Essays in Theoretical and Applied Econometrics

Wanling Huang

A Thesis

In the Department

of

Economics

Presented in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy at

Concordia University

Montreal, Quebec, Canada

August 2010

© Wanling Huang, 2010

Library and Archives Canada
Published Heritage Branch
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada
Direction du Patrimoine de l'édition
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-71147-7
Our file Notre référence
ISBN: 978-0-494-71147-7

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

ABSTRACT

Essays in Theoretical and Applied Econometrics

Wanling Huang, Ph.D.
Concordia University, 2010

This thesis investigates three topics in theoretical and applied econometrics: Bartlett-type correction of the Distance Metric (DM) test, a Generalized Method of Moments (GMM) study of the effect of the North American Free Trade Agreement (NAFTA) on Quebec manufacturing industries, and a goodness-of-fit test for copulas.

The first topic derives an Edgeworth approximation of the distribution of the DM test statistic and obtains a Bartlett-type correction factor, then it uses examples of covariance structures to illustrate the theoretical results and applies the theoretical results to study the covariance structure of earnings. The second topic calculates Canadian tariff rates over the period 1991-2007 for manufacturing industries, classified using the North American Industry Classification System (NAICS), proposes a simulation-based moment selection procedure to improve the properties of the system GMM estimator, and analyzes the effect of NAFTA on earnings of Quebec manufacturing industries. The third topic proposes a new rank-based goodness-of-fit test for copulas, conducts a power study to show that the new test has reasonable properties, and presents an application.

To *my son*

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Artem Prokhorov, for teaching me everything in research, for spending so much time discussing my work, for his creative thinking, for his patience, for his approachability, for his generous financial support of my travel to various conferences, and for many other research-related things not listed here. I had never imagined that I could do such difficult work before I started it. It was Professor Prokhorov's supervision that made everything possible. Without his supervision, my hard work would have resulted in little.

I would also like to thank other members of my dissertation committee, Professor Nikolay Gospodinov and Professor Gordon Fisher, for their valuable advice and helpful suggestions. Professor Gospodinov taught me various econometrics courses, and his lecture notes have been important reference for my research. Professor Fisher taught me to write a research proposal entitled "Effect of NAFTA on Quebec Manufacturing Industries: GMM Methods," which was accepted by FQRSC, a granting agency of the Quebec Government. A research-based scholarship from FQRSC sponsored the remaining years of my PhD study.

I am grateful for the helpful comments on parts of my dissertation from Professors Christian Genest, Jorgen Hansen, Lynda Khalaf, Eric Renault, Paul Rilstone, Jeffrey M. Wooldridge; seminar participants at the Department of Economics, Ryerson University; and participants of the 43rd Conference of the Canadian Economics Association, the 26th annual meeting of Canadian Econometric Study Group, the 44th Annual Conference of the Canadian Economics Association, the International Symposium on Econometrics of Specification Tests in 30 Years, and the 16th International Conference on Panel Data. I am also grateful to Professors Seung Chan Ahn, Gerdie Everaert, Lorenzo Pozzi, Tae-Hwy Lee and Xiangdong Long for helpful codes.

I am naturally grateful to Professor Jaleel Ahmad for initially supervising my research before I entered the PhD program and to Professor William A. Sims and other professors and staff members at the Department of Economics, Concordia University for all the years I stayed at Concordia. My thanks are also due to Professor John W. Galbraith in the Department of Economics, McGill University and Professor David A. Stephens in the Department of Mathematics and Statistics, McGill University for allowing me to sit in their classes and treating me like their own students. What I learned there was really helpful to my dissertation. The same goes to the Concordia Librarians for helpful guidance in obtaining research data. My thanks are also due to the World Bank for the financial contribution which made it possible for me to use WITS and access UNCTAD-TRAINS database.

I would like to thank late Dr. Naomi Bercusson for taking care of me over the years at the Concordia Health Services and the International Students Office for their helpful advice during the starting years of my PhD program.

I would like to thank the FQRSC Doctoral Scholarship and the J.W. McConnell Memorial Graduate Fellowship, which allowed me to work full time on the dissertation, and the CIREQ Travel Grants and the CEA Travel Grant, which allowed me to present my work at conferences. I am deeply grateful to Professor Dipjyoti Majumdar for his generous financial support during the last year of my study.

Finally, I want to thank my husband for encouraging me to enter into the PhD program and especially my son. My son's standard of living in Canada is lower than in China because a student mom is not rich, but he understands everything and becomes an excellent boy.

I want to thank my mom, and the first thing I want to do after graduation is to see my mom.

Contents

List of Figures

List of Tables

Introduction

This thesis investigates three topics in theoretical and applied econometrics: Bartlett-type correction of the Distance Metric (DM) test, a Generalized Method of Moments (GMM) study of the effect of the North American Free Trade Agreement (NAFTA) on Quebec manufacturing industries, and a goodness-of-fit test for copulas. These are organized in Chapters 1, 2 and 3 respectively.

Asymptotically, the DM test statistic has a chi-squared distribution. In practice, however, this is infeasible since the sample size is finite. It is expected that after Edgeworth expansion, the distribution of the corrected DM test statistic would be closer to a chi-squared distribution than the uncorrected one. Chapter 1 mainly has three parts: in the theoretical part, the Edgeworth approximation of the distribution of the DM test statistic is derived and a Bartlett-type correction factor is obtained; in the simulation part, examples of covariance structures are given to illustrate the theoretical results; in the application part, the theoretical results are applied to study the covariance structure of earnings. The contributions of this chapter are: (i) it can be viewed as complementary to both Phillips and Park (1988) and Hansen (2006) in that it relaxes the basic requirement of nonlinear restrictions in some sense; (ii) it extends Hansen (2006) to multiple restrictions (possibly large number of degrees of freedom) and various models; (iii) it explains and provides a solution to the long-existing "troublesome" discrepancy puzzle in the labor economics literature that a longer panel reverses the original inference; (iv) the theoretical results are distribution-free.

Using GMM methods, Chapter 2 examines the effect of NAFTA on earnings of the Quebec manufacturing industries. It calculates Canadian tariff rates over the period 1991-2007 for manufacturing industries classified using the North American Industry Classification System (NAICS) and analyzes the effect of Canadian tariff concessions granted to the United States and Mexico. The system GMM method is used to estimate a two-way dynamic panel data model. In order for the system GMM estimator to have best possible properties, it proposes a simulation-based moment selection procedure, uses the procedure to select instruments, and shows that the system GMM estimator is not worse than the bootstrap-based bias-corrected least-squares dummy variable (LSDVb) estimator (Everaert and Pozzi, 2007) if moments are properly chosen. In other words, we use simulations to choose instruments so that the finite sample properties of the GMM are optimized, then we compare this "optimal" GMM estimator with the "best" non-GMM alternative. Finally, using this "optimal" GMM estimator, we find that the Canadian tariff concessions' effect on Quebec manufacturing earnings is statistically significant but economically very small.

Chapter 3 proposes a new rank-based goodness-of-fit test for copulas. It uses the information matrix equality and so relates to the White (1982) specification test. The test avoids parametric specification of marginal distributions, it does not involve kernel weighting, bandwidth selection or any other strategic choices, and it is relatively simple compared to available alternatives. The finite-sample size of this type of tests is known to deviate from their nominal size based on asymptotic critical values, and bootstrapping critical values could be a preferred alternative. A power study shows that, in a bivariate setting, the test has reasonable properties compared to its alternatives in samples as small as $1,000$ observations.

Chapter 1

Bartlett-type Correction of Distance Metric Test

The Distance Metric (DM) test of Newey and West (1987) is commonly used in econometrics to assess competing specifications. This is a simple test – the DM test statistic is usually calculated as the sample size times the difference in the criterion function evaluated at the restricted and the unrestricted estimate. At the same time, unlike some classical tests, this test is invariant to different but equivalent formulations of the restriction and robust to autocorrelation and heteroskedasticity of unknown form provided that the criterion function uses a heteroskedasticity-consistent estimate of the covariance matrix (see, e.g., Newey and McFadden, 1994). This makes the test popular among applied researchers. For example, this test has been widely used in covariance structure analysis in the context of asymptotic distribution-free estimation (see, e.g., Browne, 1984; Satorra and Bentler, 2001).

