# On a paradox in GMM estimation with nuisance parameters

Georgios Skoulakis *
University of Maryland

October 13, 2006

**Abstract**

This paper examines an apparent paradox in GMM parameter estimation in the presence of nuisance parameters. Conventional wisdom might suggest that the asymptotic efficiency of the estimator of the parameter of interest should increase when the true value of the nuisance parameter is known and this knowledge is incorporated in the estimation. It turns out, however, that the opposite phenomenon, arguably characterized as counterintuitive, is possible under certain scenarios. We provide necessary and sufficient conditions for the paradox to occur. Examples are given to illustrate the several possible cases. An example of an asset pricing factor model shows that the risk premium estimator that uses the true and known factor mean could be more or less efficient than the estimator that assumes no knowledge of the factor mean, depending on the underlying parameters.

*Department of Finance, Robert H. Smith School of Business, University of Maryland. Comments and suggestions are welcome at gskoulak@rhsmith.umd.edu.

# 1 Introduction

Consider the estimation of a parameter $\theta = (\theta_1', \theta_2')'$ based on a set of moment conditions where $\theta_1$ is the parameter of interest and $\theta_2$ is a nuisance parameter. When a parametric model is available, maximum likelihood estimation can be described as a moment-based estimation procedure that uses the loglikelihood function. Suppose that the true value of the nuisance parameter $\theta_2$ can be assumed known. In this case, and under certain regularity conditions, the estimator of $\theta_1$ that uses the known value of $\theta_2$ is at least as asymptotically efficient as the estimator of $\theta_1$ that does not make use of the known value of $\theta_2$ and is obtained through joint estimation of $\theta_1$ and $\theta_2$. This fact is in line with the common intuition that suggests incorporating additional (valid) information into the estimation should lead to efficiency gains at least in asymptotic terms.

Nevertheless, when more general moment conditions are utilized, the opposite counter-intuitive phenomenon might take place in which case using the known (and true) value of the nuisance parameter $\theta_2$ makes the estimator of $\theta_1$ less asymptotically efficient. An early example of such paradoxical behavior in which a statistic becomes more efficient when the true value of a parameter is replaced by an estimator can be traced back to Pierce (1982). Particular examples for which the aforementioned paradox occurs can be found in Robins et al. (1992), Robins et al. (1994), Lawless et al. (1999) and Zou and Fine (2002). In a recent paper, Henmi and Eguchi (2004) provide a thorough analysis of this paradox in the context of estimating functions. Using geometric arguments and projected estimating functions, they obtain a common lower bound for the asymptotic covariance matrices of the two estimators of $\theta_1$ obtained when the nuisance parameter $\theta_2$ is assumed known and unknown respectively. Since the lower bound is common, a ranking of the two estimators in terms of asymptotic efficiency is obtained whenever the lower bound is achieved for one of the inequalities. In the setup of Henmi and Eguchi (2004), observations are i.i.d. and the number of estimating functions equals the dimension of the parameter $\theta$. This environment corresponds to an exactly identified GMM estimation with i.i.d. data. We generalize their results to a more general GMM setup incorporating over-identifying restrictions and non-i.i.d. time series observations. Using simple matrix algebraic arguments we obtain the common lower bound for the two estimators of interest and use it to rank them when certain conditions hold. Although this result is useful towards understanding the paradox, it is not powerful enough in the sense that there exist cases in which the bound is not achieved in either inequality. Therefore, a conclusion cannot be reached as to which estimator is more efficient. To characterize the paradox, we obtain a matrix that serves as a "sufficient statistic" for addressing the paradox. The positive (or negative) definiteness of this matrix determines whether the paradox occurs (or not). Specific examples are provided to illustrate cases in which the paradox does or does not occur. In particular, we offer an interesting example of an asset pricing factor model for which the risk premium is the parameter of interest and the factor mean is the nuisance parameter. Incorporating the known (and true) value of the factor mean does not necessarily improve the asymptotic efficiency of the risk premium estimate. This example is interesting because, in contrast to other instances of the paradox, the model is linear in the parameters and further the ranking of the two estimators changes depending on the values of the underlying parameters.

Other research on GMM estimation in the presence of nuisance parameters includes the work of Crepon, Kramarz and Trognon (1997). They show how to estimate the parameter of interest following a two-step hybrid GMM procedure while replacing the nuisance parameter by an empirical counterpart. The resulting estimator of the parameter of interest is as efficient as the estimator obtained using the complete set of moment conditions and thus it entails no loss of efficiency. Another issue studied in the literature that might appear related to our paper is the concept of partial redundancy of moment conditions. This subject has thoroughly been studied by Breusch et al. (1999) and Qian (2002). Roughly speaking, a subset of moment conditions is called partially redundant for the parameter $\theta_1$ if there is no efficiency loss in estimating $\theta_1$ associated with ignoring this set of moment conditions. The difference with our framework is that when we estimate $\theta_1$ based on a subset of conditions we use the assumed known (and true) value of the nuisance parameter $\theta_2$ instead of an estimator.

The rest of the paper is organized as follows. In section 2, we provide the motivation for the subject of the paper. In section 3, we describe the estimation environment, state the main question we seek to address, and provide a theoretical result characterizing the paradox. In section 4, we offer three examples to illustrate some of the possible scenarios. The final section concludes. The proofs not contained in the main text are collected in the Appendix.

## 2 Motivation

The subject of this paper was motivated by the following example from the asset pricing literature. Linear factor models are widely used in asset pricing to describe the trade-off between risk and return. Given a pervasive factor of small dimension, a factor model prescribes a linear relationship between expected asset returns and betas. Specifically, let $f_t$ be the $k$-dimensional factor. Assume there are $n$ assets in the economy and denote by $R_t$ the vector of asset returns in excess of the risk-free rate. Denote by $\mu_f$ and $\mu_R$ the expected values of $f_t$ and $R_t$ respectively. The beta matrix is the $n \times k$ matrix defined by $B = \Sigma_{Rf}\Sigma_f^{-1}$ where $\Sigma_{Rf} = E[(R_t - \mu_R)(f_t - \mu_f)']$ and $\Sigma_f = E[(f_t - \mu_f)(f_t - \mu_f)']$. The factor model, in its most familiar form, is described by the pricing equation $\mu_R = B\lambda$, where $\lambda$ is the vector of risk premia. In the beta representation above, the parameter of interest is the vector of risk premia $\lambda$. Alternatively, we can define $b = \Sigma_f^{-1}\lambda$ and rewrite the pricing equation as $\mu_R = \Sigma_{Rf}b$. In this represenation $b$ measures covariance risk, whereas $\lambda$ measures beta risk in the standard beta representation. The GMM approach to estimating $b$ is based on the asset pricing equation $\mu_R = \Sigma_{Rf}b$, which is equivalent to the moment condition

