Department of Economics

Working Paper Series

Likelihood Based Estimation in a Panel Setting: Robustness, Redundancy and Validity of Copulas

09-002

Artem Prokhorov
Concordia University

Peter Schmidt
Michigan State University

Department of Economics, 1455 De Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8
Tel 514–848–2424 # 3900 · Fax 514–848–4536 · econ@alcor.concordia.ca · alcor.concordia.ca/~econ/repec

Likelihood Based Estimation in a Panel Setting: Robustness, Redundancy and Validity of Copulas

Artem Prokhorov* Peter Schmidt†

Abstract

This paper considers estimation of likelihood-based models in a panel setting. That is, we have panel data, and for each time period separately we have a correctly specified model that could be estimated by MLE. We want to allow non-independence over time. This paper shows how to improve on the QMLE. It then considers MLE based on joint distributions constructed using copulas. It discusses the efficiency gain from using the true copula, and shows that knowledge of the true copula is redundant only if the variance matrix of the relevant set of moment conditions is singular. It also discusses the question of robustness against misspecification of the copula, and proposes a test of the validity of the copula. GMM methods are argued to be useful analytically, and also for reasons of efficiency if the copula is robust but not correct.

JEL Classification: C13

Keywords: Copula, MLE, QMLE, GMM, panel data

*Department of Economics, Concordia University, Montreal, PQ H3G1M8 Canada; email: artem.prokhorov@concordia.ca

†Department of Economics, Michigan State University, East Lansing, MI 48824 USA; email: schmidtp@msu.edu

1 Introduction

In this paper we suppose that we have panel data, with individuals $i = 1, \ldots, N$ and time periods $t = 1, \ldots, T$. We presume that N is large and T is small, and we will consider asymptotic arguments as $N \to \infty$ with T fixed. In this setting, suppose that we have a correctly specified likelihood-based model (e.g. a logit model, or a parametric duration model, or a stochastic frontier model, etc.) which could be consistently estimated from any of the T cross sections. However, we would like to be able to use the data from the whole panel. So, for example, we would be perfectly comfortable estimating our logit model from a single cross section, but we wish to combine information over time to get a more efficient estimate.

In this paper the motivation for using the whole panel is solely to improve efficiency of estimation, and not to correct bias caused by unobservables. An obvious limitation of the paper is that we cannot accommodate fixed individual effects. The paper is firmly rooted in a likelihood approach, and at out level of generality the so-called incidental parameters problem that fixed effects would cause cannot be addressed. The methods we discuss below (based on copulas) for allowing correlations over time can be viewed as an alternative to the traditional random effects approach, which is obviously cumbersome except under joint normality.

For a single cross section at time t, we would have a log-likelihood of the generic form

$$\ln L = \sum_{i=1}^{N} \ln f(y_{it}, \theta),$$

where y_{it} represents the data for person i at time t, and this is what we would like to extend to the panel. The obvious extension (in principle) is to specify a joint log-likelihood of the generic form

$$\ln L = \sum_{i=1}^{N} \ln h(y_{i1}, \ldots, y_{iT}, \theta).$$

The problem, of course, is that there are many joint (all T time periods) distributions consistent with the marginal (individual time period) distributions. How to go from the marginal distributions to the joint distribution is clear under independence, and more or less standard under normality (where the multivariate normal is the obvious extension of marginal normals), but not in general. So there are issues of whether we wish to specify a joint distribution and how to do so.

In Section 2 of the paper we ask what can be done without specifying a joint distribution. It is well known that the QMLE based on the likelihood that would be correct under independence (over t) is robust to non-independence. Using GMM methods, we suggest an improved QMLE estimator (IQMLE) that dominates the QMLE, and we derive the condition under which the efficiency gain is positive.

Section 3 considers construction of the joint distribution using copulas. We provide a GMM interpretation of the MLE estimator based on the joint distribution: it adds another set of moment conditions, which we call the "copula score," to the moment conditions that the IQMLE uses.

In Section 4, under the assumption that the copula is correct, we ask under what circumstances the MLE based on the joint distribution is more efficient than the IQMLE. This is so when the copula score is not redundant. We show that in this setting redundancy can only occur when the covariance matrix of the full set of moment conditions is singular.

In Section 5 we discuss the question of robustness. In general, if we misspecify the copula we have misspecified the joint distribution and the MLE will be inconsistent. However, we show that it is possible that the copula score has mean zero even if the copula is incorrect, so that the MLE is robust to the misspecification of the copula. In this case the GMM estimator based on the full set of moment conditions dominates the (pseudo) MLE based on the incorrectly specified joint likelihood.

In Section 6 we provide a test of the validity of the copula, in the sense that the copula score has mean zero. The copula is valid if it is correctly specified, but also if it is misspecified but the MLE based on it is robust. Our test is a conditional moment test of whether the copula score moment conditions hold, assuming that the moment conditions based on the marginal scores do hold.

Finally, Section 7 contains our concluding remarks.

2 QMLE and Improved QMLE

In this section we ask what can be done using only the marginal distributions. It is known that the QMLE (which maximizes the likelihood that would arise if there were independence over t) is consistent even if there is dependence over t. We discuss this result briefly, and show how to define an improved QMLE (IQMLE).

In this section, and throughout the paper, we will assume that $(y_{i1}, \ldots, y_{iT})$ is i.i.d. over i, but we do not generally assume independence over t.

The following discussion is textbook material (e.g., Hayashi, 2000, section 8.7). The QMLE is the value of θ that maximizes the quasi-likelihood

$$\ln L^Q = \sum_i \sum_t \ln f(y_{it}, \theta). \tag{1}$$

The expression in (1) is the likelihood if we have independence over i and t. However, the QMLE remains consistent if we have dependence over t. To discuss this point, we define a little notation. We define the score functions

$$\begin{aligned} s_{it}(\theta) &= \nabla_\theta \ln f(y_{it}, \theta), &(2)\\ s_i(\theta) &= \sum_t s_{it}(\theta), &(3) \end{aligned}$$

where "∇_θ" means partial derivative with respect to θ. Then the QMLE $\hat{\theta}$ solves the first-order condition

$$\sum_i s_i(\hat{\theta}) = 0. \tag{4}$$

As such, it is a GMM estimator based on the moment condition

$$\mathbb{E} s_i(\theta_o) = 0. \tag{5}$$

But this condition holds so long as the marginal density $f(y_{it}, \theta)$ is correctly specified, because (subject to the usual regularity conditions for MLE) correct specification of the marginal density $f(y_{it}, \theta)$ implies that $\mathbb{E} s_{it}(\theta_o) = 0$ for all t, and therefore $\mathbb{E} s_i(\theta_o) = \sum_t \mathbb{E} s_{it}(\theta_o) = 0$.

We further define the Hessian

$$H_{it}(\theta) = \nabla_\theta s_{it}(\theta) = \nabla_\theta^2 \ln f(y_{it}, \theta) \tag{6}$$

and correspondingly $H_i(\theta) = \sum_t H_{it}(\theta)$ and $\mathbb{H} = \mathbb{E} H_i(\theta_o)$. Also we define the variance matrix of the score:

$$\mathbb{V} = \mathbb{E} s_i(\theta_o) s_i(\theta_o)'. \tag{7}$$

Then under suitable regularity conditions the asymptotic variance of $\hat{\theta}$ is $(\mathbb{H}\mathbb{V}^{-1}\mathbb{H})^{-1} = \mathbb{H}^{-1}\mathbb{V}\mathbb{H}^{-1}$. (We use the standard terminology that "the asymptotic variance of $\hat{\theta}$ is Σ" means that $\sqrt{N}(\hat{\theta} - \theta_o)$ converges in distribution to $N(0, \Sigma)$.) The "sandwich form" is necessary because $\mathbb{H} = -\mathbb{V}$ under independence over t, but not (in general) otherwise.

We now make an observation which, though very simple, appears to be original. Except under independence, summation is not generally the optimal way to combine the observations.

That is, instead of summing over t, we can stack all T values and let GMM perform the optimal weighting. So instead of $\mathbb{E}s_i(\theta_o) = 0$ as in (5), we use the stacked moment conditions

$$\mathbb{E}s_i^*(\theta_o) = 0, \tag{8}$$

where

$$s_i^*(\theta) = \begin{bmatrix} s_{i1}(\theta) \\ s_{i2}(\theta) \\ \vdots \\ s_{iT}(\theta) \end{bmatrix}. \tag{9}$$

We will call the optimal GMM estimator based on the moment conditions in (8) the improved QMLE (IQMLE).[1] Like the QMLE, it should be consistent so long as the marginal distributions are correctly specified, since $\mathbb{E}s_i^*(\theta_o) = 0$ if $\mathbb{E}s_{it}(\theta_o) = 0$ for all t. Define $\mathbb{H}_* = \mathbb{E}\nabla_\theta s_i^*(\theta_o)$ and $\mathbb{V}_* = \mathbb{E}s_i^*(\theta_o)s_i^*(\theta_o)'$. Then, if $\tilde{\theta}$ is the IQMLE estimator, standard results would indicate that the asymptotic variance of $\tilde{\theta}$ is $(\mathbb{H}_*'\mathbb{V}_*^{-1}\mathbb{H}_*)^{-1}$.

It is obvious from basic principles that the IQMLE estimator is efficient relative to the QMLE estimator (optimal weighting is optimal), and the only remaining question is when they are equally efficient. The following result states the first result formally, and answers the remaining question.

To state the result, we define some additional notation. Let $A = 1_T' \otimes I_p$, where 1_T is a $T \times 1$ vector of ones, and where $p = \dim(\theta)$. The matrix A arises naturally because $s_i = As_i^*$, and correspondingly $\mathbb{H} = A\mathbb{H}_*$, and $\mathbb{V} = A\mathbb{V}_*A'$.

Theorem 1 *(a) The IQMLE estimator is efficient relative to the QMLE estimator. (b) The two estimators are equally efficient if and only if $\mathbb{H}_*$ is in the space spanned by $(\mathbb{V}_*A')$.*

Proof. See the Appendix for all proofs that are not given in the main text. □

The condition in part (b) of Theorem 1 is not very intuitive. However, we can identify two cases when it holds.

Theorem 2 *The QMLE and IQMLE estimators are equally efficient if either of the two following conditions holds.*

[1] Because we deal only in asymptotics here, we will not be explicit about issues like how to estimate the optimal GMM weighting matrix. Nor would it matter if instead of GMM we considered other asymptotically equivalent estimates, such as empirical likelihood, exponential tilting, etc.

(a) y_{it} *is i.i.d. over both* i *and* t.

(b) y_{it} *is identically distributed over* t *and the scores are "equicorrelated", in the sense that* $\mathbb{V}_* = E \otimes V_o$, *where*

$$E = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}, \quad T \times T$$

and V_o *is a positive definite* $p \times p$ *matrix and* $\rho > -\frac{1}{T-1}$.

Condition 2(a) is obvious. Condition 2(b) is less so. However, it could arise in a variety of random effects models (where the random effects error structure generates E) with regressors that are the same for all t (so that V_o would be the variance matrix of the marginal score for all t).

3 Copulas, Likelihoods, and a GMM Interpretation of MLE

In this section we summarize the relationship between copulas and likelihoods, and we consider MLE from a GMM perspective. We show that the difference between the MLE based on the joint distribution and the IQMLE estimator based on the marginal scores lies in a term we call the copula score.

A standard definition is that a copula is a joint distribution whose marginals are uniform on [0,1]. That is, it is the distribution of a set of generally non-independent uniform random variables. The connection between a copula and a joint distribution with specified marginals is given by the following well-known theorem.

Theorem 3 *(Sklar, 1959, p.229-230) Let* H *be an* T*-dimensional distribution function with marginals* $F_1, \ldots, F_T$. *Then there exists a* T*-dimensional copula* C *such that for all* $(y_1, \ldots, y_T)$,

$$H(y_1, \ldots, y_T) = C(F_1(y_1), \ldots, F_T(y_T)). \tag{10}$$

If $F_1, \ldots, F_T$ *are continuous, then* C *is unique. Conversely, if* C *is an* T*-dimensional copula and* $F_1, \ldots, F_T$ *are distribution functions, then the function* H *in (10) is an* T*-dimensional distribution function with marginals* $F_1, \ldots, F_T$.