It is well known that the DM test statistic asymptotically has the chi-square distribution with r degrees of freedom, where r is the number of restrictions (see, e.g., Newey and McFadden, 1994). However, the sampling distribution of the test statistic is poorly approximated by the asymptotic distribution if samples are small (see, e.g., Clark, 1996). Edgeworth expansions can deal with this problem by expanding the sampling density of test statistics around the asymptotic density in decreasing powers of $N^{-\frac{1}{2}}$, with N being the sample size. This may improve the accuracy of the asymptotic approximation. Surveys of Edgeworth expansion methods, including the theory of their validity, are provided by Phillips (1977, 1978); Kallenberg (1993); Rothenberg (1984); Reid (1991); Sargan and Satchell (1986), among others.

However, Edgeworth expansion methods have not yet been applied to the most general version of the DM test. Most of known results concern the LR, Wald and the score test (see, e.g., Cribari-Neto and Cordeiro, 1996; Phillips and Park, 1988; Magee, 1989; Linton, 2002; Hausman and Kuersteiner, 2008). Hansen (2006) is the only (known to us) application of the Edgeworth correction to the DM test but it is restricted to the setting of a normal linear regression with a single constraint. Moreover, it is well known that Edgeworth expansions do not always improve the quality of first-order asymptotic approximations (see, e.g., Phillips, 1983). The main contribution of the paper is that we derive the Edgeworth correction, also known as the Bartlett-type correction, for the DM test in its most general form and illustrate in simulations that this corrected approximation does work better, often surprisingly better, than the uncorrected test.

We do not consider alternative ways to remedy the inaccuracy of first-order asymptotic approximations. Such alternatives include resampling techniques and other types of asymptotic approximations, e.g., saddle-point (tilted Edgeworth) or Cornish-Fisher expansions. Validity of the former is usually based on existence of an asymptotic approximation in the first place (see, e.g., Hall, 1992) and the various forms of the latter are substantially more complicated than the classical Edgeworth expansion (see, e.g., Barndorff-Nielsen and Cox, 1979).

The paper can be viewed as a generalization of the results by Hansen (2006), obtained for linear regressions with one restriction, to most of the extremum and minimum distance estimators and to multiple linear and nonlinear restrictions. We also draw on the results by Phillips and Park (1988) and Kollo and Rosen (2005). Phillips and Park (1988) investigate how higher-order terms in the asymptotic approximation of the Wald test are affected by various formulations of the null hypothesis. The DM test is invariant to such reformulations, however, their theorem on asymptotic expansion of the distribution provides a useful shortcut that substantially facilitates our proof. Kollo and Rosen (2005) provide general forms of Taylor series expansions for vector-valued functions, applicable in our setting.

In the application section, we consider a covariance structure model of Abowd and Card (1989). We address the question at what sample sizes would the proposed asymptotic correction make a difference for the empirical conclusions of that paper. It turns out that this happens at sample sizes as large as 900-1,000 observations. An interesting by-product of the application is that it explains the old puzzle in labor economics that longer panels reverse the original inference.

The DM test statistic is defined in Section 1.1. In Section 1.2 we derive the asymptotic expansion to order $O(N^{-1})$ of the DM test statistic, and in Section 1.3 we give the higher-order approximation of its distribution. Simple simulations are provided in Section 1.4, and an empirical illustration is presented in Section 1.5. Section 1.6 contains brief concluding remarks.

1.1 Distance Metric Test

For a family of distributions $\{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$, Θ compact, consider the test

$$H_0 : g(\theta) = 0,$$

$$H_1 : g(\theta) \neq 0,$$

where $g : R^p \to R^r$ is a continuously differentiable function with the first derivative defined by

$$\underset{p\times r}{A(\theta)} \equiv \frac{dg(\theta)}{d\theta}.$$

Denote $A = A(\theta_0)$. We assume that underlying the test is a parametric model that can be written in terms of the moment conditions

$$\mathbb{E}m(Z_i, \theta) = 0 \qquad \text{iff } \theta = \theta_0, \tag{1.1}$$

where $m(\cdot, \cdot)$ is a continuous k-valued function, Z_i, $i = 1, \dots, N$, is a data vector, and θ_0 is the true value of the parameter vector. We assume that the moments identify θ_0. In covariance structure models, for example, $m(Z_i, \theta) = vechZ_iZ_i' - vech\Sigma(\theta)$, where $vech$ denotes vertical vectorization of the lower triangle of a matrix and $\Sigma(\theta)$ is a model for the covariance matrix, in which $k \geq p$.

For some positive definite weighting matrix W_N, define the criterion function

$$-Q_N(\theta) \equiv \frac{1}{2} m_N'(\theta) W_N \, m_N(\theta), \tag{1.2}$$

where $\underset{k\times 1}{m_N(\theta)} \equiv \frac{1}{N} \sum_{i=1}^{N} m(Z_i, \theta)$. In covariance structure literature, the estimator that minimizes this function is known as the asymptotically distribution free (ADF) or weighted least squared (WLS) estimator (see, e.g., Browne, 1984). It is well known that efficient weighting of $m(\cdot, \cdot)$ requires that

$$\underset{k\times k}{W_N} \xrightarrow{p} W \equiv \left\{ \mathbb{E}[m(Z_i, \theta_0) m'(Z_i, \theta_0)] \right\}^{-1}.$$

We assume efficient weighting.

The test statistic we consider is based on the value of $Q_N(\theta)$ for two competing models, one that satisfies H_0 and the other that is unrestricted. Let $\bar{\theta}_N$ and $\hat{\theta}_N$ denote the corresponding estimators:

$$\bar{\theta}_N = \arg\max_{\theta\in\Theta} Q_N(\theta), \text{ subject to } g(\theta) = 0;$$

$$\hat{\theta}_N = \arg\max_{\theta\in\Theta} Q_N(\theta).$$

Then, the DM test statistic is defined (see, e.g., Newey and McFadden, 1994, p. 2222) as

$$DM \equiv -2N[Q_N(\bar{\theta}_N) - Q_N(\hat{\theta}_N)]. \tag{1.3}$$

1.2 Asymptotic Expansion of DM Test Statistic

Let

$$\mathbb{M}_N(\theta) = W_N^{1/2} m_N(\theta),$$

then the quadratic form in (1.2) becomes

$$-Q_N(\theta) = \frac{1}{2}\mathbb{M}'_N(\theta)\mathbb{M}_N(\theta),$$

and the DM test statistic in (1.3) becomes

$$DM = N[\mathbb{M}'_N(\bar{\theta}_N)\mathbb{M}_N(\bar{\theta}_N) - \mathbb{M}'_N(\hat{\theta}_N)\mathbb{M}_N(\hat{\theta}_N)]. \tag{1.4}$$

Note that, due to the efficient weighting,

$$-\sqrt{N}\mathbb{M}_N(\theta_0) \equiv \bar{q}_N \xrightarrow{d} \underset{k\times 1}{\bar{q}} \sim N(0, I). \tag{1.5}$$

Assume $\mathbb{M}_N(\theta)$ is three-times continuously differentiable. We follow Kollo and Rosen (2005, Definition 1.4.1) and define the derivative matrices as follows

$$\underset{p\times k}{G_N(\theta)} \equiv \frac{\partial \mathbb{M}'_N(\theta)}{\partial \theta},$$

$$\underset{p\times pk}{D_N(\theta)} \equiv \frac{\partial vec' G_N(\theta)}{\partial \theta},$$

$$\underset{p\times p^2k}{C_N(\theta)} \equiv \frac{\partial vec' D_N(\theta)}{\partial \theta}.$$

Also, let $G = G(\theta_0)$, $D = D(\theta_0)$, and $C = C(\theta_0)$. Then, it is easy to show (see, e.g., Newey and McFadden, 1994, p. 2219) that, under very general conditions,

$$\sqrt{N}(\hat{\theta}_N - \theta_0) = B^{-1}G\bar{q}_N + o_p(1), \tag{1.6}$$

$$\sqrt{N}(\bar{\theta}_N - \hat{\theta}_N) = -\mathbb{H}G\bar{q}_N + o_p(1), \tag{1.7}$$

where

$$B^{-1} = (GG')^{-1}$$

is the asymptotic variance matrix of $\hat{\theta}_N$, and

$$\underset{p\times p}{\mathbb{H}} \equiv B^{-1}A(A'B^{-1}A)^{-1}A'B^{-1}.$$

By Theorem 3.1.1 of Kollo and Rosen (2005, p. 280), which we provide in Appendix A for reference, the Taylor expansion of $\mathbb{M}_N(\bar{\theta}_N)$ about $\hat{\theta}_N$ can be written as follows

$$\mathbb{M}_N(\bar{\theta}_N) = \mathbb{M}_N(\hat{\theta}_N)+G'_N(\hat{\theta}_N)(\bar{\theta}_N-\hat{\theta}_N)+\frac{1}{2}[I_k\otimes(\bar{\theta}_N-\hat{\theta}_N)']D'_N(\hat{\theta}_N)(\bar{\theta}_N-\hat{\theta}_N)+o_p(N^{-1}).$$