$$E[R_t - R_t(f_t - \mu_f)'b] = 0_n. \tag{1}$$

We are primarily interested in estimating the parameter $b$, and in that sense $\mu_f$ is a nuisance parameter. Since the factor mean $\mu_f$ is typically unknown, we also need to make use of the moment condition

$$E[f_t - \mu_f] = 0_k. \tag{2}$$

Estimation then proceeds by applying GMM on the vector moment condition

$$E[g(x_t;\theta)] = 0_{n+k}, \tag{3}$$

where $x_t = (R_t', f_t')'$, $\theta = (b', \mu_f')'$, and

$$g(x_t;\theta) = \begin{bmatrix} R_t - R_t(f_t - \mu_f)'b \\ f_t - \mu_f \end{bmatrix} \equiv \begin{bmatrix} g_1(x_t;\theta) \\ g_2(x_t;\theta) \end{bmatrix}. \tag{4}$$

We denote the GMM estimator of $\theta$ based on the moment condition (3) by $\theta_T = (b_T', \mu_T')'$.

Let us now assume that the factor mean $\mu_f$ is known. Under this assumption, one can use GMM to obtain an estimator of the parameter $b$, denoted by $\tilde{b}_T$, based on the moment condition (1) and the known and true value of $\mu_f$. The following question then arises: is estimator $\tilde{b}_T$ more efficient than estimator $b_T$? At first glance, it would appear that $\tilde{b}_T$ should be more efficient since it uses the true value of $\mu_f$ while the procedure used to obtain $b_T$ instead treats $\mu_f$ as unknown and estimates it. It turns out, that the answer to the question is actually not straightforward and it depends on the underlying parameters. Indeed, there are cases in which $b_T$ is more efficient than $\tilde{b}_T$. We view this situation as counterintuitive and paradoxical. In the next section, we address this question in a general GMM setup and obtain a theoretical result characterizing the paradox.

# 3 Analysis of the paradox

Consider using GMM to estimate the parameter $\theta \in \mathbb{R}^k$ based on the moment conditions

$$E[g(x_t;\theta)] = 0_n \tag{5}$$

where $g(\cdot;\cdot)$ is an $n$-dimensional function with $n \geq k$, and $x_t$ is the observed time series data. The following standard notation is used throughout the paper: $0_n$ denotes the $n \times 1$ zero vector, $I_m$ denotes the $m \times m$ identity matrix, and $0_{m\times r}$ denotes the $m \times r$ zero matrix. Further, for square matrices $A$ and $B$, we write $A \succeq B$ when $A - B$ is positive semidefinite. For the following analysis, we assume that the appropriate regularity conditions are satisfied so that standard GMM results, such as consistency and asymptotic normality, apply. Moreover, all GMM estimators we consider are based on the optimal GMM weighting matrix, that is the inverse of the asymptotic covariance matrix of the moment conditions. Hence, we only consider asymptotically efficient GMM estimators given a set of moment conditions. Define the $n \times k$ matrix

$$D = E\left[\frac{\partial g}{\partial \theta'}(x_t;\theta)\right] \tag{6}$$

and assume $D$ is of full rank equal to $k$. This condition ensures identification of the parameter $\theta$. The time series $\{g(x_t;\theta)\}$ is assumed to satisfy a central limit theorem so that

$$\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T} g(x_t;\theta) \xrightarrow{d} N(0_n, S) \text{ with } S = \sum\nolimits_{k=-\infty}^{\infty} E[g(x_t;\theta)g(x_{t-k};\theta)']. \tag{7}$$

The asymptotically efficient GMM estimate of $\theta$ based on the moment condition (5) is denoted by $\theta_T$. Under appropriate regularity conditions, $\theta_T$ is consistent and asymptotically normal with

$$\sqrt{T}(\theta_T - \theta) \xrightarrow{d} N(0_k, \left(D'S^{-1}D\right)^{-1}).$$

Suppose the parameter $\theta$ is partitioned as $\theta = [\ \theta_1' \quad \theta_2'\ ]'$, with $\theta_i \in \mathbb{R}^{k_i}$, $i = 1, 2$ such that $k_1 + k_2 = k$. Consider a scenario under which the parameter $\theta_1$ is of primary interest while $\theta_2$ is a nuisance parameter in the sense that, although we are not interested in $\theta_2$, we cannot obtain an estimate of $\theta_1$ unless we jointly estimate $\theta_1$ and $\theta_2$. If $\theta_2$ is assumed to be known we can use the known and true value to estimate $\theta_1$.

Suppose that $g(x_t; \theta)$ in the moment condition in (5) can be divided in two parts as follows

$$g(x_t; \theta) = \begin{bmatrix} g_1(x_t; \theta) \\ g_2(x_t; \theta) \end{bmatrix}$$

where the dimension of $g_i(x_t; \theta)$ equals $n_i$, $i = 1, 2$ with $n_i \geq k_i$ so that $n_1 + n_2 = n$. If we assume that the true value of $\theta_2$ is known, we can estimate $\theta_1$ using the moment condition

$$E[g_1(x_t; \theta)] = 0_{n_1}. \tag{8}$$

Consider the following decompositions of the matrices $D$ and $S$:

$$D = [\ D_1 \quad D_2\ ] = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$$

where

$$D_i = \begin{bmatrix} D_{1i} \\ D_{2i} \end{bmatrix},\ D_{ij} = E\left[\frac{\partial g_i}{\partial \theta_j'}(x_t; \theta)\right],\ i, j = 1, 2$$

and

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix},\ S_{ij} = \sum_{k=-\infty}^{\infty} E[g_i(x_t; \theta) g_j(x_{t-k}; \theta)'].$$

We assume that the matrix $D_{11}$ is of full rank equal to $k_1$ so that $\theta_1$ is identified by the moment condition (8) when $\theta_2$ is known. Denote by $\tilde{\theta}_{1,T}$ the estimator of $\theta_1$ obtained from the moment condition (8) using the true value of $\theta_2$. Also denote $\theta_T = (\theta_{1,T}', \theta_{2,T}')'$; that is, $\theta_{1,T}$ is the estimator of $\theta_1$ obtained from joint estimation based on the moment condition (5). The theoretical question we would like to answer is whether knowledge of the true value $\theta_2$ can improve the asymptotic efficiency of the estimator of $\theta_1$.