Thus, a copula is a multivariate distribution function that connects two or more marginal distributions to form the joint distribution. A copula thus completely parameterizes the entire dependence structure between two or more random variables. It is important to note that a given joint distribution function H defines a unique set of marginal distribution functions F_t, $t = 1, \ldots, T$, and a unique copula C, whereas given marginal distributions do not determine a unique joint distribution (and the implied copula).

Most uses of copulas in economics and related fields have used the "converse" part of Sklar's theorem. That is, you have a set of marginal cdf's $F_1, \ldots, F_T$ implied by some model, but you want a joint cdf H. So you pick a copula and it generates a joint cdf consistent with the marginals. Lee (1983) appears to be the earliest application of this approach in econometrics. Copulas have recently received more attention in the finance literature. They are used to model dependence in financial time series (e.g., Patton, 2006; Breymann et al., 2003) and in risk management applications (e.g., Embrechts et al., 2003, 2002). Cherubini et al. (2004) and Bouyé et al. (2000) cover a wide range of copula applications in finance. Use of copulas in other subfields of econometrics still appears rather limited. Smith (2003) incorporates a copula in selectivity models and provides applications to labor supply and duration of hospitalization; Cameron et al. (2004) use a copula to develop a bivariate count data model with an application to the number of doctor visits. Zimmer and Trivedi (2006) use copulas in a selection model with count data. Trivedi and Zimmer (2007) consider benefits of copula-based estimation relative to simulation-based approaches. A series of works (see, e.g., Chen and Fan, 2006) consider semiparametric estimation of copulas when unknown marginals are estimated nonparametrically in the first step.

On the other hand, our interest in copulas stems from the first part of Sklar's theorem. This says that any continuous joint distribution uniquely implies the marginal distributions and the copula (the case of non-unique copulas corresponding to discrete joint distributions is not considered here). Therefore we can examine the properties of the MLE based on the joint distribution of $(y_{i1}, \ldots, y_{iT})$ in terms of its components, the marginal distributions of the y_{it}, and the copula cdf C.

In what follows we will restrict our exposition to the bivariate case, $T = 2$. (Most of the copula literature follows this expositional convention, just to keep the notation under control.) Also, for notational simplicity, we will henceforth dispense with the cross-sectional index i except where needed. Thus (y_1, y_2) will be the data for which we seek a joint distribution.

Suppose that the joint cdf of (y_1, y_2) is $H(y_1, y_2; \theta, \rho)$, the marginal cdf's are $F_1(y_1; \theta)$ and

$F_2(y_2;\theta)$, and the copula cdf is $C(\cdot,\cdot;\rho)$. Note that the parameters of interest are the θ's. The nuisance parameter "ρ" is present only in the copula.[2] Then Sklar's theorem says:

$$H(y_1,y_2;\theta,\rho) = C(F_1(y_1;\theta),F_2(y_2;\theta);\rho). \tag{11}$$

Differentiating with respect to (y_1,y_2), we have the corresponding expression in terms of densities:

$$h(y_1,y_2;\theta,\rho) = c(F_1(y_1;\theta),F_2(y_2;\theta);\rho)\cdot f_1(y_1;\theta)\cdot f_2(y_2;\theta), \tag{12}$$

where $c(\cdot,\cdot;\rho)$ is the "copula density," the density function corresponding to the copula cdf $C(\cdot,\cdot;\rho)$. Finally, taking logs, we obtain

$$\ln h(y_1,y_2;\theta,\rho) = \ln c(F_1(y_1;\theta),F_2(y_2;\theta);\rho) + \ln f_1(y_1;\theta) + \ln f_2(y_2;\theta). \tag{13}$$

Now consider summing this expression over the suppressed index i. The left hand side would be the log of the joint likelihood. The first two terms on the right hand side would be the quasi-log-likelihood. The difference between the joint log-likelihood and the quasi-log-likelihood is the sum of the log copula density terms.

Finally, we consider the score functions (with respect to θ and to ρ) corresponding to (13). It is well known that MLE can be viewed as GMM based on the score function (see Godambe, 1960, 1976). The expected value of the score function for the correctly specified joint log-likelihood is zero at the true value of parameters. Furthermore, if the marginal densities are correctly specified, the same is true for the marginal log-likelihoods. Hence, under classical regularity conditions, the following four sets of moment conditions hold at the true values of the parameters (θ_o,ρ_o):

$$\begin{aligned}
&\mathbb{E}\nabla_\theta \ln f_1(y_1;\theta_o) = 0, && \text{(A)}\\
&\mathbb{E}\nabla_\theta \ln f_2(y_2;\theta_o) = 0, && \text{(B)}\\
&\mathbb{E}\nabla_\theta \ln c(F_1(y_1;\theta_o),F_2(y_2;\theta_o);\rho_o) = 0, && \text{(C)}\\
&\mathbb{E}\nabla_\rho \ln c(F_1(y_1;\theta_o),F_2(y_2;\theta_o);\rho_o) = 0. && \text{(D)}
\end{aligned} \tag{14}$$

We call (A) and (B) the "marginal scores" and (C) and (D) the "copula scores". Note that the GMM problem as stated in (14) is overidentified. If the dimension of θ is $p\times 1$, and the dimension of ρ is $q\times 1$, we have $p+q$ parameters and $3p+q$ moment conditions.

[2]Note that "ρ" need not be a correlation. Most copulas contain some parameters that govern the dependence between the marginals, and "ρ" is a generic notation for those parameters.

We can now consider how the various estimators that we consider relate to each other.

The QMLE is based on a moment condition that equals (A)+(B), the sum of the two marginal scores.

The IQMLE is based on the moment conditions (A)&(B). Here "&" is used to indicate that the set of moment conditions includes (A) *and* (B), which will be weighted as appropriate by the GMM machinery.

If there are no parameters in the copula, (D) does not exist. In this case the MLE is based on (A)+(B)+(C). It is asymptotically equivalent to the GMM estimator based on (A)&(B)&(C), since (assuming the copula is correct) the optimal weighting is indeed summation. The efficiency gain from MLE, as opposed to IQMLE, is due to the additional information in the copula score (C).

If there are parameters (ρ) in the copula, (D) does exist. Now the MLE is based on [(A)+(B)+(C)]&(D), whereas the asymptotically equivalent GMM estimator is based on (A)&(B)&(C)&(D). Once again the difference between these estimators and the IQMLE lies in the copula scores, (C) and (D).

4 Efficiency and Redundancy

In this section, we assume that the joint distribution is correctly specified, so that both the marginal distributions and the copula are correct. Therefore the MLE is efficient. The question we address is under what circumstances the IQMLE is efficient; that is, under what circumstances the MLE is no more efficient than the IQMLE. This is the question of when the copula scores are redundant in the sense of Breusch et al. (1999). Perhaps surprisingly, this turns out to be the case only when the full set of scores (A) - (D) is linearly dependent.

We first prove a lemma that reveals the structure of the variance and derivative matrices of the moment functions in (14). We repeat that correct specification of the copula is assumed.

Lemma 1 *Denote the covariance matrix of the moment functions in (14) by* $\mathbf{V}$*, their expected derivative matrix with respect to* (θ, ρ) *by* $\mathbf{D}$*. Then,*

$$\mathbf{V} = \left[\begin{array}{cc|cc} \mathbf{A} & \mathbf{G} & -\mathbf{G} & \mathbf{0} \\ \mathbf{G}' & \mathbf{B} & -\mathbf{G}' & \mathbf{0} \\ \hline -\mathbf{G}' & -\mathbf{G} & \mathbf{J} & \mathbf{E} \\ \mathbf{0} & \mathbf{0} & \mathbf{E}' & \mathbf{F} \end{array}\right] \tag{15}$$

and

$$\mathbf{D} = \left[\begin{array}{c|c} -\mathbf{A} & \mathbf{0} \\ -\mathbf{B} & \mathbf{0} \\ \hline \mathbf{G}+\mathbf{G}'-\mathbf{J} & -\mathbf{E} \\ -\mathbf{E}' & -\mathbf{F} \end{array}\right], \tag{16}$$

where $\mathbf{A}, \mathbf{B}, \mathbf{E}, \mathbf{F}, \mathbf{G}, \mathbf{J}$ *are matrix-functions of* (θ, ρ) *defined in the Appendix.*

Several interesting observations follow from the Lemma. First, the covariance of the first marginal score with the copula score with respect to θ equals minus the covariance of the first marginal score with the second marginal score. Thus the marginal scores are uncorrelated with the copula score with respect to θ if and only if they are uncorrelated with each other. Second, both marginal scores are uncorrelated with the copula score with respect to ρ. Third, it is easy to see that, if $\mathbf{V}$ is nonsingular, the optimal GMM estimate based on (14) is asymptotically equivalent to the MLE. To see this, note that the optimally weighted GMM estimate based on a set of moment conditions "g" is asymptotically equivalent to the GMM estimate based on the exactly identified set of moment conditions $\mathbf{D}'\mathbf{V}^{-1}g$. Let $\mathbb{I}$ denote the identity matrix. By Lemma 1,

$$\mathbf{D}' = -\begin{bmatrix} \mathbb{I} & \mathbb{I} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} \end{bmatrix} \mathbf{V}$$

and so

$$\mathbf{D}'\mathbf{V}^{-1} = -\begin{bmatrix} \mathbb{I} & \mathbb{I} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} \end{bmatrix} \mathbf{V}\mathbf{V}^{-1} = -\begin{bmatrix} \mathbb{I} & \mathbb{I} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} \end{bmatrix}. \tag{17}$$

Then $\mathbf{D}'\mathbf{V}^{-1}g = 0$, with "$g$" as given in (14), gives the first order conditions for the MLE.

For nonsingular $\mathbf{V}$, the asymptotic variance matrix of the optimal GMM estimator based on (14) is of the familiar form $\mathbb{V}_{\mathrm{GMM}} = (\mathbf{D}'\mathbf{V}^{-1}\mathbf{D})^{-1}$. By Lemma 1, this is identical to the asymptotic variance matrix of the MLE estimator of (θ, ρ)

$$\mathbb{V}_{\mathrm{MLE}} = -\left(\begin{bmatrix} \mathbb{I} & \mathbb{I} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} \end{bmatrix} \mathbf{D}\right)^{-1} = \left(\begin{bmatrix} \mathbb{I} & \mathbb{I} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} \end{bmatrix} \mathbf{V} \begin{bmatrix} \mathbb{I} & \mathbf{0} \\ \mathbb{I} & \mathbf{0} \\ \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbb{I} \end{bmatrix}\right)^{-1}. \tag{18}$$

In contrast to $\mathbb{V}_{\mathrm{GMM}}$, $\mathbb{V}_{\mathrm{MLE}}$ is defined even if $\mathbf{V}$ is singular. In fact the last representation in (18) involves the outer-product-of-the-score form of the information matrix, while the one before the last involves the expected-Hessian form of the information matrix. Both are nonsingular under the usual regularity conditions.

We now return to the question of when the IQMLE is as efficient as the MLE. Denote the asymptotic variance matrix of IQMLE by $\mathbb{V}_{\text{IQMLE}}$. Logically, this is the question of when the copula scores are partially redundant for θ, given the marginal scores.[3] Breusch et al. (1999) developed a very useful toolbox for analyzing redundancy of a set of moment conditions given another set of moment conditions. However, their analysis assumes nonsingular $\mathbf{V}$. For this reason, we do not employ their results here but instead compare $\mathbb{V}_{\text{IQMLE}}$ with the relevant block of $\mathbb{V}_{\text{MLE}}$ directly.

Theorem 4 $\mathbb{V}_{\text{MLE}}$ *for* θ *and* $\mathbb{V}_{\text{IQMLE}}$ *are equal if and only if*

$$\mathbf{J} - \mathbf{V}^{\theta}_{\mathbf{21}}\mathbf{V}^{-\mathbf{1}}_{\mathbf{11}}\mathbf{V}^{\theta}_{\mathbf{12}} - \mathbf{E}\mathbf{F}^{-\mathbf{1}}\mathbf{E}' = \mathbf{0}, \tag{19}$$

where $\mathbf{V}^{\theta}_{\mathbf{21}} = \mathbf{V}^{\theta\,\prime}_{\mathbf{12}} = [-\mathbf{G}' \quad -\mathbf{G}]$ *and* $\mathbf{V}_{\mathbf{11}} = \begin{bmatrix} \mathbf{A} & \mathbf{G} \\ \mathbf{G}' & \mathbf{B} \end{bmatrix}$.

