Efficiency Comparisons for a System GMM Estimator in Dynamic Panel Data Models

Frank Windmeijer
Institute for Fiscal Studies
7 Ridgmount Street
London WC1E 7AE

IFS Working Paper Series No. W98/1

Abstract

The system GMM estimator in dynamic panel data models combines moment conditions for the differenced equation with moment conditions for the model in levels. An initial optimal weight matrix under homoscedasticity and non-serial correlation is not known for this estimation procedure. It is common practice to use the inverse of the moment matrix of the instruments as the initial weight matrix. This paper assesses the potential efficiency loss from the use of this weight matrix using the efficiency bounds as derived by Liu and Neudecker (1997).

1 Introduction

A standard practice to estimate the parameters in dynamic panel data models is to take first differences to eliminate the correlated individual specific effects, and estimate the differenced model by Generalised Method of Moments (GMM) using appropriately lagged level variables as instruments. As the information of the instruments for the differenced model decreases as the series become more persistent, Arellano and Bover (1995) and Blundell and Bond (1997) have proposed use of a system GMM estimator that combines the differenced equation with the level equation. The instruments for the level equation are lagged differences of the variables, which are valid when these differences are uncorrelated with the individual effects. Blundell and Bond (1997) show that the system estimator has superior properties in terms of small sample bias and RMSE, especially for persistent series.

The GMM estimator is a two-step estimator. In the first step, an initial positive semidefinite weight matrix is used to obtain consistent estimates of the parameters. Given these consistent estimates, a weight matrix can be constructed that is consistent for the efficient weight matrix, and this weight matrix is used for the asymptotically efficient two-step estimates. It is well known, see e.g. Arellano and Bond (1991), that the two-step estimated standard errors have a small sample downward bias in this dynamic panel data setting, and one-step estimates with robust standard errors are often preferred. Although an efficient weight matrix for the differenced model with errors that are homoscedastic and that are not serially correlated is easily derived, this is not the case for the system estimator, combining differences and levels information.

It is common practice to use the inverse of the moment matrix of the instruments as an initial weight matrix. In this paper the potential efficiency loss will be considered in a model with homogeneous and non-serially correlated errors. To do this, upper bounds for the efficiency loss will be calculated as derived by Liu and Neudecker (1997) based on the Kantorovich Inequality (KI). These upper bounds indicate that the efficiency loss could potentially be quite severe. When the variance of the individual unobserved heterogeneity is small, efficiency can be gained by using a weight matrix that is optimal under the assumption that the variance of the unobserved heterogeneity is equal to zero.

In section 2, an AR(1) dynamic panel data model is considered and a description of the system GMM estimator is given. In section 3 the upper bounds of the efficiency loss are calculated for 3 and 4 time periods respectively. Section 4 concludes.

2 Model and System GMM Estimator

Consider the AR(1) panel data specification

$$y_{it} = \alpha_0 y_{it-1} + \eta_i + \varepsilon_{it} \tag{1}$$

for $i = 1, ..., N$, $t = 2, ..., T$, with N large, and T fixed. The error terms follow the error components structure in which

$$E(\eta_i) = 0 \quad ; \quad E(\varepsilon_{it}) = 0,$$
$$E\left(\varepsilon_{it}^2\right) = \sigma_\varepsilon^2 \quad ; \quad E\left(\eta_i^2\right) = \sigma_\eta^2,$$
$$E(\eta_i \varepsilon_{it}) = 0 \quad ; \quad E(\varepsilon_{it}\varepsilon_{is}) = 0,\ t \neq s.$$

The y_{it} series are assumed stationary with an infinite time horizon and therefore the series can alternatively be written as

$$y_{it} = \frac{\eta_i}{1 - \alpha_0} + \sum_{j=0}^{\infty} \alpha_0^j \varepsilon_{it-j}. \tag{2}$$