Our first result derives a common lower bound for the asymptotic covariance matrices $\tilde{V}_1$ and $V_1$ of the two estimators $\tilde{\theta}_{1,T}$ and $\theta_{1,T}$ respectively. The lower bound is achievable under certain conditions which we identify. When either of these conditions is satisfied, a ranking of the two estimators in terms of asymptotic efficiency obtains. The following two conditions turn out to be pivotal for the question we seek to address in this paper.

**Condition A**. The following equality holds: $D_{21} = S_{21}S_{11}^{-1}D_{11}$.

**Condition B**. The following equality holds: $D_2' S^{-1} D_1 = 0_{k_2 \times k_1}$.

First, we obtain the lower bound for $\tilde{V}_1$. It follows from standard GMM results that

$$\tilde{V}_1 = \left(D_{11}' S_{11}^{-1} D_{11}\right)^{-1}. \tag{9}$$

Define the $n_1 \times n$ auxiliary matrix

$$J = \left[\begin{array}{cc} I_{n_1} & 0_{n_1 \times n_2} \end{array}\right] \tag{10}$$

and note that $JD_1 = D_{11}$ and $JSJ' = S_{11}$. It follows that

$$\left(S^{-1}D_1 - J'S_{11}^{-1}D_{11}\right)' S \left(S^{-1}D_1 - J'S_{11}^{-1}D_{11}\right) = D_1'S^{-1}D_1 - D_{11}'S_{11}^{-1}D_{11}$$

which in turn implies

$$D_{11}'S_{11}^{-1}D_{11} = D_1'S^{-1}D_1 - \left(S^{-1}D_1 - J'S_{11}^{-1}D_{11}\right)' S \left(S^{-1}D_1 - J'S_{11}^{-1}D_{11}\right). \tag{11}$$

Hence $D_1'S^{-1}D_1 \succeq D_{11}'S_{11}^{-1}D_{11}$ since $S$ is positive definite. Applying Corollary 7.7.4 in Horn and Johnson (1990) we obtain $\tilde{V}_1 = \left(D_{11}'S_{11}^{-1}D_{11}\right)^{-1} \succeq \left(D_1'S^{-1}D_1\right)^{-1}$. Moreover, due to positive definiteness of $S$, it follows from (11) that $\left(D_{11}'S_{11}^{-1}D_{11}\right)^{-1} = \left(D_1'S^{-1}D_1\right)^{-1}$ is equivalent to $S^{-1}D_1 - J'S_{11}^{-1}D_{11} = 0_{n \times n_1}$. Simple algebra shows that the last equation is equivalent to Condition A. Thus, we have established the following lemma.

**Lemma 1** *Denote by $\tilde{V}_1$ the asymptotic covariance matrix of the estimate $\tilde{\theta}_{1,T}$ obtained when we use the known (true) value of $\theta_2$. Then*

$$\tilde{V}_1 \succeq \left(D_1'S^{-1}D_1\right)^{-1}$$

*and the equality $\tilde{V}_1 = \left(D_1'S^{-1}D_1\right)^{-1}$ holds if and only if Condition A is satisfied.*

The lower bound for the asymptotic covariance $V_1$ is obtained as follows. First, note that $V_1$ equals the $k_1 \times k_1$ upper-left block of the matrix $(D'S^{-1}D)^{-1}$. Since

$$D'S^{-1}D = \begin{bmatrix} D_1' \\ D_2' \end{bmatrix} S^{-1} \begin{bmatrix} D_1 & D_2 \end{bmatrix} = \begin{bmatrix} D_1'S^{-1}D_1 & D_1'S^{-1}D_2 \\ D_2'S^{-1}D_1 & D_2'S^{-1}D_2 \end{bmatrix}$$

we obtain from the formula for the inverse of a partitioned matrix that

$$V_1 = ((D_1'S^{-1}D_1) - (D_2'S^{-1}D_1)'(D_2'S^{-1}D_2)^{-1}(D_2'S^{-1}D_1))^{-1}. \tag{12}$$

The matrix $(D_2'S^{-1}D_2)^{-1}$ is positive definite since $D_2$ has full rank. Hence, equation (12) implies $D_1'S^{-1}D_1 \succeq V_1^{-1}$ which is equivalent to $V_1 \succeq (D_1'S^{-1}D_1)^{-1}$. Furthermore, the equality $V_1 = (D_1'S^{-1}D_1)^{-1}$ is equivalent to

$$(D_2'S^{-1}D_1)'(D_2'S^{-1}D_2)^{-1}(D_2'S^{-1}D_1) = 0_{k_1 \times k_1}$$

which, due to the positive definiteness of $(D_2'S^{-1}D_2)^{-1}$, is equivalent to Condition B. The following lemma states the obtained result.

**Lemma 2** *Denote by $V_1$ the asymptotic covariance matrix of the estimate $\theta_{1,T}$ obtained when we assume that $\theta_2$ is unknown. Then*

$$V_1 \succeq \left(D_1'S^{-1}D_1\right)^{-1}$$

*and the equality $V_1 = \left(D_1'S^{-1}D_1\right)^{-1}$ holds if and only if Condition B is satisfied.*

Combining Lemmas 1 and 2 yields the following theorem which is our first result towards answering the question raised in this paper.

**Theorem 3** *Under Condition A, the estimator $\tilde{\theta}_{1,T}$ that uses the known value of $\theta_2$ is asymptotically more efficient than the estimator $\theta_{1,T}$ that treats $\theta_2$ as unknown:*

$$avar[\theta_{1,T}] \succeq avar[\tilde{\theta}_{1,T}].$$

*Under Condition B, $\theta_{1,T}$ is asymptotically more efficient that the estimator $\tilde{\theta}_{1,T}$:*

$$avar[\tilde{\theta}_{1,T}] \succeq avar[\theta_{1,T}].$$

Although useful, the above theorem is not sharp enough to always provide an answer to the main question posed in the paper. There exist cases in which the lower bound in Lemmas 1 and 2 is not achieved and therefore we cannot compare the two estimators. The following theorem provides a matrix that serves as a "sufficient statistic" in the sense that knowing whether this matrix is positive (negative) semidefinite allows one to rank the two estimators in terms of asymptotic efficiency.