The cumbersome expression in (19) has a simple interpretation in terms of singularity of $\mathbf{V}$. It states that the error in the linear projection of moment condition (C) on moment conditions (A), (B) and (D) is uncorrelated with moment condition (C). More specifically, (19) can be rewritten as follows

$$\mathbb{E}\left\{\left(\nabla_\theta \ln c - \mathbf{\Omega_{21}}\mathbf{\Omega^{-1}_{11}} \begin{bmatrix} \nabla_\theta \ln f_1 \\ \nabla_\theta \ln f_2 \\ \nabla_\rho \ln c \end{bmatrix}\right) \nabla'_\theta \ln c \right\} = \mathbf{0},$$

where

$$\mathbf{\Omega_{21}} = [-\mathbf{G}' \quad -\mathbf{G} \quad \mathbf{E}], \quad \mathbf{\Omega_{11}} = \begin{bmatrix} \mathbf{A} & \mathbf{G} & \mathbf{0} \\ \mathbf{G}' & \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{F} \end{bmatrix}$$

and the arguments of the moment functions have been suppressed for brevity. In other words, (C) has to be a linear combination of (A), (B) and (D) for the copula information to be redundant in terms of asymptotic efficiency of estimation of θ. Thus $\mathbf{V}$ has to be singular.

[3] Suppose that θ is identified by some moment conditions g_1. Now we consider estimation based on g_1 and some additional moment condition g_2. Then g_2 is partially redundant for θ given g_1 if the estimate of θ based on g_1 and g_2 is no more efficient than the estimate of θ based on g_1 only. The word "partially" refers to the possibility that g_2 may increase efficiency of estimation for some parameter other than θ (in our case, nuisance parameters in the copula scores).

To understand why this is an important result, we consider a distinction not made in the Breusch et al. (1999) paper. Consider estimation of the mean of y, $\mathbb{E}y = \mu_o$. Suppose that we have moment conditions

$$\begin{aligned}\mathbb{E}(y - \mu_o) &= 0 \qquad [\text{"}g_1\text{"}] &(20)\\ \mathbb{E}[3(y - \mu_o)] &= 0 \qquad [\text{"}g_2\text{"}] &(21)\end{aligned}$$

Clearly g_2 is redundant given g_1 (or vice versa). This is a case of "numerical redundancy" in that one of the moment conditions is a linear combination of the others. The variance matrix $\mathbf{V}$ is singular.

This can be contrasted with "statistical redundancy" (but not "numerical redundancy") in the next example. Now suppose we have

$$\begin{aligned}\mathbb{E}(y - \mu_o) &= 0 \qquad [\text{"}g_1\text{"}] &(22)\\ \mathbb{E}[(y - \mu_o)^2 - \sigma_o^2] &= 0 \qquad [\text{"}g_2\text{"}] &(23)\end{aligned}$$

where σ_o^2 is known. The matrix $\mathbf{V}$ is not singular. Here g_2 is redundant given g_1 if and only if $\mathbb{E}(y - \mu_o)^3 = 0$.

The point of the distinction is that numerical redundancy is obvious, while statistical (but not numerical) redundancy is subtle. One can be seen by inspection whereas the other requires calculation. According to Theorem 4 and the subsequent discussion, the only way the copula score can be redundant is numerical redundancy. If there are any algebraic terms in the copula score with respect to θ that are not a linear combination of the other scores, they cannot be redundant, and the MLE must be strictly more efficient than the IQMLE.

In some cases the copula may not contain parameters. In this case in (15) and (16) the rows and columns of $\mathbf{D}$ and $\mathbf{V}$ that contain the terms $\mathbf{E}$ and $\mathbf{F}$ do not exist, and (C) is redundant given (A) and (B), so that the IQMLE is efficient, if and only if (C) is a linear combination of (A) and (B). A third possibility is that the copula does contain parameters (ρ) but they are "known" (specified). The following Corollary indicates that the same singularity result holds.

Corollary 1 *If* (C) *is a linear combination of* (A) *and* (B) *with ρ known then*

1. $\mathbf{E} = \mathbf{0}$;
2. $\mathbf{J} - \mathbf{V}_{21}^{\theta}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}^{\theta} = \mathbf{0}$;
3. *IQMLE is efficient.*

We now present five examples that show how the redundancy results can be used in practice.

Example 1 ***Bivariate Normal with common mean.*** *Assume Normal marginal densities with* $\sigma_1^2 = \sigma_2^2 = 1$ *and* $\mu_1 = \mu_2 = \mu$

$$f_1(x_1;\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2}},$$

$$f_2(x_2;\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_2-\mu)^2}{2}}.$$

Let the true joint density be Bivariate Normal, i.e.,

$$h(x_1, x_2;\mu,\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x_1-\mu)^2+(x_2-\mu)^2-2\rho(x_1-\mu)(x_2-\mu)}{2(1-\rho^2)}}.$$

Then, the implied copula is the Normal copula

$$c(F_1(x_1;\mu), F_2(x_2;\mu);\rho) = \frac{1}{\sqrt{1-\rho^2}} e^{-\frac{\rho\left(\rho(x_1-\mu)^2+\rho(x_2-\mu)^2-2(x_1-\mu)(x_2-\mu)\right)}{2(1-\rho^2)}},$$

where ρ *is the copula dependence parameter (Pearson's correlation coefficient).*

The relevant moment conditions are

$$\mathbb{E}\left\{X_1-\mu\right\} = 0 \qquad (A)$$
$$\mathbb{E}\left\{X_2-\mu\right\} = 0 \qquad (B)$$
$$\mathbb{E}\left\{-\frac{((X_1-\mu)+(X_2-\mu))\rho}{\rho+1}\right\} = 0 \qquad (C)$$
$$\mathbb{E}\left\{-\frac{\rho(X_1^2+X_2^2)+\mu(1-\rho)^2(X_1+X_2)-\left(1+\rho^2\right)X_1X_2+\rho(\rho^2-1)-\mu^2(1-\rho)^2}{(\rho-1)^2(\rho+1)^2}\right\} = 0 \qquad (D)$$

Clearly (C) is a linear combination of (A) and (B), so it is a linear combination of (A), (B) and (D). Therefore the IQMLE is efficient.

Example 2 ***T-variate Normal with common mean.*** *This example is an extension of the previous example. It shows how our efficiency results generalize to* $T > 2$*. As above assume Normal marginals with* $\sigma_t^2 = 1$ *and* $\mu_t = \mu$, $t = 1,\ldots,T$*. Let the true joint density be* T*-variate normal*

$$h(\mathbf{x};\mu,\Sigma) = \frac{1}{(2\pi)^{T/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)},$$

where $\mathbf{x} = (x_1, \ldots, x_T)'$ *and* Σ *is the correlation matrix of* $\mathbf{x}$*. Then, the implied copula is Normal:*

$$\begin{aligned} c(F_1(x_1;\mu), \ldots, F_T(x_T;\mu); \Sigma) &= \tfrac{h(\mathbf{x};\mu,\Sigma)}{\prod_{t=1}^{T} f_t(x_t;\mu)} \\ &= \tfrac{1}{|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'(\Sigma^{-1}-\mathbb{I})(\mathbf{x}-\mu)}, \end{aligned}$$

where $\mathbb{I}$ *is the identity matrix of dimension* T*.*

The first $T+1$ *moment conditions are*

$$\begin{aligned} &\mathbb{E}\left\{\mathbf{x}-\mu\right\} = 0 \\ &\mathbb{E}\left\{1_T'(\Sigma^{-1}-\mathbb{I})\left(\mathbf{x}-\mu\right)\right\} = 0, \end{aligned}$$

where 1_T *denotes a* $T \times 1$ *vector of ones.*

Here again, the copula score for μ *is a linear combination of the marginal scores. Even without writing out the copula score for* Σ *we can conclude that the IQMLE based on the first* T *moment conditions is efficient for* μ*.*

Example 3 ***Bivariate Normal regression.*** *Let* $\mathbf{y} = \mathbf{x}\beta + \epsilon$*, where* $\mathbf{y} = (y_1, y_2)'$ *is* 2×1 *and* $\mathbf{x} = (x_1, x_2)'$ *is* $2 \times k$*. Suppose* $\mathbf{x}$ *is non-random. Let* $\epsilon = (\epsilon_1, \epsilon_2)' \sim \mathbb{N}(\mathbf{0}, \mathbf{\Sigma})$*, where*

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix}.$$

For simplicity we consider the case that σ_1^2 *and* σ_2^2 *are known. Then,*

$$f_1(y_1; x_1, \beta) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(y_1-x_1\beta)^2}{2\sigma_1^2}},$$

$$f_2(y_2; x_2, \beta) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(y_2-x_2\beta)^2}{2\sigma_2^2}},$$

$$h(\mathbf{y}; \mathbf{x}, \beta, \rho) = \frac{1}{2\pi\sqrt{|\mathbf{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{x}\beta)'\mathbf{\Sigma}^{-1}(\mathbf{y}-\mathbf{x}\beta)}.$$

Then, the implied copula is Normal,

$$\begin{aligned} c(F_1(y_1; x_1, \beta), F_2(y_2; x_2, \beta); \rho) &= \frac{\sqrt{\sigma_1^2\sigma_2^2}}{\sigma_1^2\sigma_2^2 - \rho^2} e^{-\frac{\epsilon_1}{2}\left(\frac{\epsilon_1\sigma_2^2-\epsilon_2\rho}{\sigma_1^2\sigma_2^2-\rho^2}\right)-\frac{\epsilon_2}{2}\left(\frac{\epsilon_2\sigma_1^2-\epsilon_1\rho}{\sigma_1^2\sigma_2^2-\rho^2}\right)} \\ &\quad \times e^{\frac{1}{2}\left(\frac{\epsilon_1^2}{2\sigma_1^2}+\frac{\epsilon_2^2}{2\sigma_2^2}\right)}, \end{aligned}$$

where $\epsilon_i = y_i - x_i\beta$, $i = 1, 2$.

The relevant moment conditions are

$$\mathbb{E}\left\{\frac{x_1\epsilon_1}{\sigma_1^2}\right\} = 0 \qquad (A)$$
$$\mathbb{E}\left\{\frac{x_2\epsilon_2}{\sigma_2^2}\right\} = 0 \qquad (B)$$
$$\mathbb{E}\left\{-\frac{\rho(\sigma_1^2\sigma_2^2x_1\epsilon_2+\sigma_1^2\sigma_2^2x_2\epsilon_1-\sigma_1^2\rho x_2\epsilon_2-\sigma_2^2\rho x_1\epsilon_1)}{\sigma_1^2\sigma_2^2(\sigma_1^2\sigma_2^2-\rho^2)}\right\} = 0 \qquad (C)$$
$$\mathbb{E}\left\{\frac{\sigma_1^2\sigma_2^2\epsilon_1\epsilon_2+\rho^2\epsilon_1\epsilon_2-\sigma_2^2\rho\epsilon_1^2-\sigma_1^2\rho\epsilon_2^2+\rho\sigma_1^2\sigma_2^2-\rho^3}{(\sigma_1^2\sigma_2^2-\rho^2)^2}\right\} = 0 \qquad (D)$$

Now (C) is not a linear combination of (A), (B) and (D), because it contains terms $x_1\epsilon_2$ *and* $x_2\epsilon_1$ *that are not in (A), (B) or (D). So the copula scores are non-redundant and the MLE is strictly more efficient than the IQMLE. There are two exceptions. The first is the case that* $\rho = 0$*, in which case these terms disappear from the copula. The second is the "common regressors" case that* $x_1 = x_2$*, in which case these terms are equal to terms that do appear in (A) and (B). In either of these two cases the IQMLE is efficient.*

Example 4 ***Bivariate Normal with common variance.*** *Assume Normal marginal densities with* $\sigma_1^2 = \sigma_2^2 = \sigma^2$ *and* $\mu_1 = \mu_2 = 0$*. We want to estimate* σ^2*.*

$$f_1(x_1;\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x_1^2}{2\sigma^2}},$$

$$f_2(x_2;\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x_2^2}{2\sigma^2}}.$$