Substituting this into (1.4) and using (1.7) yield

$$\begin{aligned} DM &= \bar{q}'G'\mathbb{H}G_N(\hat{\theta}_N)G'_N(\hat{\theta}_N)\mathbb{H}G\bar{q} \\ &\quad + \mathbb{M}'_N(\hat{\theta}_N)(I_k \otimes \bar{q}'G'\mathbb{H})D'_N(\hat{\theta}_N)\mathbb{H}G\bar{q} \\ &\quad - N^{-1/2}\bar{q}'G'\mathbb{H}G_N(\hat{\theta}_N)(I_k \otimes \bar{q}'G'\mathbb{H})D'_N(\hat{\theta}_N)\mathbb{H}G\bar{q} \\ &\quad + \frac{1}{4}N^{-1}\bar{q}'G'\mathbb{H}D_N(\hat{\theta}_N)(I_k \otimes \mathbb{H}G\bar{q})(I_k \otimes \bar{q}'G'\mathbb{H})D'_N(\hat{\theta}_N)\mathbb{H}G\bar{q} + o_p(N^{-1}). \end{aligned} \quad (1.8)$$

We will now expand at θ_0 all functions of $\hat{\theta}_N$ contained in (1.8). We wish to use Theorem 3.1.1 of Kollo and Rosen (2005) to do that. So we need to transform the current representation into the one based on vector functions. Specifically, we need the vectorized versions of matrices $G_N(\hat{\theta}_N)$ and $D_N(\hat{\theta}_N)$. Using the facts that

$$vec(ABC) = (C' \otimes A)vecB,$$
$$(A \otimes B)' = A' \otimes B',$$

we obtain the following equations

$$\bar{q}'G'\mathbb{H}G_N(\hat{\theta}_N) = vec'G_N(\hat{\theta}_N)(I_k \otimes \mathbb{H}G\bar{q}),$$
$$D'_N(\hat{\theta}_N)\mathbb{H}G\bar{q} = (I_{pk} \otimes \bar{q}'G'\mathbb{H})vecD_N(\hat{\theta}_N).$$

Equation (1.8) can now be rewritten as

$$\begin{aligned} DM = & vec'G_N(\hat{\theta}_N)M_1 vecG_N(\hat{\theta}_N) \\ & + \mathbb{M}'_N(\hat{\theta}_N)M_2 vecD_N(\hat{\theta}_N) \\ & - N^{-1/2} vec'G_N(\hat{\theta}_N)M_3 vecD_N(\hat{\theta}_N) \\ & + N^{-1}\frac{1}{4} vec'D_N(\hat{\theta}_N)M_4 vecD_N(\hat{\theta}_N) + o_p(N^{-1}), \end{aligned} \tag{1.9}$$

where

$$M_1 = (I_k \otimes \mathbb{H}G\bar{q})(I_k \otimes \bar{q}'G'\mathbb{H}),$$

$$M_2 = I_k \otimes \bar{q}'G'\mathbb{H} \otimes \bar{q}'G'\mathbb{H},$$

$$M_3 = (I_k \otimes \mathbb{H}G\bar{q})(I_k \otimes \bar{q}'G'\mathbb{H} \otimes \bar{q}'G'\mathbb{H}),$$

$$M_4 = I_k \otimes \mathbb{H}G\bar{q}\bar{q}'G'\mathbb{H} \otimes \mathbb{H}G\bar{q}\bar{q}'G'\mathbb{H}.$$

Substituting the Taylor expansions at θ_0 of $\mathbb{M}_N(\hat{\theta}_N)$, $vecG_N(\hat{\theta}_N)$ and $vecD_N(\hat{\theta}_N)$ into (1.9) gives the asymptotic expansion of the DM test statistic, which is summarized in the following theorem.

Theorem 1.1. *The asymptotic expansion of the DM test statistic is given by*

$$DM = \bar{q}'P\bar{q} + N^{-1/2}u(\bar{q}) + N^{-1}v(\bar{q}) + o_p(N^{-1}), \tag{1.10}$$

where

$$\underset{k\times k}{P} \equiv G'\mathbb{H}G, \tag{1.11}$$

$$u(\bar{q}) = u_1(\bar{q}) + u_2(\bar{q}) + u_3(\bar{q}),$$

$$v(\bar{q}) = v_1(\bar{q}) + v_2(\bar{q}) + v_3(\bar{q}) + v_4(\bar{q}),$$

with $u_i(\bar{q})$ $(i = 1, 2, 3)$ *and* $v_i(\bar{q})$ $(i = 1, 2, 3, 4)$ *specified by*

$$u_1(\bar{q}) = 2\bar{q}'G'B^{-1}DM_1 vecG, \tag{1.12}$$

$$u_2(\bar{q}) = \bar{q}'(G'B^{-1}G - I_k)M_2 vecD, \tag{1.13}$$

$$u_3(\bar{q}) = -vec'GM_3 vecD; \tag{1.14}$$

$$v_1(\bar{q}) = \bar{q}'G'B^{-1}DM_1D'B^{-1}G\bar{q} + \bar{q}'G'B^{-1}C(I_{pk} \otimes B^{-1}G\bar{q})M_1vecG, \tag{1.15}$$

$$v_2(\bar{q}) = \bar{q}'(G'B^{-1}G - I_k)M_2C'B^{-1}G\bar{q} + \frac{1}{2}\bar{q}'G'B^{-1}D(I_k \otimes B^{-1}G\bar{q})M_2vecD, \tag{1.16}$$

$$v_3(\bar{q}) = -\bar{q}'G'B^{-1}CM_3'vecG - \bar{q}'G'B^{-1}DM_3vecD, \tag{1.17}$$

$$v_4(\bar{q}) = \frac{1}{4}vec'DM_4vecD. \tag{1.18}$$

Proof. See Appendix A for all proofs.

1.3 Distribution of DM Test Statistic

In this section we follow Phillips and Park (1988) and use the Taylor expansion of DM to derive the Edgeworth expansion of its distribution to order $O(N^{-1})$. Theorem 2.4 of Phillips and Park (1988) allows one to skip many intermediate steps in deriving the expansion for the distribution from the expansion of the test statistics. Hansen (2006) used this approach for a single restriction DM test in a normal linear regression with known error variance.

In order to use Phillips and Park's results, we first show that $u(\bar{q})$ and $v(\bar{q})$ can be written in terms of Kronecker products of $\bar{q}$ and $\bar{q}\bar{q}'$. This is done in the following lemma.

Lemma 1.1. *$u(\bar{q})$ and $v(\bar{q})$ in Theorem 1.1 can be written as*

$$u(\bar{q}) = vec'J(\bar{q} \otimes \bar{q} \otimes \bar{q}),$$

$$v(\bar{q}) = tr[L(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')],$$

where

$$vecJ = vecJ_1 + vecJ_2 + vecJ_3, \tag{1.19}$$

with

$$vecJ_1 = 2(G'\mathbb{H}G \otimes G'\mathbb{H} \otimes G'B^{-1})vecD, \tag{1.20}$$

$$vecJ_2 = [(G'B^{-1}G - I_k) \otimes G'\mathbb{H} \otimes G'\mathbb{H}]vecD, \tag{1.21}$$

$$vecJ_3 = -(G'\mathbb{H}G \otimes G'\mathbb{H} \otimes G'\mathbb{H})vecD; \tag{1.22}$$

and

$$L = L_1 + L_2 + L_3 + L_4, \tag{1.23}$$

with

$$L_1 = (G'\mathbb{H} \otimes G'B^{-1})V_D(\mathbb{H}G \otimes B^{-1}G) + (G'\mathbb{H} \otimes G'B^{-1})M_V(I_k \otimes \mathbb{H}G), \tag{1.24}$$

$$L_2 = (G'\mathbb{H} \otimes G'\mathbb{H})M_{VI} + \frac{1}{2}(G'\mathbb{H} \otimes G'\mathbb{H})V_D(B^{-1}G \otimes B^{-1}G), \tag{1.25}$$

$$L_3 = -(G'\mathbb{H} \otimes G'\mathbb{H})M_V(I_k \otimes \mathbb{H}G) - (G'\mathbb{H} \otimes G'\mathbb{H})V_D(\mathbb{H}G \otimes B^{-1}G), \tag{1.26}$$

$$L_4 = \frac{1}{4}(G'\mathbb{H} \otimes G'\mathbb{H})V_D(\mathbb{H}G \otimes \mathbb{H}G), \tag{1.27}$$

where V_D, M_V and M_{VI} are given in Appendix A.