**Theorem 4** *Define the $k_1 \times k_1$ matrix*

$$F = D_1' S^{-1} \left[ D_2 (D_2' S^{-1} D_2)^{-1} D_2' - G \right] S^{-1} D_1 \tag{13}$$

*where*

$$G = \begin{bmatrix} 0_{n_1 \times n_1} & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & S_{22} - S_{21} S_{11}^{-1} S_{12} \end{bmatrix}. \tag{14}$$

*Then, the following classification holds:*

$$avar[\theta_{1,T}] \succeq avar[\tilde{\theta}_{1,T}] \Leftrightarrow F \textit{ is positive semidefinite},$$

$$avar[\tilde{\theta}_{1,T}] \succeq avar[\theta_{1,T}] \Leftrightarrow -F \textit{ is positive semidefinite},$$

$$avar[\tilde{\theta}_{1,T}] = avar[\theta_{1,T}] \Leftrightarrow F = 0_{n_1 \times n_1}.$$

A comment is in order on how Conditions A and B relate to the positive definiteness of $F$ and $-F$ and the ranking of the two estimators $\theta_{1,T}$ and $\tilde{\theta}_{1,T}$. When Condition A holds we have

$$GS^{-1}D_1 = \begin{bmatrix} 0_{n_1 \times n_1} & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & (S^{22})^{-1} \end{bmatrix} \begin{bmatrix} S^{11} & S^{12} \\ S^{21} & S^{22} \end{bmatrix} \begin{bmatrix} D_{11} \\ D_{21} \end{bmatrix} = 0_{n \times n_1}$$

since $(S^{22})^{-1} S^{21} = -S_{21} S_{11}^{-1}$. Therefore, under Condition A,

$$F = D_1' S^{-1} D_2 (D_2' S^{-1} D_2)^{-1} D_2' S^{-1} D_1$$

which is obviously positive semidefinite and so $avar[\theta_{1,T}] \succeq avar[\tilde{\theta}_{1,T}]$. When Condition B holds, we have

$$F = -D_1' S^{-1} G S^{-1} D_1$$

and so $-F$ is positive semidefinite since $G$ is positive semidefinite. Thus, under Condition B, it follows that $avar[\tilde{\theta}_{1,T}] \succeq avar[\theta_{1,T}]$.

# 4 Examples

In the section we provide three examples to illustrate the several possible cases. The first example represents what can be thought of as the regular situation in which the paradox does not occur. In the second example, we have a case in which the two estimators are asymptotically equivalent in terms of efficiency and knowledge of the true value of the nuisance parameter does not enhance the efficiency of the estimator of the parameter of interest. This example is interesting in the sense that neither of the bound obtained in Lemmas 1 and 2 is achieved and therefore one has to resort to Theorem 4 for an answer. The third example is the asset pricing example used in section 2 to motivate the paper. This is the most intriguing example since it provides a model for which either estimator can be more efficient than the other depending on the underlying parameters.

**Example 1:** Consider the standard linear regression model with two regressors

$$y_t = \beta z_t + \gamma w_t + \varepsilon_t$$

where

$$E[\varepsilon_t z_t] = E[\varepsilon_t w_t] = 0.$$

We assume that $(z_t, w_t, \varepsilon_t)'$ are i.i.d. and that the disturbances are homoscedastic with $E[\varepsilon_t | z_t, w_t] = \sigma^2$. We consider $\beta$ to be the parameter of primary interest while $\gamma$ is a nuisance parameter. Let $\theta = (\beta, \gamma)'$ and $x_t = (y_t, z_t, w_t)'$. Using the notation developed in the previous section, we write the moment condition as

$$E[g(x_t;\theta)] = 0_2 \text{ with } g(x_t;\theta) = \begin{bmatrix} (y_t - \beta z_t - \gamma w_t) z_t \\ (y_t - \beta z_t - \gamma w_t) w_t \end{bmatrix}.$$

Simple algebra yields

$$D = -\begin{bmatrix} E[z_t^2] & E[z_t w_t] \\ E[z_t w_t] & E[w_t^2] \end{bmatrix} \text{ and } S = -\sigma^2 D.$$

Hence, Condition A is always satisfied since

$$D_{21} = -E[z_t w_t] = S_{21} S_{11}^{-1} D_{11}$$

while Condition B is satisfied only if $E[z_t w_t] = 0$ since

$$D_2' S^{-1} D_1 = \frac{E[z_t w_t]}{\sigma^2}.$$

Therefore, when $E[z_t w_t] \neq 0$, we can estimate $\beta$ more efficiently if we use the known value of $\gamma$. If $E[z_t w_t] = 0$ the two estimators are asymptotically equivalent.

**Example 2:** Consider the heteroscedastic regression model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

with

$$E[\varepsilon_t] = E[x_t] = E[x_t \varepsilon_t] = 0, \text{ and } E[\varepsilon_t^2 | x_t] = v(x_t).$$

We assume that $(x_t, \varepsilon_t)'$ are i.i.d. and denote

$$v_x = E[x_t^2], \ v_\varepsilon = E[\varepsilon_t^2] \text{ and } \tau_k = E[x_t^k v(x_t)], \ k = 0, 1, 2.$$

We consider $\beta$ to be the parameter of interest and $\alpha$ to be the nuisance parameter. The moment conditions used to estimate $\theta = (\beta, \alpha)'$ are

$$E[g(x_t; \theta)] = 0_2 \text{ where } g(x_t; \theta) = \begin{bmatrix} (y_t - \alpha - \beta x_t) x_t \\ y_t - \alpha - \beta x_t \end{bmatrix}.$$

Moreover, we have

$$D_1 = -\begin{bmatrix} v_x \\ 0 \end{bmatrix}, \ D_2 = -\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \ S = \begin{bmatrix} \tau_2 & \tau_1 \\ \tau_1 & \tau_0 \end{bmatrix}.$$