Again, let the true joint distribution be Bivariate Normal, i.e.,

$$h(x_1,x_2;\sigma,\rho) = \frac{1}{2\pi\sqrt{\sigma^4-\rho^2}}e^{-\frac{x_1^2\sigma^2-2x_1x_2\rho+x_2^2\sigma^2}{2(\sigma^4-\rho^2)}}.$$

Then, the implied copula is Normal,

$$c(F_1(x_1;\sigma),F_2(x_2;\sigma);\rho) = \frac{\sigma^2}{\sqrt{\sigma^4-\rho^2}}e^{-\frac{\rho\left(x_1^2\rho+x_2^2\rho-2\sigma^2x_1x_2\right)}{2(\sigma^4-\rho^2)\sigma^2}}.$$

The relevant moment conditions are

$$\mathbb{E}\left\{\frac{{x_1}^2-\sigma^2}{2\sigma^4}\right\} = 0 \qquad (A)$$
$$\mathbb{E}\left\{\frac{{x_2}^2-\sigma^2}{2\sigma^4}\right\} = 0 \qquad (B)$$
$$\mathbb{E}\left\{-\frac{\left(\left(3\rho\sigma^4-\rho^3\right)(x_1^2+x_2^2)-4\sigma^6x_1x_2-2\sigma^2\rho(\sigma^4-\rho^2)\right)\rho}{2(\sigma^2-\rho)^2(\sigma^2+\rho)^2\sigma^2}\right\} = 0 \qquad (C)$$
$$\mathbb{E}\left\{-\frac{\rho\sigma^2(x_1^2+x_2^2)-(\rho^2+\sigma^4)x_1x_2-\rho(\sigma^4-\rho^2)}{(\sigma^2+\rho)^2(\sigma^2-\rho)^2}\right\} = 0 \qquad (D)$$

In this example (C) is not a linear combination of (A) and (B), so we cannot claim that the IQMLE is efficient with ρ known. However, if ρ is unknown, (C) is a linear combination of (A), (B), and (D), so the variance matrix $\mathbf{V}$ *is singular, and the IQMLE is efficient.*

Example 5 ***Farlie-Gumbel-Morgenstern copula with general marginals.*** *For* $i = 1, 2$ *denote the marginal pdf's and cdf's by*

$$f_i \equiv f_i(x_i; \theta)$$

and

$$F_i \equiv F_i(x_i; \theta) = \int_{-\infty}^{x_i} f_i(z; \theta) dz,$$

respectively.

Assume the FGM copula. Then

$$c(u, v; \rho) = 1 + \rho - 2\rho u - 2\rho v + 4\rho uv.$$

Our moment conditions are now

$$\begin{aligned} \mathbb{E}\left\{\tfrac{1}{f_1}\tfrac{\partial f_1}{\partial \theta}\right\} &= 0, && (A) \\ \mathbb{E}\left\{\tfrac{1}{f_2}\tfrac{\partial f_2}{\partial \theta}\right\} &= 0, && (B) \\ \mathbb{E}\left\{\tfrac{2\rho f_1 + 2\rho f_2 - 4\rho f_1 F_2 - 4\rho F_2 F_1}{1+\rho-2\rho F_1 - 2\rho F_2 + 4\rho F_1 F_2}\right\} &= 0, && (C) \\ \mathbb{E}\left\{\tfrac{1 - F_1 - F_2 + 4F_1F_2}{1+\rho-2\rho F_1 - 2\rho F_2 + 4\rho F_1 F_2}\right\} &= 0. && (D) \end{aligned}$$

In general, (C) is not a linear combination of (A), (B) or (A), (B) and (D). So the copula scores are not redundant in general and IQMLE is generally inefficient.

5 Validity and Robustness of Copulas

5.1 Introduction

In this section, as previously, we assume that the marginal likelihoods are correctly specified and that the parameters of interest are the parameters (θ) that appear in the marginal likelihoods. However, we now consider the possibility that the copula is specified incorrectly, so that the joint likelihood is specified incorrectly.

We will say that a copula is *valid* if the expectation of the copula scores [(C) and (D) in equation (14)] is zero, when evaluated at θ_o and some value of ρ. The true copula is always valid. But an incorrect copula may also be valid. If an incorrect copula is valid, then we will say that it is *robust*.

We will use the terminology "pseudo MLE" (PMLE) to refer to the estimator obtained by maximizing the (incorrect) likelihood based on the incorrect copula.[4] In general we would expect the PMLE to be inconsistent, because the model is misspecified. However, there are exceptions. A trivial example is the QMLE based on the independence copula, which is consistent given correct specification of the marginal likelihoods even when there is non-independence. So the QMLE based on independence is robust against all possible forms of dependence. Cases that the PMLE is robust correspond to cases where the assumed copula is robust, as defined in the previous paragraph.

Whether an incorrect copula is robust depends on the nature of the marginal likelihoods and of the true copula. Specifically, the validity of a copula depends on the nature of the model, since that determines the marginal likelihoods, and a copula that is robust in the context of one model may not be robust in the context of a different model.

We can note a few simple results.

(i) The independence copula is always valid, and therefore it is robust against all true copulas. It is also always redundant. (The copula score is identically zero.)

(ii) If a copula is redundant when it is true, then it is also robust against all true copulas. The reason is that the copula score with respect to θ must be a linear combination of the remaining scores; the marginal scores evaluated at θ_o have mean zero by assumption, and the copula score with respect to ρ has zero mean when evaluated at any value of θ, including θ_o, for some value of ρ.

(iii) If the true copula is redundant, then any other valid copula is also redundant. This is true because if the true copula is redundant, the IQMLE is efficient.

(iv) If the copula is valid, the (P)MLE will be consistent.

In subsection 5.2, we show that a non-redundant robust copula is possible. In subsection 5.3, we discuss the fact that the PMLE is dominated by the optimal GMM estimator.

5.2 An Example of a Non-Redundant Robust Copula

In this subsection we give an example of a non-redundant robust copula.

We consider scalar random variables Y_1 and Y_2, with means μ_1 and μ_2, which we wish to

[4] Another possible terminology is quasi MLE (QMLE) but that is usually reserved for cases where the estimator is consistent.

estimate. Suppose that (Y_1, Y_2) have joint cdf H, marginal cdfs F_1 and F_2, and copula C. The corresponding joint density, marginal densities and copula densities will be h, f_1 and f_2, and c.

Definition 1 (Y_1, Y_2) *is* ***radially symmetric (RS) about*** (μ_1, μ_2) *if*

$$H(\mu_1 + y_1, \mu_2 + y_2) = 1 - F_1(\mu_1 - y_1) - F_2(\mu_2 - y_2) + H(\mu_1 - y_1, \mu_2 - y_2),$$

or, equivalently,

$$h(\mu_1 + y_1, \mu_2 + y_2) = h(\mu_1 - y_1, \mu_2 - y_2), \tag{24}$$

for all (y_1, y_2).

Definition 2 Y_1 *is* ***marginally symmetric about*** μ_1 *if*

$$F_1(\mu_1 + y_1) = 1 - F_1(\mu_1 - y_1),$$

or, equivalently,

$$f_1(\mu_1 + y_1) = f_1(\mu_1 - y_1). \tag{25}$$

Definition 3 C *is* ***radially symmetric*** *if*

$$C(1 - u, 1 - v) = 1 - u - v + C(u, v),$$

or, equivalently,

$$c(1 - v, 1 - u) = c(v, u), \tag{26}$$

for all (u, v) *in* $[0, 1] \times [0, 1]$.

It is well known (e.g., Nelsen, 1999, p.33) that the joint distribution is radially symmetric if and only if the marginal distributions are symmetric and the copula is radially symmetric. Many commonly used distributions are RS. For example, bivariate Normal, bivariate Student-t, bivariate Cauchy and other elliptically contoured distributions are RS. For a discussion of the elliptically contoured family of distributions, see Mardia et al. (1979, Section 2.7.2). With reference to the other commonly-considered families of copulas, the independence, FGM, Normal, Plackett, and Frank families are RS, while the Logistic, Ali-Mikhail-Haq (AMH), Joe, Clayton and Gumbel families are not. Interestingly, Frank (1979) shows that the only Archimedean copula family that satisfies Definition 3 is the Frank family. Joe, AMH, Clayton and Gumbel are all Archimedean copulas that are not RS.

We now have the following result.

Theorem 5 *Suppose that the distribution of* (Y_1, Y_2) *is RS about* $\theta = (\mu_1, \mu_2)$*. Then any RS copula is robust for estimation of* θ*.*

To state the result a bit more explicitly, let k be any RS copula density (we use the notation "k" to distinguish it from the true copula "c".) Then, with $\theta = (\mu_1, \mu_2)$,

$$\mathbb{E}\nabla_\theta \ln k(F_1(\mu_1 + Y_1), F_2(\mu_2 + Y_2), \rho) = 0.$$

This is true for any value of ρ (the nuisance parameters in the assumed copula k).

Therefore, so long as the marginal distributions are correctly chosen and the true joint distribution is RS, the misspecified copula k can be used to consistently estimate θ.

The robustness result given in Theorem 5 does not address the issue of whether a non-redundant robust copula can exist. We now give an example that shows that it can. The example consists of logistic marginals, each of which contains a common location parameter μ which is the parameter of interest, and the FGM copula. Because the FGM copula is RS it satisfies the conditions of the theorem above. We will show that the FGM copula, if correct, is not redundant for this problem.

Example 6 ***Farlie-Gumbel-Morgenstern copula and logistic marginals with common mean.*** *Consider logistic marginals with the common location parameter* μ*. Let the true value of* μ *be zero. For* $i = 1, 2$ *the marginal pdf's and cdf's are, respectively,*

$$f_i(y_i; \mu) = \frac{e^{-y_i+\mu}}{(1 + e^{-y_i+\mu})^2}$$

and

$$F_i(y_i; \mu) = \frac{1}{(1 + e^{-y_i+\mu})}.$$

Suppose the true copula is the FGM copula. The copula pdf is

$$c(u, v; \rho) = 1 + \rho - 2\rho u - 2\rho v + 4\rho uv.$$

Note that the logistic distribution is symmetric about zero and that the FGM copula is RS.

Our moment conditions are now

$$\begin{array}{rr}
\mathbb{E}\frac{1-e^{-y_1+\mu}}{1+e^{-y_1+\mu}} = 0 & (A) \\
\mathbb{E}\frac{1-e^{-y_2+\mu}}{1+e^{-y_2+\mu}} = 0 & (B) \\
\mathbb{E}\left\{\frac{2\rho[(-1+2F_1)f_2+(-1+2F_2)f_1]}{1+\rho-2\rho F_1-2\rho F_2+4\rho F_1F_2}\right\} = 0 & (C) \\
\mathbb{E}\left\{\frac{1-F_1-F_2+4F_1F_2}{1+\rho-2\rho F_1-2\rho F_2+4\rho F_1F_2}\right\} = 0 & (D)
\end{array}$$

Let a, b, c and d denote the moment functions in (A), (B), (C) and (D), respectively. Then, simple algebra shows that

$$c = \frac{\rho d}{2ab}[b(1-a^2) + a(1-b^2)]$$

and

$$d = \frac{ab}{1+\rho ab}.$$

Clearly, (C) is not *a linear combination of (A), (B) or of (A), (B) and (D).*

We conclude that in this problem the FGM copula is non-redundant.

In the above example, the FGM copula is non-redundant when it is true. By Theorem 5, it is robust against any radially symmetric true copula. So, while redundancy of a copula implies robustness, the converse is not true.

5.3 Efficiency and Redundancy with a Misspecified but Robust Copula

We now consider the question of efficient estimation when the assumed copula is misspecified but robust. The main thing we wish to point out is that the PMLE is dominated by the efficient GMM estimator. The reason is that summation is no longer the appropriate weighting of the scores with respect to θ. This is the same logical point as was made in Section 2, and indeed Section 2 is therefore a special case of this Section.

We go back to the set of moment conditions (A) - (D) as in equation (14). However, the copula scores (C) and (D) are based on an incorrect but robust copula.