We can now apply the results of Phillips and Park (1988, p. 1069-1072) (see also Hansen, 2006, Theorem 3) to prove our main theorem.

Theorem 1.2. *The asymptotic expansion to $O(N^{-1})$ of the distribution function of DM is given by*

$$F_{DM}(x) = F_r\left(x - N^{-1}(\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3)\right) + o(N^{-1}) \tag{1.28}$$

where F_r denotes the distribution function of a χ_r^2 variate and

$$\alpha_1 = (4a_1 - b_2)/4r,$$

$$\alpha_2 = (4a_2 + b_2 - b_3)/4r(r+2),$$

$$\alpha_3 = b_3/4r(r+2)(r+4),$$

with a_i $(i = 1,2)$ and b_i $(i = 1,2,3)$ defined in Appendix A.

The Edgeworth correction factor that follows from (1.28) can be written as

$$1 - N^{-1}(\alpha_1 + \alpha_2 DM + \alpha_3 DM^2) \tag{1.29}$$

where DM is the original (uncorrected) DM test statistic. If multiplied by the correction factor, the DM test statistic should be better approximated by the χ_r^2 distribution than the uncorrected statistic. Strictly speaking, the correction cannot be called "Bartlett" because it depends on the uncorrected statistic DM. However, it is common to call such corrections Bartlett-type due to their similarity to the classical Bartlett (1937) correction (see, e.g., Cribari-Neto and Cordeiro, 1996, for a review of Bartlett and Bartlett-type corrections of common tests).

Note that increasing the number of restrictions r does not necessarily result in a bigger correction factor because α_i $(i = 1, 2, 3)$ may be negative. Moreover, it is important to note that, even if the restrictions are linear, the Bartlett-type correction factor in (1.29) will be different from one so long as $\mathbb{M}_N(\theta)$ is nonlinear in parameters. The theorem imposes no constraint on the number of restrictions tested or on the specific estimator represented by the moment condition (1.1).

1.4 Illustrative Simulations

In this section, we use simulations to illustrate the theoretical results obtained in Section 1.3 in the settings of a simple covariance structure model. Consider a random vector $Z \in \mathcal{Z} \subset \mathbb{R}^q$ from P_{θ_0}, $\theta_0 \in \Theta \subset \mathbb{R}^p$. Assume that $\mathbb{E}[Z] = 0$, $\mathbb{E}\{\|Z\|^4\} < \infty$ and $\mathbb{E}[ZZ'] = \Sigma(\theta_0)$. The matrix function $\Sigma(\theta)$ may come from a variety of models, e.g., LISREL, MIMIC, factor model, random effects or simultaneous equations model. For a random sample $(Z_1, \cdots, Z_N)$, let

$$S_i \equiv Z_i Z_i'$$

and

$$S \equiv \frac{1}{N}\sum_{i=1}^{N} S_i.$$

Then, S satisfies a central limit theorem:

$$\sqrt{N}(vechS - vech\Sigma(\theta_0)) \longrightarrow N(0, \Delta(\theta_0)),$$

where

$$\Delta(\theta_0) = \mathbb{V}(vechS_i) = \mathbb{E}[vechS_i vech'S_i] - vech\Sigma(\theta_0)vech'\Sigma(\theta_0).$$

Assume $p \leq \frac{1}{2}q(q+1)$. Then, in terminology of covariance structure literature, the degrees of freedom of the model are equal to $\frac{q(q+1)}{2} - p$, and they will be increased by one for each independent restriction imposed on $\Sigma(\theta)$ by the model. We can write all distinct moment functions as follows

$$\underset{\frac{1}{2}q(q+1)\times 1}{m_N(\theta)} \equiv \frac{1}{N}\sum_{i=1}^{N} m(Z_i, \theta) = vechS - vech\Sigma(\theta)$$

where

$$\underset{\frac{1}{2}q(q+1)\times 1}{m(Z_i, \theta)} = vechS_i - vech\Sigma(\theta).$$

The sample covariance matrix of the moments is

$$\begin{aligned}\underset{\frac{1}{2}q(q+1)\times\frac{1}{2}q(q+1)}{W_N^{-1}} &= \frac{1}{N}\sum_{i=1}^{N}[m(Z_i, \theta)m'(Z_i, \theta)] \\ &= \frac{1}{N}\sum_{i=1}^{N}[vechS_i vech'S_i - vechS_i vech'\Sigma(\theta) \\ &\quad - vech\Sigma(\theta)vech'S_i + vech\Sigma(\theta)vech'\Sigma(\theta)].\end{aligned}$$

We are interested in testing $H_0 : \Sigma(\theta) = \Sigma(c)$ against $H_1 : \Sigma(\theta) \neq \Sigma(c)$, where c is a constant vector. This type of test is fundamental in covariance structure analysis. Known as the ADF test, it has been studied by Korin (1968); Sugiura

(1969); Nagarsenker and Pillai (1973); Browne (1984); Chou et al. (1991); Muthen and Kaplan (1991); Yuan and Bentler (1997); Satorra and Bentler (2001); Yanagihara et al. (2004), among others. The literature has focused on three dimensions of the test behavior: (i) what is the effect of the sample size; (ii) how the sample size requirements change for different nonnormal distributions; (iii) how the sample size requirements change for models of different size. We wish to apply our Bartlett-type correction to the DM test of this restriction and study its behavior along the same dimensions.

For simplicity, we consider a bivariate problem (i.e. $q = 2$) in which

$$\Sigma(\theta) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix},$$

$\theta' = (\sigma_1, \sigma_{12}, \sigma_2)$, $c' = (1, 0, 1)$ and $p = k = r = 3$. So the restricted model has 3 degrees of freedom. Write the null hypothesis as

$$H_0 : \underset{3\times 1}{g(\theta)} = 0,$$

where

$$g(\theta) = vech\Sigma(\theta) - vech\Sigma(c) = \begin{bmatrix} \sigma_1^2 - 1 \\ \sigma_{12} - 0 \\ \sigma_2^2 - 1 \end{bmatrix}.$$

In order to demonstrate the effect of the Bartlett-type correction, we generate a sample of varying size from normal, Student-t and uniform distributions and compute the uncorrected and corrected versions of the DM test statistics. This is done 1,000 times. Then we plot the quantiles of the resulting bootstrap distributions. These are displayed on Figures 1.1-1.3. The quantile curve of the chi-square distribution, marked "chi^2", is drawn as a benchmark. The uncorrected and corrected versions of the DM test statistic are marked "DM" and "DM_star," respectively.

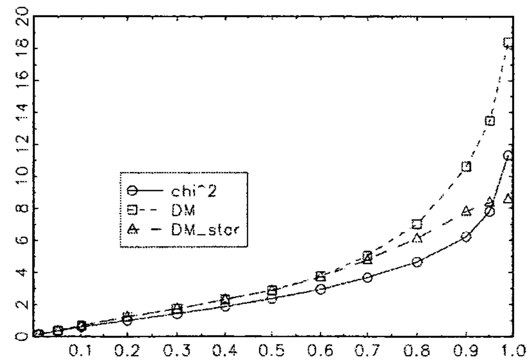
All figures show severe over-rejection of the uncorrected DM test statistic. The fact that the size of the DM test is substantially greater in small samples than the asymptotic size is well documented (see, e.g., Clark, 1996), and our results agree with that. Our corrected statistic performs much better for all distributions and all sample sizes. Of course, the corrected distribution is not identical to the chi-square distribution and the corrected test exhibits over- and under-rejection at times, but the deviations are substantially smaller than for the uncorrected test. It is perhaps surprising how much improvement one can obtain using the corrected statistic in the area close to the 95th percentile, which corresponds to the commonly used 5% significance level. At that level, the correction is almost perfect.

