y_{-2} \sim} \frac{\sigma^2}{1-\alpha^4}.
\]

Since the first equality is obvious, we prove only the second one. Before going into the proof, we introduce useful expressions of \(y_{-2}\) and \(\hat{Q}^{-1}\), which will be extensively used in discussions below. First, we express \(y_{-2}\) as

\[
y_{-2} = \tilde{y}_{-2} + M u_{-2},
\]

where

\[
y_{-2} = [y_0, y_1, \ldots, y_{r-2}]', \quad \tilde{y}_{-2} = [y_0, \alpha y_0, \ldots, \alpha^{r-2} y_0]',
\]

\[
u_{-2} = [0, u_1, \ldots, u_{r-2}]', \quad M = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
\alpha & 1 & \cdots & 0 \\
\alpha & \cdots & \cdots & \cdots \\
\alpha^{r-2} & \cdots & \cdots & \alpha \\
\alpha^{r-2} & \cdots & \cdots & \alpha
\end{bmatrix}.
\]

Second, we note that, ignoring the end effect, we can express \(\hat{Q}^{-1}\) as

\[
\hat{Q}^{-1} = K^{\prime -1} K^{-1},
\]

where

---

\(^6\) See also Amemiya and Fuller [1], Maddala [11], and Dhrymes [3, p. 182 and p. 200].

\(^7\) Note the rather misleading proof in Maddala and statement in Dhrymes [3, p. 182], which seems to imply the contemporaneous correlation between explanatory variables and error term is at stake. The correlation is a sufficient condition for non-vanishing (2.13), but not necessary.
April 1977  T. Yamamoto: A Note on the Use of Two-Step Aitken Method

Then, we have

\[ K^{-1} = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
-\alpha & 1 & \cdots & 0 \\
& & \ddots & \vdots \\
& & & 1
\end{bmatrix}. \]

From (A.4) and (A.5), we complete the proof of (A.1) by Chebyshev’s inequality.

Appendix B

In this appendix, we show that

\[ \lim_{T \to \infty} \frac{1}{T^2} E(y_{-2}' \Omega^{-1} y_{-2}) = \frac{\sigma^2}{1-\alpha^T}. \]

Further, we find that

\[ \lim_{T \to \infty} \frac{1}{T^2} E(y_{-2}' \Omega^{-1} y_{-2})^2 = \frac{\sigma^2}{(1-\alpha^T)^2}. \]

From (A.4) and (A.5), we complete the proof of (A.1) by Chebyshev’s inequality.
Following the similar argument in Appendix A, it is easily verified that

\[
\lim_{T \to \infty} \frac{1}{T} E\left(y_{t'} \frac{\partial Q^{-1}}{\partial \alpha} w\right) = \frac{\alpha \sigma^2}{1 - \alpha^2}. \tag{B.2}
\]

Thus, from (B.2) and (B.3), we complete the proof of (B.1) by Chevyshev's inequality.

Appendix C

In this appendix, we show that, as \( T \) goes to infinity,

\[
\frac{1}{\sqrt{T}} y_{t'} \tilde{Q}^{-1} w \sim N\left(0, \frac{\sigma^4}{(1 - \alpha^2)^2} (1 - \alpha^2)(1 + 4\alpha^2)\right). \tag{C.1}
\]

From (2.14), \( (1/\sqrt{T}) y_{t'} \tilde{Q}^{-1} w \) is given by

\[
\frac{1}{\sqrt{T}} y_{t'} \tilde{Q}^{-1} w = \frac{1}{\sqrt{T}} \left(y_{t'} Q^{-1} w + \frac{\alpha(1 - \alpha^2)}{(1 - \alpha^2)} y_{t'}, u\right).
\]

From (A.2), (A.3) and \( w = Ku \), we can rewrite the above when \( T \) goes to infinity as

\[
\frac{1}{\sqrt{T}} y_{t'} \tilde{Q}^{-1} w = \frac{1}{\sqrt{T}} \left(u_{t'} M' K^{-1} + \frac{\alpha(1 - \alpha^2)}{(1 - \alpha^2)} \sum_{i=2}^{T} s_{2i}\right) = \frac{1}{\sqrt{T}} \left(\sum_{i=2}^{T} s_{2i} - \frac{\alpha(1 - \alpha^2)}{(1 - \alpha^2)} \sum_{i=2}^{T} s_{2i}\right) = \frac{1}{\sqrt{T}} \left(\sum_{i=2}^{T} s_{2i} + \frac{\alpha(1 - \alpha^2)}{(1 - \alpha^2)} s_{2i}\right), \tag{C.2}
\]

where \( s_{2i} = \sum_{m=0}^{m^*} \alpha^{2m} u_{t-2-2m} \), and \( m^* \) is the maximum integer of \( (t-2)/2 \), and \( s_{2i} = \sum_{m=0}^{m^*} \alpha^{2m} u_{t-2-2m} u \). The central limit theorem of \( m \)-dependent variables by Hoeffding and Robbins [7] is readily applicable to the above, and we immediately obtain the desired result.

(Soka University)

REFERENCES

April 1977  

T. Yamamoto: A Note on the Use of Two-Step Aitken Method