Lemma 2 *Denote the covariance matrix of the moment functions in (14) by* $\mathbf{C}$*, and their expected derivative matrix by* $\mathbf{D}$*. Then,*

$$\mathbf{V} = \left[\begin{array}{cc|cc} \mathbf{A} & \mathbf{G} & -\mathbf{K} & -\mathbf{P} \\ \mathbf{G}' & \mathbf{B} & -\mathbf{L}' & -\mathbf{Q}' \\ \hline -\mathbf{K}' & -\mathbf{L} & \mathbf{N} & \mathbf{Z} \\ -\mathbf{P}' & -\mathbf{Q} & \mathbf{Z}' & \mathbf{W} \end{array}\right]$$

and

$$\mathbf{D} = \left[\begin{array}{c|c} -\mathbf{A} & \mathbf{0} \\ -\mathbf{B} & \mathbf{0} \\ \hline \mathbf{K}' + \mathbf{L} - \mathbf{M} & -\mathbf{S} \\ -\mathbf{S}' & -\mathbf{T} \end{array}\right],$$

where $\mathbf{A}, \mathbf{B}, \mathbf{G}$ *are as in Lemma 1, and* $\mathbf{K}, \mathbf{L}, \mathbf{M}, \mathbf{N}, \mathbf{P}, \mathbf{Q}, \mathbf{S}, \mathbf{T}, \mathbf{W}, \mathbf{Z}$ *are matrix-functions of* (θ, ρ) *defined in the Appendix.*

These expressions are not nearly as simple as the corresponding expressions in (15) and (16). Less information equalities apply in the present case than in the case that the copula is correctly specified.

Lemma 2 can be used to make the following important observation. The optimal GMM estimator using the moment conditions (14), where the assumed copula is robust but not correct, is not the same as the PML estimator. This is in contrast to the case that the copula is correctly specified, in which case the optimal GMM estimator and the PMLE were asymptotically equivalent. The simple form of the optimal set of linear combinations, $\mathbf{D}'\mathbf{V}^{-1}$, as given in (17) does not hold in the present case. In other words, the optimal weighting now does not correspond to summation of (A), (B) and (C), which is what PMLE does.

We will call the optimal GMM estimator based on (14) the *Improved PML Estimator* (IPMLE). The following Theorem formally states its dominance of the PMLE in terms of efficiency.

Theorem 6 *Let* $\mathbb{V}_{\text{IPMLE}}$ *and* $\mathbb{V}_{\text{PMLE}}$ *denote the asymptotic variance matrices of the IPMLE and PMLE of* (θ_o, ρ_o)*, respectively. Then,* $\mathbb{V}_{\text{PMLE}} - \mathbb{V}_{\text{IPMLE}}$ *is positive semi-definite.*

Proof. Define

$$\mathbb{A} = \begin{bmatrix} \mathbb{I} & \mathbb{I} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} \end{bmatrix}.$$

Then, the moment conditions used by PMLE are those given in (14), pre-multiplied by $\mathbb{A}$. Correspondingly, the variance matrix of these moment functions can be expressed as $\mathbb{A}\mathbf{V}\mathbb{A}'$, and their expected derivative matrix can be expressed as $\mathbb{A}\mathbf{D}$. Then,

$$\mathbb{V}_{\text{PMLE}} = [(\mathbb{A}\mathbf{D})'(\mathbb{A}\mathbf{V}\mathbb{A}')^{-1}(\mathbb{A}\mathbf{D})]^{-1}, \tag{27}$$

while

$$\mathbb{V}_{\text{IPMLE}} = [\mathbf{D}'(\mathbf{V})^{-\mathbf{1}}\mathbf{D}]^{-1}. \tag{28}$$

$\mathbb{V}_{\text{PMLE}} - \mathbb{V}_{\text{IPMLE}}$ is PSD if and only if $\mathbb{V}^{-1}_{\text{IPMLE}} - \mathbb{V}^{-1}_{\text{PMLE}} = \mathbf{D}'\mathbf{V}^{-\mathbf{1}}\mathbf{D} - \mathbf{D}'\mathbb{A}'(\mathbb{A}\mathbf{V}\mathbb{A}')^{-1}\mathbb{A}\mathbf{D}$ is PSD. Rewrite the last expression as

$$\mathbf{D}'(\mathbf{V})^{-\mathbf{1/2}}[\mathbb{I} - (\mathbf{V})^{\mathbf{1/2}}\mathbb{A}'(\mathbb{A}(\mathbf{V})^{\mathbf{1/2}}(\mathbf{V})^{\mathbf{1/2}}\mathbb{A}')^{-1}\mathbb{A}(\mathbf{V})^{\mathbf{1/2}}](\mathbf{V})^{-\mathbf{1/2}}\mathbf{D}.$$

This is PSD because the matrix in brackets is the PSD projection matrix orthogonal to $(\mathbf{V})^{\mathbf{1/2}}\mathbb{A}'$. $\square$

We now revisit the redundancy question that was addressed in Section 4 for correctly specified copulas. The IPMLE estimator must dominate the IQMLE estimator in terms of efficiency, but the question is when the efficiency difference is zero. Since $\mathbb{V}_{\text{IPMLE}}$ is only defined when $\mathbf{V}$ is nonsingular, thus we can apply the redundancy toolbox of Breusch et al. (1999).

In stating the next result, we must distinguish the covariance matrix of the moment conditions based on the incorrect but robust copula (which we will call $\mathbf{V^k}$) from the covariance matrix of the moment conditions based on the true copula (which we will call $\mathbf{V}$). The true copula is not involved in estimation, but it is involved in evaluating expectations because it is part of the true joint distribution.

Theorem 7 $\mathbb{V}_{\text{IPMLE}}$ *for* θ *and* $\mathbb{V}_{\text{IQMLE}}$ *are equal if and only if*

$$\mathbf{M} - \mathbf{V}^{\theta\mathbf{k}}_{\mathbf{21}}\mathbf{V}^{-\mathbf{1}}_{\mathbf{11}}\mathbf{V}^{\theta}_{\mathbf{12}} - \mathbf{S}\mathbf{T}^{-\mathbf{1}}(\mathbf{R} - \mathbf{V}^{\rho\mathbf{k}}_{\mathbf{21}}\mathbf{V}^{-\mathbf{1}}_{\mathbf{11}}\mathbf{V}^{\theta}_{\mathbf{12}}) = \mathbf{0}, \tag{29}$$

where $\mathbf{V}^{\theta\mathbf{k}}_{\mathbf{21}} = [-\mathbf{K}' \quad -\mathbf{L}]$, $\mathbf{V}^{\rho\mathbf{k}}_{\mathbf{21}} = [-\mathbf{P}' \quad -\mathbf{Q}]$, *and* $\mathbf{V}^{\theta}_{12} = \begin{bmatrix} -\mathbf{G} \\ -\mathbf{G}' \end{bmatrix}$.

In (29), $\mathbf{M} - \mathbf{V}^{\theta\mathbf{k}}_{\mathbf{21}}\mathbf{V}^{-\mathbf{1}}_{\mathbf{11}}\mathbf{V}^{\theta}_{\mathbf{12}}$ and $\mathbf{R} - \mathbf{V}^{\rho\mathbf{k}}_{\mathbf{21}}\mathbf{V}^{-\mathbf{1}}_{\mathbf{11}}\mathbf{V}^{\theta}_{\mathbf{12}}$ can be viewed as covariance matrices between copula moments based on the incorrect but robust copula and the error in the linear projection of the true copula moment on the marginal moments (A-B). More explicitly,

$$\begin{aligned} \mathbf{M} - \mathbf{V}^{\theta\mathbf{k}}_{\mathbf{21}}\mathbf{V}^{-\mathbf{1}}_{\mathbf{11}}\mathbf{V}^{\theta}_{\mathbf{12}} &= \mathbb{E}\left\{\nabla_\theta \ln k \left(\nabla_\theta \ln c - \mathbf{V_{21}}\mathbf{V}^{-\mathbf{1}}_{\mathbf{11}} \begin{bmatrix} \nabla_\theta \ln f_1 \\ \nabla_\theta \ln f_2 \end{bmatrix}\right)'\right\}, \\ \mathbf{R} - \mathbf{V}^{\rho\mathbf{k}}_{\mathbf{21}}\mathbf{V}^{-\mathbf{1}}_{\mathbf{11}}\mathbf{V}^{\theta}_{\mathbf{12}} &= \mathbb{E}\left\{\nabla_\rho \ln k \left(\nabla_\theta \ln c - \mathbf{V_{21}}\mathbf{V}^{-\mathbf{1}}_{\mathbf{11}} \begin{bmatrix} \nabla_\theta \ln f_1 \\ \nabla_\theta \ln f_2 \end{bmatrix}\right)'\right\}, \end{aligned} \tag{30}$$

where k is the incorrect copula pdf. Clearly, when both of these matrices are zero, (29) holds for any $\mathbf{S}$. Also, if only (30) is zero and $\mathbf{S} = \mathbf{0}$, (29) holds for any $\mathbf{R}$ and $\mathbf{V}^{\rho\mathbf{k}}_{\mathbf{21}}$.

Corollary 2 *If the true copula score with respect to* θ *is a linear combination of (A) and (B) with* ρ *known then*

1. $\mathbf{M} - \mathbf{V}^{\theta\mathbf{k}}_{\mathbf{21}}\mathbf{V}^{-\mathbf{1}}_{\mathbf{11}}\mathbf{V}^{\theta}_{\mathbf{12}} = \mathbf{0}$;

2. $\mathbf{R} - \mathbf{V}_{21}^{\rho\mathbf{k}}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}^{\theta} = \mathbf{0}$*;*

3. IQMLE and IPMLE for θ are equally efficient.

Basically, Corollary 2 repeats fact (iii) noted in subsection 5.1. If the true copula is redundant, then any other valid copula must also be redundant. What is important to note is that this does *not* imply that the copula score based on the incorrect but robust copula must be a linear combination of the other scores. In other words, with a correctly specified copula, the only way redundancy can occur is when there is linear dependency among the scores. This is not the case when the copula is incorrect but robust.

6 Testing the Validity of the Copula

Suppose that we are willing to assert the correctness of the marginal distributions but we are doubtful about the correctness of the joint distribution. More precisely, we are willing to assert that the marginal moment conditions (A) and (B) in (14) hold, but we are doubtful about the validity of the copula moment conditions (C) and (D). We wish to test the validity of the copula by testing the validity of the copula moment conditions. We will discuss two ways of doing this. It is worth noting that in either case we are testing the *validity* of the copula as opposed to the *correctness* of the copula. The tests we discuss would not distinguish the case of a true copula from the case of an incorrect but robust copula. This should be possible in principle because the truth of the copula imposes a number of "information equalities" that imply restrictions on the matrices $\mathbf{D}$ and $\mathbf{V}$ discussed above, and those restrictions should be testable. We do not pursue that detail here, however.

One test that can be used is the usual test of overidentifying restrictions. The GMM problem (14) is overidentified since we have $3p+q$ moment conditions (where p is the dimension of θ and q is the dimension of ρ) and $p+q$ parameters. Therefore we can test the validity of the full set of moment conditions by the test of overidentifying restrictions (see, e.g., Hansen, 1982; Newey and West, 1987). If we assert that the marginal moment conditions are correct, then this is a test of the validity of the copula moment conditions.