Figure 1.1 shows the quantiles for various sample sizes from $\mathcal{N}(0,1)$. One can clearly see from the figure how the uncorrected curve deviates from the chi-square quantiles as the sample size decreases while the degree of model complexity does not change ($q = 2$). At the same time, the corrected curve consistently provides a great deal of improvement.

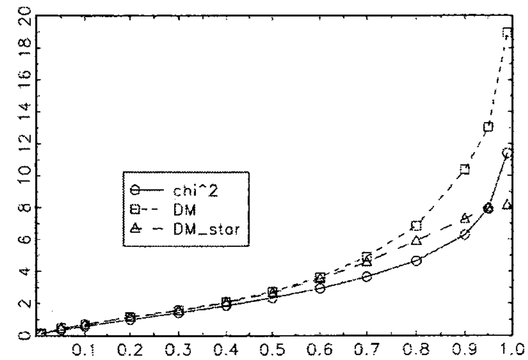
In Figure 1.2 we show the behavior of the corrected and uncorrected test statistics for two distributions, Student-t and uniform, and two sample sizes, $N = 25$ and $N = 65$. As expected, the test (and its correction), being distribution-free, exhibits similar behavior under the two distributions. The figures also show that the benefit of a larger sample size varies for the two distributions. For other distributions (not reported here), the sample size needed to obtain a similar level of approximation accuracy as in panel (d) was several hundred observations. For some distributions, the correction may be trivial even when samples are small while for others it may produce a large correction even when samples are large.

In Figure 1.3, in addition to the bivariate case, we consider a univariate ($q = 1$) model in which $\Sigma(\theta) = \sigma^2$. The null is $\sigma = c$, and the restricted model has one degree of freedom. This is done to show how model size (as measured by the degrees of freedom of the model) affects the performance of the test statistics. In the larger model ($q = 2$), the gap between the sampling and the asymptotic χ_3^2 distribution is much larger than between the sampling and the asymptotic χ_1^2

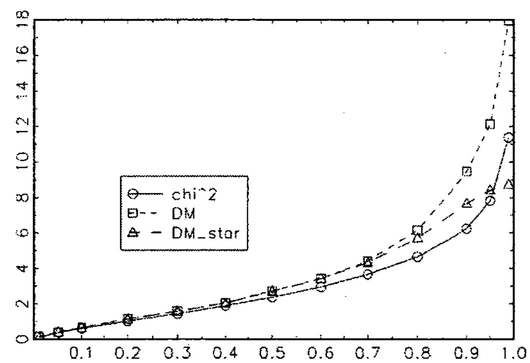
distribution in the smaller model. It is interesting to note that the model size plays as important a role in accuracy of asymptotic approximations as the sample size: we more than double the sample size between panel (b) and panel (d), and this has a similar effect on the larger model accuracy as replacing it by a model with 2 fewer degrees of freedom. This is consistent with the findings of Hoogland and Boomsma (1998) that the chi-square statistics are sensitive to model size (as measured by the degrees of freedom of the model). A bigger model requires a larger sample size to ensure good behavior of the statistics. At the same time, for the smaller models (panels (a) and (c)), larger sample sizes do not improve the asymptotic approximation by much – the approximation error is small to start with. The corrected statistic displays an improved behavior for both model sizes and both sample sizes.



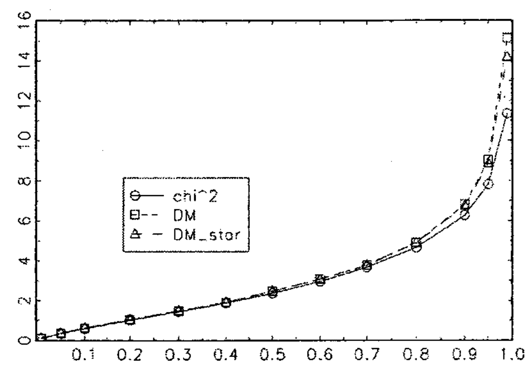
(a) $N = 25$



(b) $N = 35$

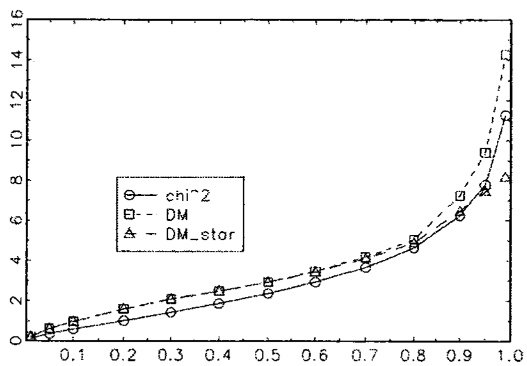


(c) $N = 65$

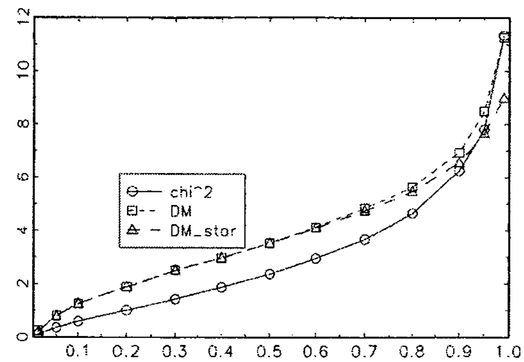


(d) $N = 200$

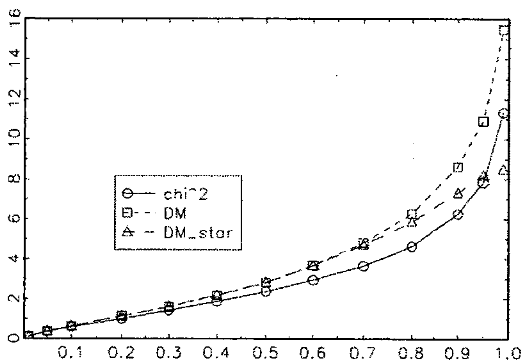
Figure 1.1: Quantiles of chi-square and bootstrap distribution of uncorrected and corrected DM test statistics for various sample sizes; $q = 2$.



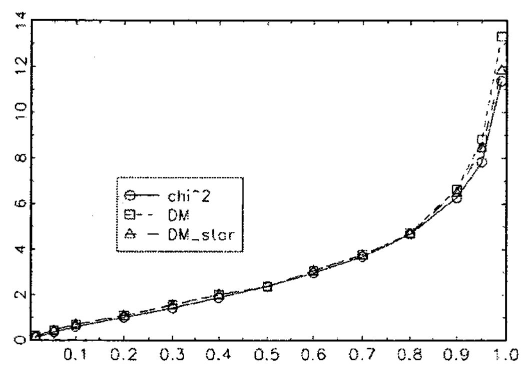
(a) $N = 25$, Student-t with 9 df.



(b) $N = 65$, Student-t with 9 df.