Hence,

$$D_2' S^{-1} D_1 = \frac{-\tau_1 v_x}{\tau_2 \tau_0 - \tau_1^2}, \ D_{21} = 0, \ S_{21} S_{11}^{-1} D_{11} = \frac{-\tau_1 v_x}{\tau_2 \tau_0}$$

implying that Conditions A and B are simultaneously satisfied if and only if $\tau_1 = 0$. Therefore, if $\tau_1 = 0$ then $\beta_T$ and $\tilde{\beta}_T$ are asymptotically equivalent while if $\tau_1 \neq 0$ Conditions A and B provide no conclusion. However, we can provide an answer to whether $\beta_T$ or $\tilde{\beta}_T$ is more asymptotically efficient by computing the matrix $F$ from Theorem 4. Simple algebra yields

$$G = \begin{bmatrix} 0 & 0 \\ 0 & \tau_0 - \frac{\tau_1^2}{\tau_2} \end{bmatrix} \text{ and } D_2 (D_2' S^{-1} D_2)^{-1} D_2 = \left( \frac{\tau_2}{\tau_0 \tau_2 - \tau_1^2} \right)^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} = G.$$

Therefore, we obtain $F = 0_{2\times 2}$ and so the two estimators are always asymptotically equivalent.

**Example 3:** We now revisit the asset pricing factor model introduced in section 2. The goal is to compare the relative efficiency of the estimators $b_T$, which is based on the moment condition (3), and $\tilde{b}_T$, which uses the known and true value of $\mu_f$ and is based on the moment condition (1). The associated time-series regression of returns on factors

$$R_t = \mu_R + B(f_t - \mu_f) + \varepsilon_t, \text{ with } E[\varepsilon_t] = 0_n, \ E[\varepsilon_t f_t'] = 0_{n\times k} \tag{15}$$

will be used in the sequel. To reduce notational complexity we make the following simplifying assumption.

**Assumption I** The following conditions hold: (i) $(f_t, \varepsilon_t')'$ are i.i.d., (ii) $f_t$ is symmetric around its mean $\mu_f$, and (iii) $E[\varepsilon_t | f_t] = 0_n$.

The asymptotic covariance of $g(x_t;\theta)$ is the $(n+k)\times(n+k)$ matrix

$$S = E\left[g(x_t;\theta)g(x_t;\theta)'\right] = \Sigma_g$$

which can be decomposed in a natural fashion as

$$S = \Sigma_g = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} \Sigma_h & \Sigma_{hf} \\ \Sigma_{fh} & \Sigma_f \end{bmatrix} \tag{16}$$

where

$$h_t = R_t - R_t(f_t - \mu_f)'b = g_1(x_t;\theta)$$

and

$$\Sigma_h = E[h_t h_t'], \ \Sigma_{hf} = E[h_t f_t'] = \Sigma_{fh}'.$$

In addition the $(n+k)\times 2k$ matrix $D$ is given by

$$D = E\left[\frac{\partial g(x_t;\theta)}{\partial \theta'}\right] = \begin{bmatrix} -\Sigma_{Rf} & \mu_R b' \\ 0_{k\times k} & -I_k \end{bmatrix}. \tag{17}$$

Observe that under Assumption I and using the time-series regression (15) we obtain

$$\begin{aligned} \Sigma_{hf} &= E[h_t(f_t - \mu_f)'] \\ &= E[(\mu_R + B(f_t - \mu_f) + \varepsilon_t)(1 - (f_t - \mu_f)'b)(f_t - \mu_f)'] \\ &= -\mu_R b'\Sigma_f + B\Sigma_f \end{aligned}$$

$\infty$and so

$$\Sigma_{hf} = \Sigma_{Rf}(I_k - bb'\Sigma_f). \tag{18}$$

Recall that the asymptotic covariance matrix of $\tilde{b}_T$ equals $\tilde{V}_b = (D_{11}' S_{11}^{-1} D_{11})^{-1}$ while the asymptotic covariance matrix of $b_T$ equals the upper-left $k \times k$ sub-matrix of $(D'S^{-1}D)^{-1}$.

It follows from (16) and the formula for the inverse of a partitioned matrix that

$$S^{-1} = \Sigma_g^{-1} = \begin{bmatrix} \Sigma^h & \Sigma^{hf} \\ \Sigma^{fh} & \Sigma^f \end{bmatrix} \tag{19}$$

where

$$\Sigma^h = (\Sigma_h - \Sigma_{hf}\Sigma_f^{-1}\Sigma_{fh})^{-1} = \Sigma_h^{-1} + \Sigma_h^{-1}\Sigma_{hf}\Sigma^f\Sigma_{fh}\Sigma_h^{-1}, \tag{20}$$

$$\Sigma^f = (\Sigma_f - \Sigma_{fh}\Sigma_h^{-1}\Sigma_{hf})^{-1}, \tag{21}$$

$$\Sigma^{hf} = -\Sigma_h^{-1}\Sigma_{hf}(\Sigma_f - \Sigma_{fh}\Sigma_h^{-1}\Sigma_{hf})^{-1} = -\Sigma^h\Sigma_{hf}\Sigma_f^{-1}, \tag{22}$$

$$\Sigma^{fh} = -\Sigma_f^{-1}\Sigma_{fh}(\Sigma_h - \Sigma_{hf}\Sigma_f^{-1}\Sigma_{fh})^{-1} = -\Sigma^f\Sigma_{fh}\Sigma_h^{-1}. \tag{23}$$

Next, we examine whether Conditions A and B are satisfied. Note that

$$D_1 = \begin{bmatrix} -\Sigma_{Rf} \\ 0_{k\times k} \end{bmatrix}, \; D_2 = \begin{bmatrix} \mu_R b' \\ -I_k \end{bmatrix}$$

and so using (19) we obtain

$$D_2' S^{-1} D_1 = -(b\mu_R'\Sigma^h - \Sigma^{fh})\Sigma_{Rf}.$$

Combining expressions (23), (20) and (18) yields

$$\Sigma^{fh} = -\Sigma_f^{-1}\Sigma_{fh}\Sigma^h = -\Sigma_f^{-1}(I_k - \Sigma_f bb')\Sigma_{Rf}'\Sigma^h = -\Sigma_f^{-1}\Sigma_{Rf}'\Sigma^h + b\mu_R'\Sigma^h$$

and so

$$D_2' S^{-1} D_1 = \Sigma_f^{-1}(\Sigma_{Rf}'\Sigma^h\Sigma_{Rf}).$$

Since both $\Sigma_f^{-1}$ and $\Sigma_{Rf}'\Sigma^h\Sigma_{Rf}$ are invertible we cannot have $D_2'S^{-1}D_1 = 0_{k\times k}$. Hence, Condition B is not satisfied. Furthermore, employing (18) once again, we obtain

$$S_{21}S_{11}^{-1}D_{11} = \Sigma_{fh}\Sigma_h^{-1}\Sigma_{Rf} = (I_k - bb'\Sigma_f)\Sigma_{Rf}'\Sigma_h^{-1}\Sigma_{Rf}.$$