To be explicit, we will need more notation. For $m = 1, 2$ and $i = 1, \ldots, N$, denote

$f_{mi}(\theta) = f_m(y_{mi};\theta)$, $c_i(\theta,\rho) = c(F_1(y_{1i};\theta), F_2(y_{2i};\theta);\rho)$,

$$\psi_i(\theta,\rho) = \begin{bmatrix} \nabla_\theta \ln f_{1i}(\theta) \\ \nabla_\theta \ln f_{2i}(\theta) \\ \nabla_\theta \ln c_i(\theta,\rho) \\ \nabla_\rho \ln c_i(\theta,\rho) \end{bmatrix}, \quad g_i(\theta) = \begin{bmatrix} \nabla_\theta \ln f_{1i}(\theta) \\ \nabla_\theta \ln f_{2i}(\theta) \end{bmatrix},$$

$$r_i(\theta,\rho) = \begin{bmatrix} \nabla_\theta \ln c_i(\theta,\rho) \\ \nabla_\rho \ln c_i(\theta,\rho) \end{bmatrix}.$$

Note that ψ_i is a (3p+q)-vector. Let

$$\bar{\psi}(\theta,\rho) \equiv \frac{1}{N}\sum_{i=1}^{N}\psi_i(\theta,\rho), \quad \bar{g}(\theta) \equiv \frac{1}{N}\sum_{i=1}^{N} g_i(\theta), \quad \bar{r}(\theta,\rho) \equiv \frac{1}{N}\sum_{i=1}^{N} r_i(\theta,\rho).$$

Following our previous notation, let

$$\begin{aligned}
\mathbf{V_o} &\equiv \mathbb{E}\psi(\theta_o,\rho_o)\psi(\theta_o,\rho_o)', \\
\mathbf{V^o_{11}} &\equiv \mathbb{E}g(\theta_o)g(\theta_o)', \\
\mathbf{V^o_{22}} &\equiv \mathbb{E}r(\theta_o,\rho_o)r(\theta_o,\rho_o)', \\
\mathbf{V^o_{12}} = \mathbf{V^o_{21}}' &\equiv \mathbb{E}g(\theta_o)r(\theta_o,\rho_o)', \\
\mathbf{D_o} &\equiv \mathbb{E}\nabla_{(\theta',\rho')'}\psi(\theta_o,\rho_o), \\
\mathbf{D^o_{11}} &\equiv \mathbb{E}\nabla_\theta g(\theta_o), \\
\mathbf{D^o_{21}} &\equiv \mathbb{E}\nabla_\theta r(\theta_o,\rho_o), \\
\mathbf{D^o_{22}} &\equiv \mathbb{E}\nabla_\rho r(\theta_o,\rho_o),
\end{aligned}$$

where expectations are with respect to the true joint density $h(x_1,x_2)$.

Theorem 8 *Let $(\breve{\theta},\breve{\rho})$ denote the optimal GMM estimate of (θ,ρ) based on (14). Then*

$$N\bar{\psi}(\breve{\theta},\breve{\rho})'\mathbf{V_o^{-1}}\bar{\psi}(\breve{\theta},\breve{\rho}) \overset{a}{\sim} \chi^2_{2p}. \tag{31}$$

This test is a specification test which, given that the marginal distributions are correct, should capture copula misspecification (or, more precisely, invalidity of the copula). A consistent estimator of $\mathbf{V_o}$ such as

$$\breve{\mathbf{V}}_\mathbf{o} = \frac{1}{N}\sum_{i=1}^{N}\psi_i(\breve{\theta},\breve{\rho})\psi_i(\breve{\theta},\breve{\rho})'$$

is usually used in (31). It is however important to note that the statistic in (31) can be used only if $\mathbf{V}$ is non-singular, i.e. if copula terms are not redundant.

At an intuitive level, this test is unappealing because it does not focus strongly on the moment conditions we are doubtful about. Also one could object that the number of degrees of freedom does not seem right. We maintain the correctness of the marginal moment conditions. Then the copula scores add $p+q$ moment conditions, but also add q nuisance parameters. So the number of restrictions to test, and therefore the "right" number of degrees of freedom, should be p, not $2p$. We achieve this with the following two-step procedure.

Theorem 9 *Let $\hat{\theta}$ be the optimal GMM estimate based on the marginal moment conditions $\mathbb{E}g(\theta) = 0$. Let $\hat{\rho}$ be obtained by minimizing $\bar{r}(\hat{\theta},\rho)'\mathbf{B_o^{-1}}\bar{r}(\hat{\theta},\rho)$, where*

$$\begin{aligned}\mathbf{B_o} = \mathbf{V^o_{22}} & -\mathbf{D^o_{21}}(\mathbf{D^o_{11}}\mathbf{V^o_{11}}^{-1}\mathbf{D^o_{11}})^{-1}\mathbf{D^o_{11}}'\mathbf{V^o_{11}}^{-1}\mathbf{V^o_{12}} \\ & -\mathbf{V^o_{21}}\mathbf{V^o_{11}}^{-1}\mathbf{D^o_{11}}(\mathbf{D^o_{11}}\mathbf{V^o_{11}}^{-1}\mathbf{D^o_{11}})^{-1}\mathbf{D^o_{21}}' \\ & +\mathbf{D^o_{21}}(\mathbf{D^o_{11}}\mathbf{V^o_{11}}^{-1}\mathbf{D^o_{11}})^{-1}\mathbf{D^o_{21}}'.\end{aligned}$$

Then,

$$N\bar{r}(\hat{\theta},\hat{\rho})'\mathbf{B_o^{-1}}\bar{r}(\hat{\theta},\hat{\rho}) \overset{a}{\sim} \chi^2_p. \tag{32}$$

Similarly to Theorem 8, consistent estimates of the elements of $\mathbf{V_o}$ and $\mathbf{D_o}$ will be used in practice for calculating the test statistic in (32).

Essentially the above test is a conditional moment test of the type discussed by Tauchen (1985) and Wooldridge (1991). Formally it is an extension of those tests because it needs to accommodate the presence of nuisance parameters that appear in the moments that are being tested and not in the moments that are maintained. That extension may be useful in other contexts.

7 Concluding remarks

This paper has discussed likelihood based estimation in a panel setting. The point of view is that we have a likelihood based model, like a logit model, that is correctly specified, but we want to account for non-independence over time.

The existing procedure in the literature is the QMLE. We show how this can be improved and provide conditions under which the improved QMLE (IQMLE) does or does not have a positive efficiency gain. An interesting unanswered question, which we hope to address in future work, is whether this is the best we can do using only the assumption that the marginal distributions are correct. In other words, what is the semiparametric efficiency bound for estimation of θ when the true copula is unknown? This is the converse of the question addressed by Chen and Fan (2006), who consider efficiency bounds when the copula has a known parametric form but the marginals are unknown. Given that we have correctly specified parametric marginals, we can certainly estimate the copula non-parametrically, and then the question is whether using this estimated copula improves efficiency relative to IQMLE. We hope to address this question in future work.

Next we consider MLE based on a joint distribution. Often the joint distribution will be constructed by choosing a copula. The further assumption (copula) that converts the marginal distributions into a joint distribution raises standard questions of efficiency and robustness. We address these questions in a GMM framework. The GMM approach is potentially useful in some practical ways. It leads directly to a way of testing the validity of the copula, and it leads to an improvement over the pseudo-MLE (PMLE) in the case that the copula is misspecified but robust. However, we hope that we have demonstrated also that the GMM approach is useful because it allows us to study questions of efficiency and robustness in a structured way. The MLE offers a positive efficiency gain over the IQMLE if the copula scores are not redundant, and the PMLE or GMM-based IPMLE are consistent if the copula score moment conditions are valid.

Our paper was motivated by a panel data model in which the marginal distributions are the same, and the same parameters appear in all of them. In fact, the mathematics of the paper allows the marginals to be different, but all of our results are established under the assumption that these marginals all depend on the same parameters. How many of our results can be generalized to the case that different marginals have different parameters remains to be seen.

In our view, the most important remaining question is to characterize circumstances in which a copula is robust. We provide a non-trivial example in which an incorrect copula can be robust, but this depends on the nature of the marginal models, the assumed copula and the true copula. Any more general results would be very useful. The Holy Grail would be to find a *magic copula* that is always robust and sometimes non-redundant, or to prove that a magic copula cannot exist.

References

Bouyé, E., V. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli (2000): "Copulas for Finance: A Reading Guide and Some Applications," *Crédit Lyonnais Working Paper*, gro.creditlyonnais.fr/content/wp/copula-survey.pdf.

Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski (1999): "Redundancy of moment conditions," *Journal of Econometrics*, 91, 89–111.

Breymann, W., A. Dias, and P. Embrechts (2003): "Dependence structures for multivariate high-frequency data in finance," *Quantitative Finance*, 3, 1–14, http://www.iop.org/EJ/abstract/1469-7688/3/1/301/.

Cameron, A. C., T. Li, P. K. Trivedi, and D. M. Zimmer (2004): "Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts," *Econometrics Journal*, 7, 566–84.

Chen, X. and Y. Fan (2006): "Estimation of copula-based semiparametric time series models," *Journal of Econometrics*, 130, 307–335.

Cherubini, U., E. Luciano, and W. Vecchiato (2004): *Copula Methods in Finance*, Wiley Finance Series, John Wiley and Sons Ltd.

Embrechts, P., A. Höing, and A. Juri (2003): "Using copulae to bound the Value-at-Risk for functions of dependent risks," *Finance and Stochastics*, 7, 145–167.

Embrechts, P., A. McNeil, and D. Straumann (2002): "Correlation and dependence in risk management: properties and pitfalls," in *Risk Management: Value at Risk and Beyond*, ed. by M. Dempster, Cambridge: Cambridge University Press, 176–223.

Frank, M. (1979): "On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$," *Aequationes Mathematicae*, 19, 194–226.

Godambe, V. P. (1960): "An optimum property of regular maximum likelihood estimation," *The Annals of Mathematical Statistics*, 31, 1208–1211.

——— (1976): "Conditional likelihood and unconditional optimum estimating equations," *Biometrika*, 63, 277–284.

Hansen, L. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–1054.

Hayashi, F. (2000): *Econometrics*, Princeton University Press.

Lee, L.-F. (1983): "Generalized econometric models with selectivity," *Econometrica*, 51, 507–512.

Mardia, K., J. Kent, and J. Bibby (1979): *Multivariate Analysis*, Probability and Mathematical Statistics, London: Academic Press.

Nelsen, R. B. (1999): *An Introduction to Copulas*, vol. 139 of *Lecture Notes in Statistics*, Springer.

——— (2006): *An Introduction to Copulas*, vol. 139 of *Springer Series in Statistics*, Springer, 2 ed.

Newey, W. and K. West (1987): "Hypothesis testing with efficient method of moments estimation," *International Economic Review*, 28, 777–787.

Patton, A. (2006): "Modelling Asymmetric Exchange Rate Dependence," *International Economic Review*, 47, 527–556.

Sklar, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

Smith, M. D. (2003): "Modelling sample selection using Archimedian copulas," *Econometrics Journal*, 6, 99–123.

Tauchen, G. (1985): "Diagnostic testing and evaluation of maximum likelihood models," *Journal of Econometrics*, 30, 415–43.

Trivedi, P. K. and D. M. Zimmer (2007): *Copula Modeling: An Introduction for Practitioners*, vol. 1 of *Foundations and Trends in Econometrics*.

Wooldridge, J. (1991): "Specification Testing and Quasi-Maximum Likelihood Estimation," *Journal of Econometrics*, 48, 29–55.

Zimmer, D. M. and P. K. Trivedi (2006): "Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand," *Journal of Business and Economic Statistics*, 24, 63–76.

8 Appendix: Proofs

Proof of Theorem 1: Let $\mathbb{V}_{\mathrm{QMLE}}$ and $\mathbb{V}_{\mathrm{IQMLE}}$ denote the asymptotic variance matrix of QMLE and IQMLE respectively. Then,

$$\mathbb{V}_{\mathrm{QMLE}} = [(A\mathbb{H}_*)'(A\mathbb{V}_*A')^{-1}(A\mathbb{H}_*)]^{-1}, \tag{33}$$

while

$$\mathbb{V}_{\mathrm{IQMLE}} = [\mathbb{H}_*'\mathbb{V}_*^{-1}\mathbb{H}_*]^{-1}. \tag{34}$$

But $\mathbb{V}_{\mathrm{QMLE}} - \mathbb{V}_{\mathrm{IQMLE}}$ is positive semi-definite (PSD) if and only if $\mathbb{V}_{\mathrm{IQMLE}}^{-1} - \mathbb{V}_{\mathrm{QMLE}}^{-1} = \mathbb{H}_*'\mathbb{V}_*^{-1}\mathbb{H}_* - \mathbb{H}_*'A'(A\mathbb{V}_*A')^{-1}A\mathbb{H}_*$ is PSD. The last expression can be rewritten as $\mathbb{H}_*'\mathbb{V}_*^{-1/2}[I - \mathbb{V}_*^{1/2}A'(A\mathbb{V}_*^{1/2}\mathbb{V}_*^{1/2}A')^{-1}A\mathbb{V}_*^{1/2}]\mathbb{V}_*^{-1/2}\mathbb{H}_*$. This is PSD because the matrix in brackets is the PSD projection matrix orthogonal to $\mathbb{V}_*^{1/2}A'$. The expression is zero (the two estimators are equally efficient) if and only if $\mathbb{V}_*^{-1/2}\mathbb{H}_*$ is in the space spanned by $\mathbb{V}_*^{1/2}A'$, or, equivalently, $\mathbb{H}_*$ is in the space spanned by $\mathbb{V}_*A'$. $\square$

Proof of Theorem 2: (a) If y_{it} is independent over i and t then $\mathbb{V}_* = I_T \otimes V_o$, where $V_o = \mathbb{E}s_{it}(\theta_o)s_{it}(\theta_o)'$, and $\mathbb{H}_* = 1_T \otimes H_o$, where $H_o = \mathbb{E}\nabla'_\theta s_{it}(\theta_o)$. We have $\mathbb{V}_*A' = (I_T \otimes V_o)(1_T \otimes I_p) = 1_T \otimes V_o$. Now, H_o is in the space spanned by V_o since V_o is nonsingular, and so $\mathbb{H}_*$ is in the space spanned by $\mathbb{V}_*A'$. Note that we did not need the "correct distribution" assumption that $H_o = -V_o$.