The OLS and within groups estimators of α_0 in model (1) are biased and inconsistent. A consistent estimator for α_0 is the system GMM estimator, as proposed by Arellano and Bover (1995) and Blundell and Bond (1997), utilising the following $(T+1)(T-2)/2$ moment conditions[1]

$$E\left[(\Delta y_{it} - \alpha_0 \Delta y_{it-1})(y_{it-2}, ..., y_{i1})\right] = 0 \tag{3}$$
$$E\left[(y_{it} - \alpha_0 y_{it-1}) \Delta y_{it-1}\right] = 0, \tag{4}$$

for $t = 3, ..., T$. Moment conditions (3) are for the model in first differences, utilising appropriately lagged levels information as instruments, whereas conditions (4) are for the model in levels, utilising lagged differences as instruments. As Blundell and Bond (1997) show, the system estimator is considerably more

[1] Under homoscedasticity there are additional moment conditions available to improve efficiency, see Ahn and Schmidt (1995). These extra moment conditions are not considered here.

efficient than the traditional GMM estimator utilising the moment conditions of the differenced model only.

Define

$$Z_i = \begin{bmatrix} y_{i1} & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & y_{i1} & y_{i2} & \ldots & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & y_{i1} & \ldots & y_{iT-2} & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & \Delta y_{i2} & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 & \Delta y_{i3} & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 & 0 & \ldots & \Delta y_{T-1} \end{bmatrix}$$

$$v_i = v_i(\alpha_0) = \begin{bmatrix} \Delta y_{i3} - \alpha_0 \Delta y_{i2} \\ \Delta y_{i4} - \alpha_0 \Delta y_{i3} \\ \ldots \\ \Delta y_{iT} - \alpha_0 \Delta y_{iT-1} \\ y_{i3} - \alpha_0 y_{i2} \\ y_{i4} - \alpha_0 y_{i3} \\ \ldots \\ y_{iT} - \alpha_0 y_{iT-1} \end{bmatrix},$$

and

$$f_i(\alpha_0) = Z_i' v_i.$$

Moment conditions (3) and (4) imply that $E(f_i(\alpha_0)) = 0$. The GMM estimator[2] $\hat{\alpha}$ for α_0 minimises

$$\left[\frac{1}{N}\sum_{i=1}^{N} f_i(\alpha)\right]' W_N \left[\frac{1}{N}\sum_{i=1}^{N} f_i(\alpha)\right],$$

with respect to α; where W_N is a positive semidefinite weight matrix which satisfies $\text{plim}_{N\to\infty} W_N = W$, with W a positive definite matrix. Regularity conditions are in place such that $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} f_i(\alpha) = E(f(\alpha))$ and $\frac{1}{\sqrt{N}} f_i(\alpha_0) \to N(0, \Psi)$. Let $F(\alpha) = E(\partial f_i(\alpha)/\partial\alpha)$ and $F_0 \equiv F(\alpha_0)$, then $\sqrt{N}(\hat{\alpha} - \alpha_0)$ has a limiting normal distribution, $\sqrt{N}(\hat{\alpha} - \alpha_0) \to N(0, V_W)$, where

$$V_W = (F_0' W F_0)^{-1} F_0' W \Psi W F_0 (F_0' W F_0)^{-1}. \tag{5}$$

[2] See Hansen (1982), Ogaki (1993). Here, the same notation as in Liu and Neudecker (1997) is used.

3 Efficiency Comparisons

As is clear from the expression of the asymptotic variance matrix V_W, (5), the efficiency of the GMM estimator is affected by the choice of the weight matrix W_N. An optimal choice is a weight matrix for which $W = \Psi^{-1}$. The asymptotic variance matrix is then given by $(F_0'\Psi^{-1}F_0)^{-1}$. For any other W the GMM estimator is less efficient as

$$\left(F_0'\Psi^{-1}F_0\right)^{-1} \cdot (F_0'WF_0)^{-1} F_0'W\Psi WF_0 (F_0'WF_0)^{-1}.$$