Also note that $D_{21} = 0_{k\times k}$. Hence, since $\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf}$ is invertible, Condition A holds if and only if $I_k - bb'\Sigma_f = 0_{k\times k} \Leftrightarrow \Sigma_f^{-1} = bb'$. But this is impossible since $\Sigma_f^{-1}$ has rank equal to $k$ while $bb'$ has rank equal to 1. Therefore, Condition A cannot be satisfied either.

We conclude that the lower bound obtained in Lemmas 1 and 2 does not help in determining which estimator is more efficient. However, we can directly compare the asymptotic covariance matrices of the two estimators by computing their difference. Since this derivation is no more complicated than the derivation of the matrix $F$, it seems preferable to present an explicit expression for the difference between $V_b$ and $\tilde{V}_b$. The result is captured in the following proposition.

**Proposition 5** *The difference between the asymptotic covariance matrices, $V_b$ and $\tilde{V}_b$, of the GMM estimators $b_T$ and $\tilde{b}_T$ respectively, is given by*

$$V_b - \tilde{V}_b = (2 - b'\Sigma_f b)\, bb'.$$

*Hence, the estimator $\tilde{b}_T$ that uses the known (true) value of $\mu_f$ is asymptotically more (less) efficient than the estimator $b_T$ that does not use the known (true) value of $\mu_f$ if and only if $b'\Sigma_f b \leq 2$ ($b'\Sigma_f b \geq 2$).*

# 5 Conclusion

This paper seeks to answer the following theoretical question in the context of GMM estimation when nuisance parameters are present: When is it the case that using the known value of a nuisance parameter will improve the asymptotic efficiency of the estimator of the parameter of interest? Perhaps contrary to one's intuition, it is indeed possible that incorporating the known value of a nuisance parameter leads to some efficiency loss in the estimation of the parameter of interest. We refer to such a situation as a paradox. We provide necessary and sufficient conditions for the paradox to occur. Concrete examples are presented to illustrate cases for which the paradox does or does not occur. An intriguing example of an asset pricing factor model shows that the risk premia estimator that incorporates the known value of the factor mean could be more or less asymptotically efficient than the estimator that treats the factor mean as unknown, depending on the values of the underlying parameters.

# 6 Appendix

*Proof of Theorem 4*: By Corollary 7.7.4 in Horn and Johnson (1990) we have $V_1 \succeq \tilde{V}_1 \Leftrightarrow \tilde{V}_1^{-1} \succeq V_1^{-1}$. Hence, using expressions (12), (9), and (11) we obtain that $V_1 \succeq \tilde{V}_1$ is equivalent to

$$(D_2' S^{-1} D_1)'(D_2' S^{-1} D_2)^{-1}(D_2' S^{-1} D_1) \succeq \left(S^{-1} D_1 - J' S_{11}^{-1} D_{11}\right)' S \left(S^{-1} D_1 - J' S_{11}^{-1} D_{11}\right)$$

where $J$ is defined by (10). Next, we simplify the right hand side in the above inequality. First, note that

$$S^{-1} D_1 - J' S_{11}^{-1} D_{11} = S^{-1}(D_1 - S J' S_{11}^{-1} D_{11}) = S^{-1} Q_1$$

where

$$Q_1 = \begin{bmatrix} 0_{n_1 \times k_1} \\ D_{21} - S_{21} S_{11}^{-1} D_{11} \end{bmatrix}.$$

Hence,

$$\begin{aligned} & \left(S^{-1} D_1 - J' S_{11}^{-1} D_{11}\right)' S \left(S^{-1} D_1 - J' S_{11}^{-1} D_{11}\right) \\ = & \; Q_1' S^{-1} Q_1 \\ = & \begin{bmatrix} 0_{n_1 \times k_1} \\ D_{21} - S_{21} S_{11}^{-1} D_{11} \end{bmatrix}' \begin{bmatrix} S^{11} & S^{12} \\ S^{21} & S^{22} \end{bmatrix} \begin{bmatrix} 0_{n_1 \times k_1} \\ D_{21} - S_{21} S_{11}^{-1} D_{11} \end{bmatrix} \\ = & \; (D_{21} - S_{21} S_{11}^{-1} D_{11})' S^{22} (D_{21} - S_{21} S_{11}^{-1} D_{11}). \end{aligned}$$

Moreover, we can write

$$D_{21} - S_{21} S_{11}^{-1} D_{11} = -Q_2 D_1$$

where

$$Q_2 = \begin{bmatrix} S_{21} S_{11}^{-1} & -I_{n_2} \end{bmatrix}.$$

Summarizing, we have

$$\begin{aligned} V_1 & \succeq \tilde{V}_1 \Leftrightarrow \\ D_1' S^{-1} D_2 (D_2' S^{-1} D_2)^{-1} D_2' S^{-1} D_1 & \succeq D_1' Q_2' S^{22} Q_2 D_1 \Leftrightarrow \\ D_1' S^{-1} \left[D_2 (D_2' S^{-1} D_2)^{-1} D_2' - S Q_2' S^{22} Q_2 S\right] S^{-1} D_1 & \succeq 0_{n_1 \times n_1}. \end{aligned}$$

Finally, we observe

$$Q_2 S = \left[ \begin{array}{cc} S_{21} S_{11}^{-1} & -I_{n_2} \end{array} \right] \left[ \begin{array}{cc} S^{11} & S^{12} \\ S^{21} & S^{22} \end{array} \right] = \left[ \begin{array}{cc} 0_{n_2 \times n_1} & -\left(S^{22}\right)^{-1} \end{array} \right]$$

which, in turn, implies

$$S Q_2' S^{22} Q_2 S = \left[ \begin{array}{cc} 0_{n_1 \times n_1} & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & S_{22} - S_{21} S_{11}^{-1} S_{12} \end{array} \right] = G$$

and hence completes the proof. Q.E.D.