(b) If scores are equicorrelated, $\mathbb{H}_* = 1_T \otimes H_o$ and $\mathbb{V}_*A' = (1 + \rho(T-1))1_T \otimes V_o$. So $\mathbb{H}_*$ is in the space spanned by $\mathbb{V}_*A'$ so long as V_o is nonsingular. Of course, ρ should be greater than $-\frac{1}{T-1}$ for $\mathbb{V}_{\text{QMLE}}$ (asymptotic variance of the QMLE estimator) to be positive definite. Again, note that no "correct distribution" assumption is used. $\square$

Proof of Theorem 3: See Sklar (1959, p.229-230). $\square$

Proof of Lemma 1: By the information matrix equality (IME),

$$\mathbf{A} \equiv \mathbb{E}\left\{\nabla_\theta \ln f_1(Y_1;\theta)\nabla'_\theta \ln f_1(Y_1;\theta)\right\} = -\mathbb{E}\nabla^2_\theta \ln f_1(Y_1;\theta). \tag{35}$$

Similar for $\mathbf{B}, \mathbf{F}$.

By the generalized IME (GIME - see, e.g., Tauchen, 1985),

$$\begin{aligned}\mathbf{E} &\equiv \mathbb{E}\left\{\nabla_\theta \ln c(F_1(Y_1;\theta), F_2(Y_2;\theta);\rho)\nabla'_\rho \ln c(F_1(Y_1;\theta), F_2(Y_2;\theta);\rho)\right\} \\ &= -\mathbb{E}\nabla^2_{\theta\rho} \ln c(F_1(Y_1;\theta), F_2(Y_2;\theta);\rho)\end{aligned} \tag{36}$$

and, for $i = 1, 2$,

$$\mathbb{E}\left\{\nabla_\theta \ln f_i(Y_i;\theta)\nabla'_\rho \ln c(F_1(Y_1;\theta), F_1(Y_2;\theta);\rho)\right\} = -\mathbb{E}\nabla^2_{\theta\rho} \ln f_i(Y_i;\theta) = \mathbf{0}.$$

Also by GIME and (13),

$$\begin{aligned}&\mathbb{E}\left\{\nabla_\theta \ln f_i(Y_i;\theta)\nabla'_\theta\left[\ln f_1(Y_1;\theta) + \ln f_2(Y_2;\theta) + \ln c(.,.;\rho)\right]\right\} = \\ &\qquad = -\mathbb{E}\nabla^2_\theta \ln f_i(Y_i;\theta)\end{aligned}$$

for $i = 1, 2$, which, along with (35), implies that

$$\begin{aligned}\mathbf{G} &\equiv \mathbb{E}\left\{\nabla_\theta \ln f_1(Y_1;\theta)\nabla'_\theta \ln f_2(Y_2;\theta)\right\} \\ &= -\mathbb{E}\left\{\nabla_\theta \ln f_1(Y_1;\theta)\nabla'_\theta \ln c(F_1(Y_1;\theta), F_1(Y_2;\theta);\rho)\right\}\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}\left\{\nabla_\theta \ln f_2(Y_2;\theta)\nabla'_\theta \ln f_1(Y_1;\theta)\right\} = \\ &= -\mathbb{E}\left\{\nabla_\theta \ln f_2(Y_2;\theta)\nabla'_\theta \ln c(F_1(Y_1;\theta), F_1(Y_2;\theta);\rho)\right\} = \mathbf{G}'.\end{aligned} \tag{37}$$

Finally, by GIME and (13),

$$
\begin{aligned}
&\mathbb{E}\{\nabla_\theta \ln c(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\times\\
&\times\nabla'_\theta\left[\ln f_1(Y_1;\theta)+\ln f_2(Y_2;\theta)+\ln c(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\right]\}=\\
&\qquad\qquad\qquad\qquad= -\mathbb{E}\nabla^2_\theta \ln c(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho).
\end{aligned}
$$

With $\mathbf{G}$ as defined above and

$$
\mathbf{J}\equiv\mathbb{E}\{\nabla_\theta \ln c(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\nabla'_\theta \ln c(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\}
$$

this implies that

$$
\mathbb{E}\nabla^2_\theta \ln c(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)=\mathbf{G}+\mathbf{G}'-\mathbf{J}.
$$

□

Proof of Theorem 4: From the discussion in the main text,

$$
\mathbb{V}_{\mathrm{MLE}}=\begin{bmatrix}\mathbf{A}+\mathbf{B}+\mathbf{J}-\mathbf{G}-\mathbf{G}' & \mathbf{E}\\ \mathbf{E}' & \mathbf{F}\end{bmatrix}^{-1},
$$

$$
\mathbb{V}_{\mathrm{IQMLE}}=\left(\begin{bmatrix}-\mathbf{A} & -\mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbf{A} & \mathbf{G}\\ \mathbf{G}' & \mathbf{B}\end{bmatrix}^{-1}\begin{bmatrix}-\mathbf{A}\\ -\mathbf{B}\end{bmatrix}\right)^{-1}. \tag{38}
$$

Using partitioned inverse formulas, the upper left $p\times p$ block of $\mathbb{V}_{\mathrm{MLE}}$ can be written as $\mathbf{\Sigma}^{-1}$, where $\mathbf{\Sigma}=\mathbf{A}+\mathbf{B}+\mathbf{J}-\mathbf{G}-\mathbf{G}'-\mathbf{E}\mathbf{F}^{-1}\mathbf{E}'$.

Also,

$$
\begin{aligned}
\mathbb{V}^{-1}_{\mathrm{IQMLE}} &= \left([-\mathbf{G}' \quad -\mathbf{G}]+[\mathbb{I}\quad\mathbb{I}]\begin{bmatrix}\mathbf{A} & \mathbf{G}\\ \mathbf{G}' & \mathbf{B}\end{bmatrix}\right)\begin{bmatrix}\mathbf{A} & \mathbf{G}\\ \mathbf{G}' & \mathbf{B}\end{bmatrix}^{-1}\times\\
&\quad\times\left(\begin{bmatrix}\mathbf{A} & \mathbf{G}\\ \mathbf{G}' & \mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbb{I}\\ \mathbb{I}\end{bmatrix}+\begin{bmatrix}-\mathbf{G}\\ -\mathbf{G}'\end{bmatrix}\right) \qquad (39)\\
&= [-\mathbf{G}' \quad -\mathbf{G}]\begin{bmatrix}\mathbf{A} & \mathbf{G}\\ \mathbf{G}' & \mathbf{B}\end{bmatrix}^{-1}\begin{bmatrix}-\mathbf{G}\\ -\mathbf{G}'\end{bmatrix}\\
&\quad -\mathbf{G}'-\mathbf{G}+\mathbf{A}+\mathbf{B}. \qquad (40)
\end{aligned}
$$

Thus, $\mathbb{V}^{-1}_{\mathrm{IQMLE}}=\mathbf{\Sigma}$ if and only if

$$
\mathbf{J}-\mathbf{E}\mathbf{F}^{-1}\mathbf{E}'=[-\mathbf{G}' \quad -\mathbf{G}]\begin{bmatrix}\mathbf{A} & \mathbf{G}\\ \mathbf{G}' & \mathbf{B}\end{bmatrix}^{-1}\begin{bmatrix}-\mathbf{G}\\ -\mathbf{G}'\end{bmatrix}.
$$

□

Proof of Corollary 1:

1. If (C) is a linear combination of (A) and (B) then covariances between moment functions in (D) and (C) are linear combinations of covariances between (D) and (A-B), which are all zero by Lemma 1.

2. Rewrite $\mathbf{J} - \mathbf{V}_{21}^{\theta}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}^{\theta}$ as
$$\mathbb{E}\left\{\left(\nabla_\theta \ln c - \mathbf{V}_{21}^{\theta}\mathbf{V}_{11}^{-1}\begin{bmatrix}\nabla_\theta \ln f_1 \\ \nabla_\theta \ln f_2\end{bmatrix}\right)\nabla'_\theta \ln c\right\}.$$
This is identically zero because, due to linearity of (C) in (A-B),
$$\nabla_\theta \ln c - \mathbf{V}_{21}^{\theta}\mathbf{V}_{11}^{-1}\begin{bmatrix}\nabla_\theta \ln f_1 \\ \nabla_\theta \ln f_2\end{bmatrix} = \mathbf{0}.$$

3. By Theorem 4.

□

Proof of Theorem 5: We show that $\mathbb{E}\nabla_\theta \ln k(F_1(\mu_1+Y_1), F_2(\mu_2+Y_2);\rho) = 0$, where $\theta = (\mu_1, \mu_2)'$, holds for any RS copula density k.

By the chain rule, $\nabla_\theta \ln k(F_1(\mu_1+Y_1), F_2(\mu_2+Y_2);\rho)$ contains terms of the form
$$\frac{1}{k(F_1(\mu_1+Y_1), F_2(\mu_2+Y_2);\rho)} \times \frac{\partial k(F_1(\mu_1+Y_1), F_2(\mu_2+Y_2);\rho)}{\partial F_i(\mu_i+Y_i)} \times f_i(\mu_i+Y_i), \tag{41}$$
$i = 1, 2$.

Due to MS of (Y_1, Y_2) and RS of K, $f_i(\mu_i+Y_i) = f_i(\mu_i - Y_i)$ and $k(F_1(\mu_1+Y_1), F_2(\mu_2+Y_2)) = k(1-F_1(\mu_1+Y_1), 1-F_2(\mu_2+Y_2)) = k(F_1(\mu_1-Y_1), F_2(\mu_2-Y_2))$. So the first term in (41) is the same whether evaluated at (Y_1, Y_2) or $(-Y_1, -Y_2)$. Similarly, the last term is the same whether evaluated at Y_i or $-Y_i$.

Furthermore,
$$\begin{aligned}\frac{\partial k(F_1(\mu_1+Y_1),F_2(\mu_2+Y_2);\rho)}{\partial F_i(\mu_i+Y_i)} &= \frac{\partial k(1-F_1(\mu_1+Y_1),1-F_2(\mu_2+Y_2);\rho)}{\partial(1-F_i(\mu_i-Y_i))} \\ &= -\frac{\partial k(F_1(\mu_1-Y_1),F_2(\mu_2-Y_2);\rho)}{\partial F_i(\mu_i-Y_i)}.\end{aligned}$$

Thus, $\nabla_\theta \ln k(F_1(\mu_1+Y_1), F_2(\mu_2+Y_2);\rho) = -\nabla_\theta \ln k(F_1(\mu_1-Y_1), F_2(\mu_2-Y_2);\rho)$.

Denote $g(Y_1, Y_2) \equiv \nabla_\theta \ln k(F_1(\mu_1+Y_1), F_2(\mu_2+Y_2);\rho)\cdot h(\mu_1+Y_1, \mu_2+Y_2)$. From the above, it follows with RS of (Y_1, Y_2) that $g(-Y_1, -Y_2) = -g(Y_1, Y_2)$.