In panel data models the efficient estimator is obtained in a two-step procedure. The one-step GMM estimator $\hat{\alpha}$ is obtained using an arbitrary weight matrix W_{N1}. Let $\hat{u}_i = v_i(\hat{\alpha})$. The efficient two-step estimator is then based on the weight matrix

$$W_{N2} = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'\hat{u}_i\hat{u}_i'Z_i\right)^{-1}; \quad \text{plim}\, W_{N2} = \Psi^{-1}.$$

Although the efficient estimator is easily obtained, there is a serious problem associated with it as the estimated standard errors of the two-step estimator can be biased downwards quite severely for moderate sample sizes N, as has been documented by Arellano and Bond (1991) and Blundell and Bond (1997), who performed Monte Carlo simulations with sample sizes $N = 200$. Therefore, inference based on the two-step estimator can be very unreliable. In contrast, the one-step estimated standard errors based on the asymptotic variance matrix (5), using W_{N2} as an estimate for Ψ and substituting $\hat{\alpha}$ for α_0, are found to be much less biased, and inference, like Wald tests, much more reliable. In practice therefore, one can often only rely on inference based on the less efficient one-step estimator.

For the GMM estimator that only utilises the moment conditions (3) for the differenced model, an optimal weight matrix is $\left(\frac{1}{N}\sum_{i=1}^{N} D_i'HD_i\right)^{-1}$, where D_i is the left upper block of Z_i and H is a $(T-2)$ square matrix which has 2's on the main diagonal, -1's on the first subdiagonals and zeros elsewhere. Setting $W_{N1} = \left(\frac{1}{N}\sum_{i=1}^{N} D_i' H D_i\right)^{-1}$ results therefore in an efficient one-step estimator. For the system GMM estimator such an efficient one-step weight matrix is not known, and in practice one uses as an initial weight matrix $W_{N1} = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i' Z_i\right)^{-1}$. To assess the potential loss in efficiency from using this initial weight matrix, the following expression for the upper bound of the efficiency loss has been derived by Liu and Neudecker (1997, p.350) on the basis of the Kantorovich Inequality (KI):

$$\left(F_0' W F_0\right)^{-1} F_0' W \Psi W F_0 \left(F_0' W F_0\right)^{-1} \leq \frac{\left(\lambda_1 + \lambda_p\right)^2}{4\lambda_1 \lambda_p} \left(F_0' \Psi^{-1} F_0\right)^{-1} \qquad (6)$$

where $\lambda_1 \geq \ldots \geq \lambda_p$ are the eigenvalues of the matrix ΨW.

For $T = 3$, there is one overidentifying moment condition, as the system estimator utilises the following two moment conditions

$$\begin{aligned} E\left[\left(\Delta y_{i3} - \alpha \Delta y_{i2}\right) y_{i1}\right] &= 0 \\ E\left[\left(y_{i3} - \alpha y_{i2}\right) \Delta y_{i2}\right] &= 0, \end{aligned}$$

and Ψ is given by

$$\begin{aligned} \Psi &= \operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} y_1^2 \left(\Delta\varepsilon_{i3}\right)^2 & y_{i1} \Delta y_{i2} \left(\eta_i + \varepsilon_{i3}\right) \Delta\varepsilon_{i3} \\ y_{i1} \Delta y_{i2} \left(\eta_i + \varepsilon_{i3}\right) \Delta\varepsilon_{i3} & \left(\Delta y_{i2}\right)^2 \left(\eta_i + \varepsilon_{i3}\right)^2 \end{bmatrix} \\ &= \sigma_\varepsilon^2 \begin{bmatrix} 2\sigma_y^2 & -(1-\alpha)\sigma_y^2 \\ -(1-\alpha)\sigma_y^2 & 2\frac{\sigma_\eta^2 + \sigma_\varepsilon^2}{1+\alpha} \end{bmatrix} \end{aligned}$$

where

$$\sigma_y^2 = \frac{\sigma_\eta^2}{(1-\alpha)^2} + \frac{\sigma_\varepsilon^2}{1-\alpha^2}.$$