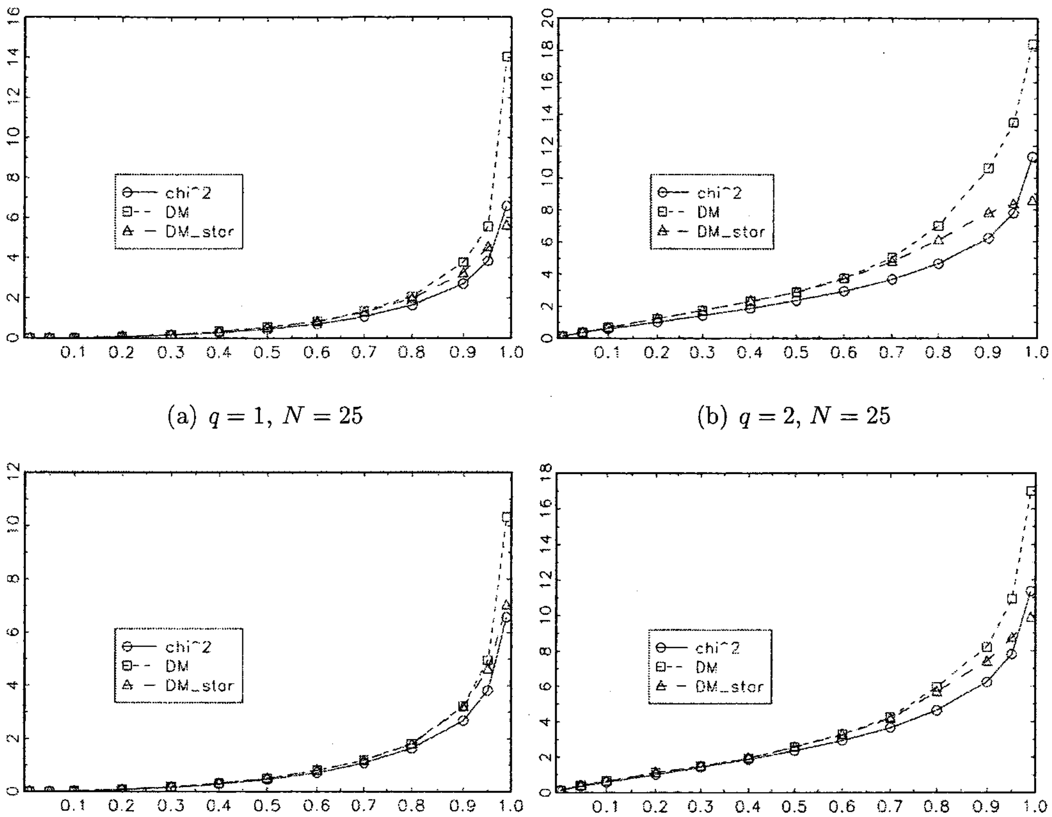


(c) $N = 25$, Uniform



(d) $N = 65$, Uniform

Figure 1.2: Quantiles of chi-square and bootstrap distribution of uncorrected and corrected DM test statistics for two data distributions and two sample sizes; $q = 2$.



(a) $q = 1$, $N = 25$

(b) $q = 2$, $N = 25$

(c) $q = 1$, $N = 65$

(d) $q = 2$, $N = 65$

Figure 1.3: Quantiles of chi-square and bootstrap distribution of uncorrected and corrected DM test statistics for two values of q and two sample sizes.

1.5 Empirical Illustration

In this section, we study applicability of the Bartlett-type correction to a covariance structure model of earnings. This type of model has been a focus of many papers in labor economics (see, e.g., MaCurdy, 1982; Abowd and Card, 1987, 1989; Topel and Ward, 1992; Baker, 1997; Baker and Solon, 2003). Among other things, the literature has been concerned with the puzzling observation that the use of longer panels results in a reversal of the original inference (see, e.g., Baker, 1997, p. 358). Longer panels are usually used to estimate higher-order autocovariances. However, the cost of longer balanced panels is a smaller number of individuals. For example, the sample sizes used by Baker (1997) in 10-year panels are 992 and 1,331 individuals for the periods 1967-76 and 1977-86, respectively; his 20-year panel contains only 534. On the other hand, as the panel gets longer (q increases), degrees of freedom grow. As mentioned earlier, this generally requires larger sample sizes for the DM statistic to remain close to the asymptotic approximation. In this section, we use parts of the sample of earnings used by Abowd and Card (1989) to demonstrate how the Bartlett-type correction affects the outcomes of a hypothesis test for various sample sizes.

The earnings data are from the Panel Study of Income Dynamics (PSID), conducted by Survey Research Center at University of Michigan. The sample consists of male heads of household, who were between the ages of 21 and 64 in the period 1969 to 1974 and who reported positive earnings in each year. The sample we use – a subsample of the data used by Abowd and Card (1989) – contains 1,578 individuals. Individuals with average hourly earnings greater than \$100 or those who reported annual hours worked greater than 4,680 were excluded. A detailed description of the PSID variables is given in Appendix A. Covariances and correlations between demeaned changes in log of real annual earnings (in 1967 dollars) are displayed in Table 1.1. Covariances are presented below the diagonal, while correlations and their two-tailed p-values are presented above the diagonal.

Table 1.1: Covariances (below diagonal) and correlations (above diagonal) between changes in log-earnings: PSID males 1967-1974

	Covariance/Correlation(with two-tailed p-value) of:				
with:	$\Delta \ln e$ 69-70	$\Delta \ln e$ 70-71	$\Delta \ln e$ 71-72	$\Delta \ln e$ 72-73	$\Delta \ln e$ 73-74
$\Delta \ln e$ 69-70	0.228	-0.204	-0.006	0.018	-0.006
		(0)	(0.827)	(0.463)	(0.823)
$\Delta \ln e$ 70-71	-0.04418	0.205	-0.415	-0.082	0
			(0)	(0.001)	(0.994)
$\Delta \ln e$ 71-72	-0.00117	-0.08345	0.197	-0.347	-0.041
				(0)	(0.101)
$\Delta \ln e$ 72-73	0.003442	-0.01447	-0.06	0.152	-0.305
					(0)
$\Delta \ln e$ 73-74	-0.00102	-0.0000303	-0.00697	-0.04518	0.144

A generic population covariance matrix for Table 1.1 can be written as

$$\Sigma(\theta) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_5^2 \end{bmatrix}, \tag{1.30}$$

where

$$\theta = vech \begin{bmatrix} \sigma_1 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_3 & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4 & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_5 \end{bmatrix}. \tag{1.31}$$

The question Abowd and Card (1989) ask is whether the information in the covariance matrix in Table 1.1 could be adequately summarized by some relatively simple statistical model. Specifically, they ask whether an MA(2) process (possibly nonstationary) can serve as the model. Indeed, there are very few covariances (correlations) that are large or statistically significant at lags greater than two. In order to address this concern, two tests were performed using the DM test statistic.

The first one is to test for a nonstationary MA(2) representation of the changes in earnings. The changes in earnings have a nonstationary MA(2) representation if the covariances at lags greater than two are zero. The null is H_0 : changes in earnings are nonstationary MA(2), and the alternative is H_1 : changes in earnings are not nonstationary MA(2). Equivalently, the null can be rewritten as

$$H_0 : \begin{bmatrix} \sigma_{41} \\ \sigma_{51} \\ \sigma_{52} \end{bmatrix} = \underset{3\times 1}{0}. \tag{1.32}$$

The second one is to test for a stationary MA(2) representation of the changes in earnings. By a stationary MA(2) representation, we mean (i) $cov(\Delta \ln e_t, \Delta \ln e_{t-j})$ depends only on j and does not change over t, and (ii) $cov(\Delta \ln e_t, \Delta \ln e_{t-j})$ is zero for $|j| > 2$. The null is H_0 : changes in earnings are stationary MA(2), and the alternative is H_1 : changes in earnings are not stationary MA(2). Equivalently, the null can be rewritten as

$$\begin{aligned} H_0 : \begin{bmatrix} \sigma_{41} \\ \sigma_{51} \\ \sigma_{52} \end{bmatrix} &= \underset{3\times 1}{0}, \\ \sigma_1 = \sigma_2 = \sigma_3 &= \sigma_4 = \sigma_5, \\ \sigma_{21} = \sigma_{32} &= \sigma_{43} = \sigma_{54}, \\ \sigma_{31} = \sigma_{42} &= \sigma_{53}. \end{aligned} \tag{1.33}$$

The test results are presented in Table 1.2. The values of the uncorrected and corrected DM test statistic (and the corresponding p-values) are very close for both tests. Not surprisingly, the corrections for this relatively large sample are minor to none. We now demonstrate the effect of the Bartlett-type correction as the sample size becomes smaller.

Table 1.2: Goodness-of-fit tests for changes in earnings: PSID males 1967-1974

Goodness-of-Fit Test	DM Test Statistic		Asy. P-Value	
N=1578	Uncorrected	Corrected	Uncorrected	Corrected
I. Nonstationary MA(2)	0.3325	0.3320	0.9538	0.9539
($df = 3$)				
II. Stationary MA(2)	19.9889	19.6262	0.0673	0.0745
($df = 12$)				

Expectedly, when the sample size becomes smaller the Bartlett-type correction becomes more important. Consider the second test as an example. The results for that test are presented in Table 1.3. We randomly select increasingly smaller subsamples of data.. As the sample size decreases from N =1,400 to 900, the correction becomes larger to the point at which the outcome of the test is reversed at conventional significance levels. For example, if N = 900, the corrected test does not reject at the 5% level while the uncorrected test does.