*Proof of Proposition 5*: First, recall that $V_b$ is the upper-left submatrix of the covariance matrix $(D' S^{-1} D)^{-1}$. Combining expressions (17) and (19) we obtain

$$D' S^{-1} D = \left[ \begin{array}{cc} U_{11} & U_{12} \\ U_{21} & U_{22} \end{array} \right]$$

where

$$U_{11} = \Sigma_{Rf}' \Sigma^h \Sigma_{Rf}, \tag{24}$$

$$U_{12} = -\Sigma_{Rf}' (\Sigma^h \Sigma_{Rf} b b' - \Sigma^{hf}), \tag{25}$$

$$U_{21} = U_{12}' = -(b b' \Sigma_{Rf}' \Sigma^h - \Sigma^{fh}) \Sigma_{Rf}, \tag{26}$$

$$U_{22} = b b' (\Sigma_{Rf}' \Sigma^h \Sigma_{Rf}) b b' - b b' \Sigma_{Rf}' \Sigma^{hf} - \Sigma^{fh} \Sigma_{Rf} b b' + \Sigma^f. \tag{27}$$

Hence, using the formula for the inverse of a partitioned matrix and the formula for the inverse of a sum of matrices, we obtain that

$$V_b = (U_{11} - U_{12} U_{22}^{-1} U_{21})^{-1} = U_{11}^{-1} + U_{11}^{-1} U_{12} P_1 U_{21} U_{11}^{-1}$$

where

$$P_1 = (U_{22} - U_{21} U_{11}^{-1} U_{12})^{-1}.$$

Combining (22) and (18) yields

$$\Sigma^h \Sigma_{Rf} b b' - \Sigma^{hf} = \Sigma^h (\Sigma_{Rf} b b' + \Sigma_{hf} \Sigma_f^{-1}) = \Sigma^h \Sigma_{Rf} \Sigma_f^{-1}$$

and so from (25) it follows

$$U_{12} = -\Sigma_{Rf}' \Sigma^h \Sigma_{Rf} \Sigma_f^{-1}. \tag{28}$$

Since $D_{11} = -\Sigma_{Rf}$ and $S_{11} = \Sigma_h$ we have $\tilde{V}_b = (D_{11} S_{11}^{-1} D_{11})^{-1} = (\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf})^{-1}$. Therefore, using expressions (24), (28), and (26) we obtain

$$\begin{aligned}
V_b - \tilde{V}_b &= U_{11}^{-1} + U_{11}^{-1} U_{12} P_1 U_{21} U_{11}^{-1} - (\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf})^{-1} \\
&= U_{11}^{-1}\Sigma'_{Rf}\Sigma^h\Sigma_{Rf}U_{11}^{-1} \\
&\quad +U_{11}^{-1}\Sigma'_{Rf}\Sigma^h\Sigma_{Rf}\Sigma_f^{-1}P_1\Sigma_f^{-1}\Sigma'_{Rf}\Sigma^h\Sigma_{Rf}U_{11}^{-1} \\
&\quad -U_{11}^{-1}\Sigma'_{Rf}\Sigma^h\Sigma_{Rf}(\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf})^{-1}\Sigma'_{Rf}\Sigma^h\Sigma_{Rf}U_{11}^{-1}
\end{aligned}$$

or equivalently

$$V_b - \tilde{V}_b = U_{11}^{-1}\Sigma'_{Rf}\Sigma^h P_2 \Sigma^h \Sigma_{Rf} U_{11}^{-1} \tag{29}$$

where

$$P_2 = (\Sigma^h)^{-1} + \Sigma_{Rf}\Sigma_f^{-1}P_1\Sigma_f^{-1}\Sigma'_{Rf} - \Sigma_{Rf}(\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf})^{-1}\Sigma'_{Rf}. \tag{30}$$

Next, we seek to simplify the expression for $P_1$. From (24), (25) and (26) it follows that $U_{12} = -U_{11}bb' + \Sigma'_{Rf}\Sigma^{hf} = U'_{21}$ and so $U_{11}^{-1}U_{12} = -bb' + U_{11}^{-1}\Sigma'_{Rf}\Sigma^{hf}$. This implies

$$\begin{aligned}
U_{21}U_{11}^{-1}U_{12} &= (-bb'U_{11} + \Sigma^{fh}\Sigma_{Rf})(-bb' + U_{11}^{-1}\Sigma'_{Rf}\Sigma^{hf}) \\
&= bb'U_{11}bb' - bb'\Sigma'_{Rf}\Sigma^{hf} - \Sigma^{fh}\Sigma_{Rf}bb' + \Sigma^{fh}\Sigma_{Rf}U_{11}^{-1}\Sigma'_{Rf}\Sigma^{hf}
\end{aligned}$$

and so

$$U_{22} - U_{21}U_{11}^{-1}U_{12} = \Sigma^f - \Sigma^{fh}\Sigma_{Rf}U_{11}^{-1}\Sigma'_{Rf}\Sigma^{hf}.$$

This, in turn, using the formula for the inverse of sum of matrices, yields

$$\begin{aligned}
P_1 &= (\Sigma^f - \Sigma^{fh}\Sigma_{Rf}U_{11}^{-1}\Sigma'_{Rf}\Sigma^{hf})^{-1} \\
&= (\Sigma^f)^{-1} + (\Sigma^f)^{-1}\Sigma^{fh}\Sigma_{Rf}(U_{11} - \Sigma'_{Rf}\Sigma^{hf}(\Sigma^f)^{-1}\Sigma^{fh}\Sigma_{Rf})^{-1}\Sigma'_{Rf}\Sigma^{hf}(\Sigma^f)^{-1}.
\end{aligned}$$