Further,

$$\begin{aligned} W_1 &= \left(\operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} Z_i' Z_i\right)^{-1} \\ &= \left(\operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} y_{i1}^2 & 0 \\ 0 & \left(\Delta y_{i2}\right)^2 \end{bmatrix}\right)^{-1} \\ &= \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & 2\frac{\sigma_\varepsilon^2}{1+\alpha} \end{bmatrix}^{-1} \end{aligned}$$

and the matrix $G = \Psi W_1$ is given by

$$G = \begin{bmatrix} 2\sigma_\varepsilon^2 & -(1-\alpha^2)\sigma_y^2/2 \\ -(1-\alpha)\sigma_\varepsilon^2 & \sigma_\varepsilon^2 + \sigma_\eta^2 \end{bmatrix}.$$

Figure 1 presents the plot of the function $b_{KI} = (\lambda_{G1} + \lambda_{Gp})^2 / 4\lambda_{G1}\lambda_{Gp}$, where the λ_G's are the eigenvalues of G, for various values of α_0 and $\sigma_\eta^2/\sigma_\varepsilon^2$. When $\sigma_\eta^2 = \sigma_\varepsilon^2$, b_{KI} is constant for different values of α and equal to $4/3$, indicating that the asymptotic variance of the one-step estimator could potentially be 33% larger than the efficient estimator. When $\sigma_\eta^2/\sigma_\varepsilon^2 < 1$, b_{KI} is declining with α_0, whereas it is increasing with α_0 when $\sigma_\eta^2/\sigma_\varepsilon^2 > 1$. The value of b_{KI} increases with $\sigma_\eta^2/\sigma_\varepsilon^2$ when $\sigma_\eta^2/\sigma_\varepsilon^2 > 1$.

[**Figure 1 about here**]

When $T = 4$, there are 4 overidentifying moment conditions, and the matrices Ψ and W_1 are given by

$$\Psi = \sigma_\varepsilon^2 \begin{bmatrix} 2\sigma_y^2 & -\sigma_y^2 & -\delta & -(1-\alpha)\sigma_y^2 & \frac{2-\alpha}{1-\alpha}\sigma_\eta^2 \\ -\sigma_y^2 & 2\sigma_y^2 & 2\delta & \frac{\sigma_\varepsilon^2}{1+\alpha} & -(1-\alpha)\delta \\ -\delta & 2\delta & 2\sigma_y^2 & -\frac{\sigma_\varepsilon^2}{1+\alpha} & -(1-\alpha)\sigma_y^2 \\ -(1-\alpha)\sigma_y^2 & \frac{\sigma_\varepsilon^2}{1+\alpha} & -\frac{\sigma_\varepsilon^2}{1+\alpha} & 2\frac{\sigma_\eta^2+\sigma_\varepsilon^2}{1+\alpha} & -\frac{1-\alpha}{1+\alpha}\sigma_\eta^2 \\ \frac{2-\alpha}{1-\alpha}\sigma_\eta^2 & -(1-\alpha)\delta & -(1-\alpha)\sigma_y^2 & -\frac{1-\alpha}{1+\alpha}\sigma_\eta^2 & 2\frac{\sigma_\eta^2+\sigma_\varepsilon^2}{1+\alpha} \end{bmatrix}$$

$$W_1 = \begin{bmatrix} \sigma_y^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_y^2 & \delta & 0 & 0 \\ 0 & \delta & \sigma_y^2 & 0 & 0 \\ 0 & 0 & 0 & 2\frac{\sigma_\varepsilon^2}{1+\alpha} & 0 \\ 0 & 0 & 0 & 0 & 2\frac{\sigma_\varepsilon^2}{1+\alpha} \end{bmatrix},$$