Table 1.3: Testing stationary MA(2) for changes in earnings: PSID males 1967-1974

Sample Size	DM Test Statistic		Asy. P-Value	
	Uncorrected	Corrected	Uncorrected	Corrected
N=1,400	22.21	21.64	0.035	0.042
N=1,200	24.15	22.83	0.019	0.029
N=1,000	25.46	22.12	0.012	0.036
N=900	25.99	20.35	0.010	0.061

Assuming that the correction does bring the sampling distribution closer to its asymptotic approximation, we conclude from this table that, for the current number of degrees of freedom, cross sections as large as 900 are not large enough to justify application of the uncorrected first-order asymptotic approximation to this covariance structure model. If used against the asymptotic critical values, the uncorrected DM test severely over-rejects.

1.6 Concluding Remarks

This paper provides the Bartlett-type correction of the DM test statistic. Our setting covers linear and nonlinear restrictions and all extremum and mimimum distance estimators that can be stated in terms of moment conditions. We also provide simple simulation evidence about the behavior of the corrected test statistic in a fairly general class of covariance structure models. In practice, it is often necessary to consider a very large (as measured by the degrees of freedom of the model) covariance structure model (see, e.g., Herzog et al., 2007; Kenny and McCoach, 2003), which makes it difficult to maintain good properties of the DM test statistic even in large samples. Moreover, large samples are not always possible to obtain and the available data are often non-normal. We show that the correction performs well in all these circumstances. The advantage of our approach over other finite-sample corrections used in the covariance structure literature (see, e.g., Yuan and Bentler, 1998, 1999) is that our results are applicable to a much wider range of models.

Chapter 2

A GMM Study of the Effect of NAFTA on Quebec Manufacturing Industries

As argued in a recent article (Postrel, 2005) in The New York Times, "Economists argue for free trade. They have two centuries of theory and experience to back them up. And they have recent empirical studies of how the liberalization of trade has increased productivity in less-developed countries like Chile and India. But still, free trade is a tough sell." Trefler (2004) argues that one reason for this is that there is not enough research on how free trade affects industrialized countries like the United States and Canada. There are a few articles addressing this question (see e.g., Romalis (2007) and references therein), but these are very rare. For research on Canada specifically, see the references listed by Trefler (2004).

In Canada, more than a half of the population live in the central area, which is made up of Ontario and Quebec. This area is the industrial and manufacturing heartland of Canada and produces more than three-quarters of all Canadian manufactured goods (*A Look at Canada*, 2007 edition). The manufacturing production of Quebec accounts for about 35% of that share (Table 304-0015, Statistics

Canada). It is therefore not surprising that this chapter focuses on the effect of the North American Free Trade Agreement (NAFTA) on Quebec manufacturing industries.

Implemented in 1994, NAFTA called for gradual reduction of tariffs among Canada, Mexico and the United States. With about one-third of the world's total GDP, NAFTA has become the world's largest free trade area, significantly larger than the European Union. Before NAFTA, there are other two major events in the history of trade liberalization among the three member countries. One is that Mexico jointed GATT in 1985, and the other is that the Canada-U.S. Free Trade Agreement (CUSFTA) came into effect in 1989. Gaston and Trefler (1997), Beaulieu (2000), and Trefler (2004) analyze the effect of tariff reductions on earnings of the Canadian industries. The data used in these articles are from 1996 and earlier, so their focus is mainly on the effect of CUSFTA. Our data cover a wider period of 1991-2007, which is far enough from CUSFTA and long enough to include the effect of NAFTA.

According to Foreign Affairs and International Trade Canada (FAITC), "NAFTA has contributed to raising standards of living" (FAITC, 2001). This chapter seeks to evaluate this statement quantitatively by looking at earnings of the Quebec manufacturing industries before and after introduction of NAFTA. Specifically, the question we ask in this chapter is: what is the tariff reductions effect of NAFTA on Quebec manufacturing earnings?

The remainder of the chapter is organized as follows. In Section 2.1, we discuss the econometric model and estimation method. Specifically, the system GMM (Arellano and Bover, 1995; Blundell and Bond, 1998) is used to estimate a two-way dynamic panel data model. Our sample is small, so in order for the system GMM estimator to have desirable properties, we need to decide which instruments to use. Thus, a simulation-based moment selection procedure is proposed in Section 2.2. Based on this procedure, the system GMM estimator is compared with the bootstrap-based bias-corrected least-squares dummy variable (LSDVb) estimator (Everaert and Pozzi, 2007), which is also described in Section 2.2. Estimation results are presented and analyzed in Section 2.3. Some brief conclusions are given in Section 2.4.

2.1 Econometric Model and Estimation Method

The main purpose of the chapter is to estimate the effect of Canadian tariff reductions on Quebec manufacturing earnings. It is well-known that earnings data usually display strong autocorrelation. The autocorrelation among Quebec manufacturing earnings data is presented in Table 2.1. The strong autocorrelations suggest that the appropriate econometric model for Quebec manufacturing earnings should be a dynamic one.

An alternative approach that accounts for strong autocorrelations is known as long double differencing. Long double differencing is essentially a difference-in-difference approach that compares changes in the dependent variable at two distant points in time, one before FTA and one after. Trefler (2004, p. 874) argues that every previous FTA study used annual data without any correction for autocorrelation. He cites Gaston and Trefler (1997), Head and Ries (1999a,b), Beaulieu (2000) and Clausing (2001). He then adopts long double-differencing models to avoid the dynamic panel estimation problems. From the perspective of interpretation, a dynamic panel data model is more appropriate for the present chapter than a long double-differencing one. As Baltagi (2005) points out, "many economic relationships are dynamic in nature and one of the advantages of panel data is that they allow the researcher to better understand the dynamics of adjustment."

Using a panel of 71 Quebec manufacturing industries over the period 1991-2007, we estimate a typical two-way dynamic panel data model. A list of these industries is provided in Table A in the Appendix. Holtz-Eakin (1988) finds evidence in the

Table 2.1: Autocorrelation among Quebec manufacturing earnings : 1991-2007

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
1991	1	.98	.96	0.95	0.96	0.95	0.93	0.91	0.87	0.87	0.86	0.84	0.81	0.74	0.71	0.70	0.70
1992	0.98	1	0.97	0.96	0.96	0.95	0.93	0.91	0.89	0.89	0.87	0.86	0.83	0.77	0.73	0.72	0.72
1993	0.96	0.97	1	0.99	0.98	0.97	0.95	0.92	0.91	0.90	0.90	0.89	0.86	0.81	0.78	0.77	0.75
1994	0.95	0.96	0.99	1	0.98	0.96	0.96	0.94	0.92	0.91	0.90	0.88	0.85	0.81	0.78	0.76	0.74
1995	0.96	0.96	0.98	0.98	1	0.98	0.96	0.93	0.92	0.92	0.91	0.89	0.86	0.82	0.78	0.76	0.75
1996	0.95	0.95	0.97	0.96	0.98	1	0.97	0.94	0.92	0.92	0.91	0.90	0.88	0.84	0.80	0.78	0.75
1997	0.93	0.93	0.95	0.96	0.96	0.97	1	0.97	0.94	0.94	0.93	0.92	0.91	0.85	0.82	0.81	0.78
1998	0.91	0.91	0.92	0.94	0.93	0.94	0.97	1	0.98	0.97	0.96	0.95	0.93	0.87	0.84	0.82	0.80
1999	0.87	0.89	0.91	0.92	0.92	0.92	0.94	0.98	1	0.99	0.98	0.96	0.95	0.90	0.87	0.85	0.82
2000	0.87	0.89	0.90	0.91	0.92	0.92	0.94	0.97	0.99	1	0.99	0.97	0.96	0.92	0.89	0.87	0.84
2001	0.86	0.87	0.90	0.90	0.91	0.91	0.93	0.96	0.98	0.99	1	0.99	0.98	0.93	0.90	0.88	0.86
2002	0.84	0.86	0.89	0.88	0.89	0.90	0.92	0.95	0.96	0.97	0.99	1	0.98	0.94	0.91	0.89	0.86
2003	0.81	0.83	0.86	0.85	0.86	0.88	0.91	0.93	0.95	0.96	0.98	0.98	1	0.97	0.93	0.92	0.88
2004	0.74	0.77	0.81	0.81	0.82	0.84	0.85	0.87	0.90	0.92	0.93	0.94	0.97	1	0.98	0.95	0.91
2005	0.71	0.73	0.78	0.78	0.78	0.80	0.82	0.84	0.87	0.89	0.90	0.91	0.93	0.98	1	0.98	0.94
2006	0.70	0.72	0.77	0.76	0.76	0.78	0.81	0.82	0.85	0.87	0.88	0.89	0.92	0.95	0.98	1	0.96
2007	0.70	0.72	0.75	0.74	0.75	0.75	0.78	0.80	0.82	0.84	0.86	0.86	0.88	0.91	0.94	0.96	1

Panel Study of Income Dynamics (PSID) data that the wage dynamics follow an AR(2) model. We initially included two lags of the dependent variable as regressors, but regressions using the present data always show that the second lag is insignificant. This is not surprising because large high-order autocorrelation in the dependent variable does not imply the same order AR model.