We thus have
$$\begin{aligned}\mathbb{E}\nabla_\theta \ln k(F_1(\mu_1+Y_1), F_2(\mu_2+Y_2);\rho) &= \textstyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(Y_1, Y_2)dY_1dY_2 \\ &= 0.\end{aligned}$$

□

Proof of Lemma 2: By construction, blocks $\mathbf{A}, \mathbf{B}, \mathbf{G}$ of matrices $\mathbf{V}$ and $\mathbf{D}$ are the same as in Lemma 1. However, GIME does not apply now. Denote the misspecified copula density by k.
$$\begin{aligned}\mathbf{Z} \equiv \mathbb{E}\left\{\nabla_\theta \ln k(F_1(Y_1;\theta), F_2(Y_2;\theta);\rho)\nabla'_\rho \ln k(F_1(Y_1;\theta), F_2(Y_2;\theta);\rho)\right\} \neq \\ \neq -\mathbb{E}\left\{\nabla^2_{\theta\rho} \ln k(F_1(Y_1;\theta), F_2(Y_2;\theta);\rho)\right\} \equiv \mathbf{S}.\end{aligned}$$

$$
\begin{aligned}
-\mathbf{P} &\equiv \mathbb{E}\left\{\nabla_\theta \ln f_1(Y_1;\theta)\nabla'_\rho \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\right\} \\
&\neq -\mathbb{E}\nabla^2_{\theta\rho} \ln f_1(Y_1;\theta) = \mathbf{0}
\end{aligned}
$$

and

$$
\begin{aligned}
-\mathbf{Q}' &\equiv \mathbb{E}\left\{\nabla_\theta \ln f_2(Y_2;\theta)\nabla'_\rho \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\right\} \\
&\neq -\mathbb{E}\nabla^2_{\theta\rho} \ln f_2(Y_2;\theta) = \mathbf{0}.
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{G} &\equiv \mathbb{E}\left\{\nabla_\theta \ln f_1(Y_1;\theta)\nabla'_\theta \ln f_2(Y_2;\theta)\right\} \neq \\
&\neq -\mathbb{E}\left\{\nabla_\theta \ln f_1(Y_1;\theta)\nabla'_\theta \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\right\} \equiv \mathbf{K}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\left\{\nabla_\theta \ln f_2(Y_2;\theta)\nabla'_\theta \ln f_1(Y_1;\theta)\right\} &\neq \\
\neq -\mathbb{E}\left\{\nabla_\theta \ln f_2(Y_2;\theta)\nabla'_\theta \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\right\} &\equiv \mathbf{L}'.
\end{aligned}
$$

However, by GIME and (13),

$$
\begin{aligned}
\mathbb{E}\nabla^2_\theta \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho) &= -\mathbb{E}\left\{\nabla_\theta \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\times\right. \\
&\quad \times [\nabla'_\theta \ln f_1(Y_1;\theta)+ \\
&\quad\quad +\nabla'_\theta \ln f_2(Y_2;\theta) + \\
&\quad\quad \left. +\nabla'_\theta \ln c(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)]\right\} \\
&\equiv \mathbf{K}' + \mathbf{L} - \mathbf{M},
\end{aligned} \tag{42}
$$

and

$$
\begin{aligned}
\mathbb{E}\nabla^2_{\rho\theta} \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho) &= -\mathbb{E}\left\{\nabla_\rho \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\right. \\
&\quad \times [\nabla'_\theta \ln f_1(Y_1;\theta)+ \\
&\quad\quad +\nabla'_\theta \ln f_2(Y_2;\theta) + \\
&\quad\quad \left. +\nabla'_\theta \ln c(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)]\right\} \\
&\equiv \mathbf{P}' + \mathbf{Q} - \mathbf{R},
\end{aligned} \tag{43}
$$

$$
\begin{aligned}
-\mathbf{T} &\equiv \mathbb{E}\nabla^2_\rho \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho) \\
&= -\mathbb{E}\left\{\nabla_\rho \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\nabla'_\rho \ln c(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
-\mathbf{S} &\equiv \mathbb{E}\nabla^2_{\theta\rho} \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho) \\
&= -\mathbb{E}\left\{\nabla_\theta \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\nabla'_\rho \ln c(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\right\}.
\end{aligned}
$$

Also,

$$
\mathbf{N} \equiv \mathbb{E}\left\{\nabla_\theta \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\nabla'_\theta \ln k(F_1(Y_1;\theta),F_1(Y_2;\theta);\rho)\right\} \neq \mathbf{M}
$$

and

$$\mathbf{W} \equiv \mathbb{E}\left\{\nabla_\rho \ln k(F_1(Y_1;\theta), F_1(Y_2;\theta);\rho)\nabla_\rho' \ln k(F_1(Y_1;\theta), F_1(Y_2;\theta);\rho)\right\} \neq \mathbf{T}.$$

Finally, by the well known algebraic property of cross-partial derivatives,

$$\mathbf{S} = -\mathbf{P} - \mathbf{Q}' + \mathbf{R}'.$$

□

Proof of Theorem 7: By Theorem 8(C) of Breusch et al. (1999), the copula scores are redundant for θ given the marginal scores if and only if

$$\begin{bmatrix} \mathbf{K}' + \mathbf{L} - \mathbf{M} \\ \mathbf{P}' + \mathbf{Q} - \mathbf{R} \end{bmatrix} - \begin{bmatrix} -\mathbf{K}' & -\mathbf{L} \\ -\mathbf{P}' & -\mathbf{Q} \end{bmatrix} \mathbf{V_{11}^{-1}} \begin{bmatrix} -\mathbf{A} \\ -\mathbf{B} \end{bmatrix} = \begin{bmatrix} -\mathbf{S} \\ -\mathbf{T} \end{bmatrix} \mathbb{B},$$

for some matrix $\mathbb{B} : q \times p$.

This is equivalent to

$$-\mathbf{M} - [-\mathbf{K}' \quad -\mathbf{L}]\mathbf{V_{11}^{-1}} \begin{bmatrix} \mathbf{G} \\ \mathbf{G}' \end{bmatrix} = -\mathbf{S}\mathbb{B},$$

$$-\mathbf{R} - [-\mathbf{P}' \quad -\mathbf{Q}]\mathbf{V_{11}^{-1}} \begin{bmatrix} \mathbf{G} \\ \mathbf{G}' \end{bmatrix} = -\mathbf{T}\mathbb{B}.$$

$\mathbf{T}$ is symmetric and invertible, so we can substitute $\mathbb{B}$ from the latter equation into the former to obtain

$$\mathbf{M} - [-\mathbf{K}' \quad -\mathbf{L}]\mathbf{V_{11}^{-1}}\mathbf{V_{12}^{\theta}} = \mathbf{S}\mathbf{T^{-1}}(\mathbf{R} - [-\mathbf{P}' \quad -\mathbf{Q}]\mathbf{V_{11}^{-1}}\mathbf{V_{12}^{\theta}}),$$

which completes the proof. □

Proof of Corollary 2:

1. By (30), $\mathbf{M} - \mathbf{V_{21}^{\theta k}}\mathbf{V_{11}^{-1}}\mathbf{V_{12}^{\theta}}$ is identically zero under linearity of (C) in (A-B).
2. As in 1.
3. By Theorem 7.

□

Proof of Theorem 8: See proof of Lemma 4.2 of Hansen (1982). □

Proof of Theorem 9: First note that, by standard optimal GMM results, $\hat{\theta}$ satisfies

$$\sqrt{N}(\hat{\theta} - \theta_o) = -(\mathbf{D_{11}^{o}}'\mathbf{V_{11}^{o}}^{-1}\mathbf{D_{11}^{o}})^{-1}\mathbf{D_{11}^{o}}'\mathbf{V_{11}^{o}}^{-1}\sqrt{N}\bar{g}(\theta_o) + o_p(1). \tag{44}$$

The first order condition for $\hat{\rho}$ can equivalently be written as

$$\begin{aligned}
[\nabla'_\rho \bar{r}(\hat{\theta},\hat{\rho})]'\mathbf{B_o}^{-1}\bar{r}(\hat{\theta},\hat{\rho}) &= 0, \\
\mathbf{D^o_{22}}'\mathbf{B_o}^{-1}\sqrt{N}\bar{r}(\hat{\theta},\hat{\rho}) &= o_p(1).
\end{aligned} \tag{45}$$

Now, by the mean-value theorem, we have

$$\sqrt{N}\bar{r}(\hat{\theta},\hat{\rho}) = \sqrt{N}\bar{r}(\theta_o,\rho_o) + \mathbf{D^o_{21}}\sqrt{N}(\hat{\theta}-\theta_o) + \mathbf{D^o_{22}}\sqrt{N}(\hat{\rho}-\rho_o) + o_p(1). \tag{46}$$

Substituting (44) into (46), pre-multiplying by $\mathbf{D^o_{22}}'\mathbf{B_o}^{-1}$, and solving for $\sqrt{N}(\hat{\rho}-\rho_o)$ using (45) yields

$$\begin{aligned}
\sqrt{N}(\hat{\rho}-\rho_o) &= -(\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\mathbf{D^o_{22}})^{-1}\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\sqrt{N}\bar{r}(\theta_o,\rho_o) \\
&\quad +(\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\mathbf{D^o_{22}})^{-1}\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\mathbf{D^o_{21}} \times \\
&\qquad \times(\mathbf{D^o_{11}}'\mathbf{V^o_{11}}^{-1}\mathbf{D^o_{11}})^{-1}\mathbf{D^o_{11}}'\mathbf{V^o_{11}}^{-1}\sqrt{N}\bar{g}(\theta_o) \\
&\quad +o_p(1).
\end{aligned} \tag{47}$$

Substituting (47) and (44) into (46) and simplifying results in

$$\sqrt{N}\bar{r}(\hat{\theta},\hat{\rho}) = \mathbf{R_o}\sqrt{N}\bar{\phi}(\theta_o,\rho_o) + o_p(1), \tag{48}$$

where

$$\begin{aligned}
\mathbf{R_o} &= \mathbb{I} - \mathbf{D^o_{22}}(\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\mathbf{D^o_{22}})^{-1}\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}, \\
\bar{\phi}(\theta_o,\rho_o) &= \bar{r}(\theta_o,\rho_o) - \mathbf{D^o_{21}}(\mathbf{D^o_{11}}'\mathbf{V^o_{11}}^{-1}\mathbf{D^o_{11}})^{-1}\mathbf{D^o_{11}}'\mathbf{V^o_{11}}^{-1}\bar{g}(\theta_o).
\end{aligned}$$

Note that $\sqrt{N}\bar{\phi}(\theta_o,\rho_o) \sim \mathbb{N}(\mathbf{0},\mathbf{B_o})$, and thus $\mathbf{B_o^{-1/2}}\sqrt{N}\bar{\phi}(\theta_o,\rho_o) \sim \mathbb{N}(\mathbf{0},\mathbb{I})$. Also, note that $\mathbf{R_o'}\mathbf{B_o^{-1}}\mathbf{R_o} = \mathbf{B_o^{-\frac{1}{2}}}[\mathbf{I} - \mathbf{B_o^{-\frac{1}{2}}}\mathbf{D^o_{22}}(\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\mathbf{D^o_{22}})^{-1}\mathbf{D^o_{22}}'\mathbf{B_o^{-\frac{1}{2}}}]\mathbf{B_o^{-\frac{1}{2}}}$.

Thus, the test statistic in (32) can be written as

$$N\bar{h}(\hat{\theta},\hat{\rho})'\mathbf{B_o^{-1}}\bar{h}(\hat{\theta},\hat{\rho}), \tag{49}$$

i.e. as a quadratic form in standard normals with the coefficient matrix

$$\mathbb{P} = \mathbf{I}_{p+q} - \mathbf{B_o^{-1/2}}\mathbf{D^o_{22}}(\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\mathbf{D^o_{22}})^{-1}\mathbf{D^o_{22}}'\mathbf{B_o^{-1/2}}. \tag{50}$$

This matrix is idempotent: it is the projection matrix orthogonal to $\mathbf{B_o^{-\frac{1}{2}}}\mathbf{D^o_{22}}$. The χ^2-test in (32) follows immediately because $tr(\mathbb{P}) = p + q - rank(D^o_{22}) = p$. $\square$