where

$$\delta = \sigma_y^2 - \frac{\sigma_\varepsilon^2}{1+\alpha}.$$

Figure 2 presents the efficiency bounds for the one-step system estimator when $T = 4$. The values for b_{KI} are larger than for the $T = 3$ case. When $\sigma_\eta^2/\sigma_\varepsilon^2 = 1$, b_{KI} is no longer constant for different values of α_0, and takes values around 3, indicating that the asymptotic variance of the one-step estimator could be 3 times the variance of the efficient estimator. Again, b_{KI} increases with $\sigma^2_\eta/\sigma^2_\varepsilon$ when $\sigma^2_\eta/\sigma^2_\varepsilon > 1$, and b_{KI} reaches the value 6 when $\sigma^2_\eta/\sigma^2_\varepsilon = 2.5$.[3]

[Figure 2 about here]

3.1 An Optimal Weight Matrix when $\sigma_\eta^2=0$

An optimal weight matrix for the system GMM estimator when $\sigma^2_\eta = 0$ is given by

$$W_{N,\sigma^2_\eta=0} = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' A Z_i \right)^{-1},$$

where

$$A = \begin{bmatrix} H & C \\ C' & I_{T-2} \end{bmatrix}$$

with H as defined above, I_{T-2} is the identity matrix of order $(T-2)$, and

$$C = \begin{bmatrix} 1 & 0 & 0 & & 0 \\ -1 & 1 & 0 & & 0 \\ 0 & -1 & 1 & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}.$$

Using this weight matrix instead of $W_{N1} = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' Z_i \right)^{-1}$ may improve on efficiency when σ^2_η is small. Figures 3 and 4 display the values for b_{KI} when $W_{N,\sigma^2_\eta=0}$ is used in the one-step estimator, for $T = 3$ and $T = 4$ respectively. Indeed, for small values of σ^2_η the potential loss in efficiency is seen to be smaller than when W_{N1} is used. However, when $\sigma^2_\eta/\sigma^2_\varepsilon$ is large, the potential efficiency loss gets larger for $W_{N,\sigma^2_\eta=0}$, which is what one would expect.

[Figures 3 and 4 about here]

One way to detect whether use of $W_{N,\sigma^2_\eta=0}$ could be beneficial, without actually calculating the variances of the components, is to calculate the efficiency bounds

[3] The value of b_{KI} increases with T as the number of moment restrictions increases. For example when $T = 6$ and $\sigma^2_\eta/\sigma^2_\varepsilon = 1$, b_{KI} is approximately 14.

b_{KI} for the efficiency difference between the one-step and two-step estimators, i.e. calculate b_{KI} from eigenvalues of the matrices $W_{N2}W_{N,\sigma_\eta^2=0}$, and $W_{N2}W_{N1}$. If the former are closer to 1 than the latter, this is an indication that there could be an efficiency gain from using $W_{N,\sigma_\eta^2=0}$ instead of W_{N1}.

4 Discussion

Upper bounds for the efficiency loss of the one-step system GMM estimator for a dynamic AR(1) panel data model as compared to the efficient two-step estimator, show that the efficiency loss could be quite severe when the weight matrix $W_{N1} = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'Z_i\right)^{-1}$ is used, especially when T gets large. When the variance of the unobserved individual effects, σ_η^2, is small, an efficiency gain can be made by using a weight matrix that is optimal under the assumption that $\sigma_\eta^2 = 0$.

The upper bounds were shown to be quite large, for example when $T = 4$, $W_{N1} = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'Z_i\right)^{-1}$ and $\sigma_\eta^2/\sigma_\varepsilon^2 = 2.5$, the ratio of the asymptotic variance of the inefficient estimator to that of the efficient estimator can be as large as 6 for high values of α_0. In Monte Carlo studies however, such large differences of the variances are not found, using normal and non-normal data generating processes. This could mean that the KI upper bounds, b_{KI}, are too large to be informative for these cases. When the b_{KI} are close to one, there is evidence of an efficient one-step estimator. The opposite statement for large b_{KI} may not be true. Further research is needed to assess whether the b_{KI} are informative to rank different one-step estimators on the basis of their relative KI -values.