Since $\Sigma^{hf}(\Sigma^f)^{-1} = -\Sigma_h^{-1}\Sigma_{hf}$ we have

$$P_1 = \Sigma_f - \Sigma_{fh}\Sigma_h^{-1}\Sigma_{hf} + \Sigma_{fh}\Sigma_h^{-1}P_3\Sigma_h^{-1}\Sigma_{hf}$$

where

$$\begin{aligned}
P_3 &= \Sigma_{Rf}\left(U_{11} - \Sigma'_{Rf}\Sigma^{hf}(\Sigma^f)^{-1}\Sigma^{fh}\Sigma_{Rf}\right)^{-1}\Sigma'_{Rf} \\
&= \Sigma_{Rf}\left(\Sigma'_{Rf}\Sigma^{h}\Sigma_{Rf} - \Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{hf}\Sigma^f\Sigma_{fh}\Sigma_h^{-1}\Sigma_{Rf}\right)^{-1}\Sigma'_{Rf} \\
&= \Sigma_{Rf}\left(\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf}\right)^{-1}\Sigma'_{Rf}
\end{aligned}$$

where the last equality obtains from (20). Therefore,

$$P_1 = \Sigma_f - \Pi$$

where $\Pi$ is defined by

$$\Pi = \Sigma_{fh}\Sigma_h^{-1}(\Sigma_h - \Sigma_{Rf}\left(\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf}\right)^{-1}\Sigma'_{Rf})\Sigma_h^{-1}\Sigma_{hf}.$$

Using (18) we obtain

$$\begin{aligned}
\Pi &= (I_k - \Sigma_f bb')\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf}(I_k - bb'\Sigma_f) \\
&\quad -(I_k - \Sigma_f bb')\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf}\left(\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf}\right)^{-1}\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf}(I_k - bb'\Sigma_f) \\
&= 0_{k\times k}.
\end{aligned}$$

Hence, $P_1 = \Sigma_f$ and so from (30) it follows

$$P_2 = \Sigma_h - \Sigma_{hf}\Sigma_f^{-1}\Sigma_{fh} - \Sigma_{Rf}(\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf})^{-1}\Sigma'_{Rf} + \Sigma_{Rf}\Sigma_f^{-1}\Sigma'_{Rf}.$$

Expression (18) yields

$$\begin{aligned}
\Sigma_{hf}\Sigma_f^{-1}\Sigma_{fh} &= \Sigma_{Rf}(I_k - bb'\Sigma_f)\Sigma_f^{-1}(I_k - \Sigma_f bb')\Sigma'_{Rf} \\
&= \Sigma_{Rf}(\Sigma_f^{-1} - 2bb' + bb'\Sigma_f bb')\Sigma'_{Rf}
\end{aligned}$$

and so

$$P_2 = \Sigma_h - \Sigma_{Rf}(\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf})^{-1}\Sigma'_{Rf} + \Sigma_{Rf}(2bb' - bb'\Sigma_f bb')\Sigma'_{Rf}$$

or equivalently

$$P_2 = P_4 + (2 - b'\Sigma_f b)\,\Sigma_{Rf}bb'\Sigma'_{Rf} \tag{31}$$

where

$$P_4 = \Sigma_h - \Sigma_{Rf}(\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf})^{-1}\Sigma'_{Rf}. \tag{32}$$

It turns out that

$$\Sigma'_{Rf}\Sigma^h P_4 \Sigma^h \Sigma_{Rf} = 0_{k\times k}. \tag{33}$$

Indeed, using expression (20) to substitute $\Sigma^h$ yields

$$\begin{aligned}
& \Sigma'_{Rf}\Sigma^h P_4 \Sigma^h \Sigma_{Rf} \\
= \; & \Sigma'_{Rf}(\Sigma_h^{-1} + \Sigma_h^{-1}\Sigma_{hf}\Sigma^f\Sigma_{fh}\Sigma_h^{-1})P_4(\Sigma_h^{-1} + \Sigma_h^{-1}\Sigma_{hf}\Sigma^f\Sigma_{fh}\Sigma_h^{-1})\Sigma_{Rf} \\
= \; & \Sigma'_{Rf}\Sigma_h^{-1}P_4\Sigma_h^{-1}\Sigma_{Rf} \\
& +\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{hf}\Sigma^f\Sigma_{fh}\Sigma_h^{-1}P_4\Sigma_h^{-1}\Sigma_{Rf} \\
& +\Sigma'_{Rf}\Sigma_h^{-1}P_4\Sigma_h^{-1}\Sigma_{hf}\Sigma^f\Sigma_{fh}\Sigma_h^{-1}\Sigma_{Rf} \\
& +\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{hf}\Sigma^f\Sigma_{fh}\Sigma_h^{-1}P_4\Sigma_h^{-1}\Sigma_{hf}\Sigma^f\Sigma_{fh}\Sigma_h^{-1}\Sigma_{Rf}.
\end{aligned}$$

Moreover,

$$\Sigma'_{Rf}\Sigma_h^{-1}P_4\Sigma_h^{-1}\Sigma_{Rf} = \Sigma'_{Rf}\Sigma_h^{-1}(\Sigma_h - \Sigma_{Rf}(\Sigma'_{Rf}\Sigma_h^{-1}\Sigma_{Rf})^{-1}\Sigma'_{Rf})\Sigma_h^{-1}\Sigma_{Rf} = 0_{k\times k}$$

and thus (33) follows from (18). Finally, combining the expressions (29), (31), and (33) we obtain

$$V_b - \tilde{V}_b = (2 - b'\Sigma_f b)\, bb'$$

completing the proof. Q.E.D.

## REFERENCES


Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski, 1999, Redundancy of moment conditions, *Journal of Econometrics* 91, 89-111.

Crepon, B., F. Kramarz, and A. Trognon, 1997, Parameters of interest, nuisance parameters and orthogonality conditions: An application to autoregressive error component models, *Journal of Econometrics* 82, 135-156.

Henmi, M., and S. Eguchi, 2004, A paradox concerning nuisance parameters and projected estimating functions, *Biometrika* 91, 929-941.

Horn, R. A., and C.R. Johnson, 1990, *Matrix Analysis*, Cambridge University Press.

Pierce, D.A., 1982, The asymptotic effect of substituting estimators for parameters in certain types of statistics, *Annals of Statistics* 10, 457-458.

Robins, J.M., S.D. Mark, and W.K. Newey, 1992, Estimating exposure effects by modelling the expectation of exposure conditional on confounders, *Biometrics* 48, 479-495.

Robins, J.M., A. Rotnitzky, and L.P. Zhao, 1994, Estimation of regression coefficients when some regressors are not always observed, *Journal of American Statistical Association* 89, 846-866.

Qian, H., 2002, Partial redundancy of moment conditions, *Econometric Theory* 18, 531-539.

Zou, F., and J.P. Fine, 2002, A note on a partial empirical likelihood, *Biometrika* 89, 958-961.