In empirical settings, one can easily compute b_{KI} from the eigenvalues of the matrix $W_{N2}W_{N1}$, where W_{N2} is the two-step efficient weight matrix.

Acknowledgments

I would like to thank Steve Bond for helpful comments and for drawing my attention to the use of $W_{N,\sigma_\eta^2=0}$. The usual disclaimer applies.

References

[1] Ahn, S.C. and P. Schmidt (1995), Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-28.

[2] Arellano, M. and S. Bond (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-98.

[3] Arellano, M. and O. Bover (1995), Another Look at the Instrumental-Variable Estimation of Error-Components Models, *Journal of Econometrics*, 68, 29-51.

[4] Blundell, R. and S. Bond (1997), Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, in press.

[5] Hansen, L.P. (1982) 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica*, 50, 1029-1054.

[6] Liu, S. and H. Neudecker (1997), Kantorovich Inequalities and Efficiency Comparisons for Several Classes of Estimators in Linear Models, *Statistica Neerlandica*, 51, 345-355.

[7] Ogaki, M. (1993) 'Generalized Method of Moments: Econometric Applications', in: Maddala, G.S., C.R. Rao, and H.D. Vinod (eds.), 'Econometrics', *Handbook of Statistics*, 11, North-Holland, Amsterdam.

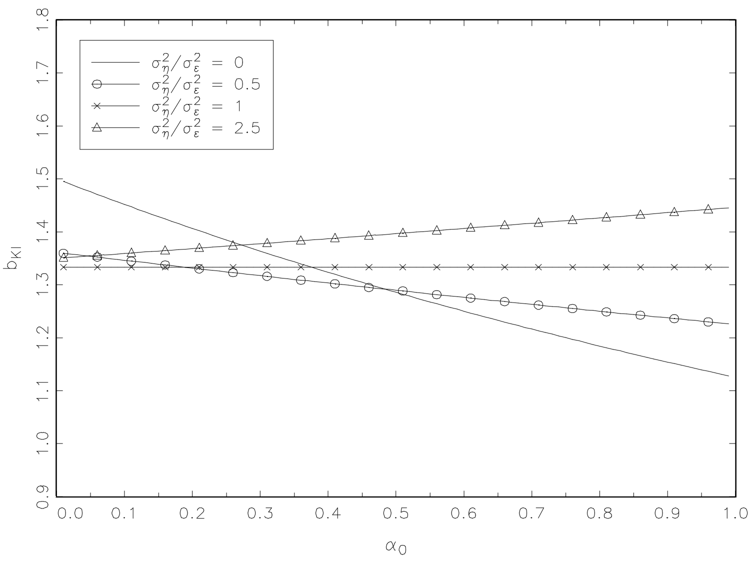

Figure 1: KI Efficiency Bounds, $T = 3$; $W_{N1} = \left(\frac{1}{N}\sum_i Z_i'Z_i\right)^{-1}$

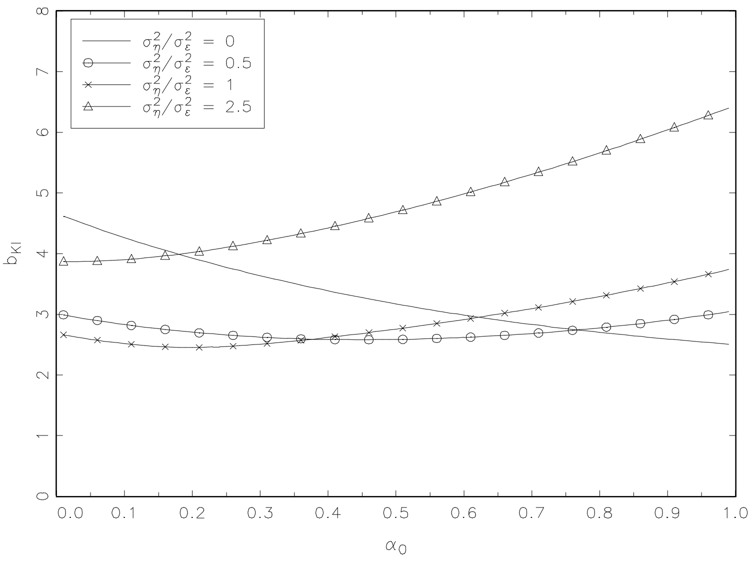

Figure 2: KI Efficiency Bounds, $T = 4$; $W_{N1} = \left(\frac{1}{N}\sum_i Z_i'Z_i\right)^{-1}$

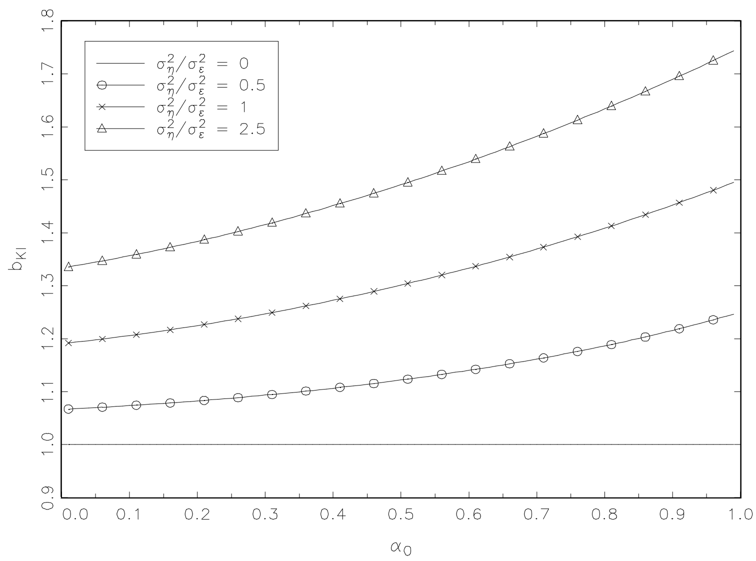

Figure 3: KI Efficiency Bounds, $T = 3$; $W_{N1} = W_{N,\sigma_\eta^2=0}$

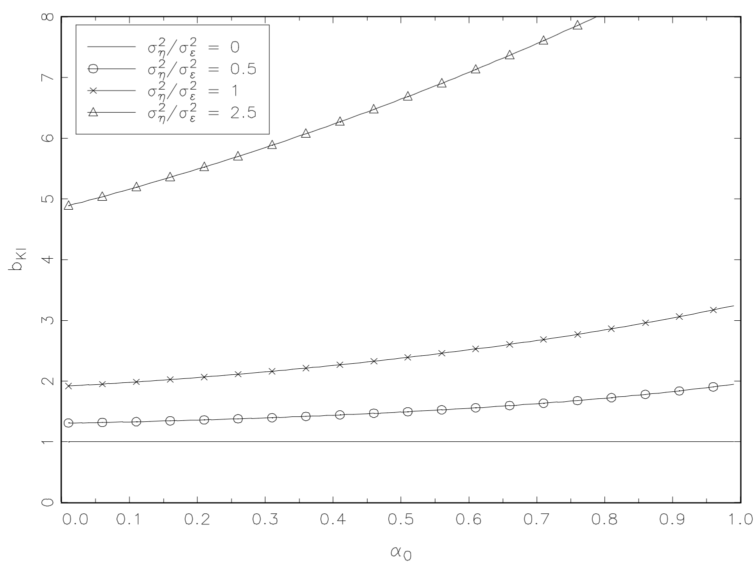

Figure 4: KI Efficiency Bounds, $T = 4$; $W_{N1} = W_{N,\sigma_\eta^2=0}$