Let i index industries ($i = 1, 2, \cdots, N; N = 71$) and t index years ($t = 1, 2, \cdots, T; T = 17$), then the model can be specified as:

$$\ln y_{it} = \alpha \ln y_{i(t-1)} + \beta'(L)x_{it} + \lambda_t + u_{it}, \qquad |\alpha| < 1, \tag{2.1}$$

where

$$u_{it} = \eta_i + v_{it}.$$

The first-differenced transformation of (2.1) can be written as:

$$\Delta \ln y_{it} = \alpha \Delta \ln y_{i(t-1)} + \beta'(L)\Delta x_{it} + \Delta\lambda_t + \Delta v_{it}. \tag{2.2}$$

Here y_{it} is CPI-adjusted Quebec manufacturing earnings; x_{it} is a vector containing a set of explanatory variables which include τ_{it}^{us}, the effective tariffs[1] imposed by Canada on imports from the United States, and τ_{it}^{mex}, the effective tariffs for imports from Mexico; $\beta'(L)$ is a vector of polynomials in the lag operator; λ_t is a time effect; and η_i is an industry-specific effect. Detailed data description is provided in the Appendix. In order to be assured of capturing a sufficiently long effect of the tariffs, the maximum number of lags in x_{it} is initially set at four. In subsequent notation, q and k denote the maximum number of lags in x_{it} and the dimension of x_{it} respectively; then $q = 4$ and $k = 2$. A panel unit root test using the cross-sectionally augmented ADF (CADF) statistic proposed by Pesaran (2007) shows that, augmented by one lag, the P-value of the test statistic is 0.001. This verifies that $|\alpha| < 1$ is satisfied.

[1]In calculating tariffs, both non-zero and zero imports are considered. See the Appendix for details.

The main reason for including a time effect λ_t is to remove the business cycle effect illustrated in Figure 2.1. Since N here is considered large and T is small, the time effects can be treated as unknown period specific parameters to be estimated, and a full set of time dummies is included for estimation (see also Arellano, 2003, p. 61 and Dahlberg and Johansson, 2000, p. 403). Alternatively, one might consider including a generated regressor in the differenced model to control for the business cycle. For example, Trefler (2004) runs a differenced time-series regression for each i to obtain the business cycle control variable. Note that this cannot be done for equations in levels, because of possible trends and unit root problems in time-series regression. However, this method has additional complications. First, the estimates of the standard deviation are wrong unless the nuisance parameters do not enter into the moment functions of interest (see, e.g., Pagan, 1984; Prokhorov and Schmidt, 2009a), so an adjustment is usually needed (see, e.g., Wooldridge, 2002). Secondly, since N is large, regressions with specification tests for each i are time-consuming. Thirdly, in the regressions for each i, sample sizes are not large and some corrections may be necessary to avoid invalid inference.

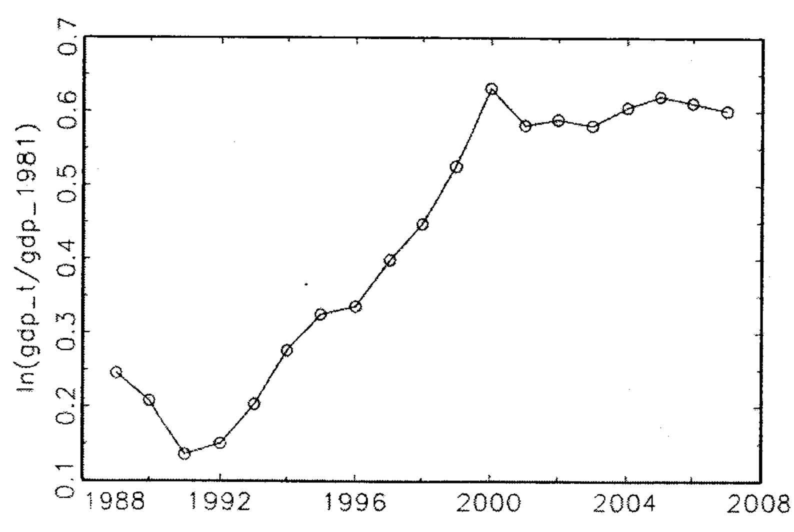


Figure 2.1: Real Canadian manufacturing GDP

The v_{it} in (2.1) is assumed to have finite moments and in particular $\mathbb{E}(v_{it}) = \mathbb{E}(v_{it}|y_{i(t-1)}, \lambda_t, \eta_i) = 0$ for all t. This assumption implies that v is serially uncorrelated but not necessarily independent over time. The OLS estimator from the first-difference specification is not consistent since $\Delta v_{it} = v_{it} - v_{i(t-1)}$ and $\Delta y_{i(t-1)} = y_{i(t-1)} - y_{i(t-2)}$ are correlated. However, $v_{it} - v_{i(t-1)}$ is uncorrelated with $y_{i(t-s)}$ for $s \geq 2$ and the differenced equation can be estimated consistently by GMM using $y_{i(t-s)} (s \geq 2)$ as instruments. It is obvious that an estimator that uses lags as instruments under the assumption of white noise errors would lose its consistency if in fact the errors in the levels equation were serially correlated. In other words, the validity of the instrumental variables hinges heavily upon lack of serial correlation in the errors. The first-order serial correlation in the first-difference errors need not be zero, but the second-order serial correlation has to be zero. Therefore, equivalently we can say that the consistency of the GMM estimators depends on the assumption $\mathbb{E}(\Delta v_{it} \Delta v_{i(t-2)}) = 0$ (Arellano and Bond, 1991, p. 281), which is exactly the null hypothesis of the Arellano-Bond specification test. Arellano and Bond (1991) show that asymptotically this test statistic has a standard normal distribution.

Note that we do not impose any restrictions on x_{it}. Let Z_i^+ denote the matrix of instruments used in estimation. Then (Blundell and Bond, 1998, p. 126),

$$Z_i^+ = \begin{bmatrix} Z_i & 0 & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & 0 & \cdots & 0 \\ 0 & 0 & \Delta y_{i3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \Delta y_{i(T-1)} \end{bmatrix}, \tag{2.3}$$

where Z_i is a matrix of instruments for the differenced equation (Arellano and Bond, 1991) and Δy_{is} $(s = 2, \cdots, T-1)$ are the instruments for the levels equation (Arellano and Bover, 1995; Blundell and Bond, 1998). The form of Z_i depends on whether x_{it} is strictly exogenous, predetermined or endogenous. If x_{it} is strictly exogenous, i.e. $\mathbb{E}(x_{it}v_{is}) = 0$ for all t and s, all the $x's$ are valid instruments and Z_i has the form:

$$Z_i = \begin{bmatrix} y_{i1} & x'_{i1} & \cdots & x'_{iT} & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & y_{i1} & y_{i2} & x'_{i1} & \cdots & x'_{iT} & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & y_{i1} & \cdots & y_{i(T-2)} & x'_{i1} & \cdots & x'_{iT} \end{bmatrix}.$$

So Z_i is a $(T-2)\times(T-2)[(T-1)+2kT]/2$ matrix. If x_{it} is not strictly exogenous, things would be a little more complicated. Consider a simple situation where no lags of x_{it} are included. Due to a possible feedback from lags of the dependent variable to the current values of x, v_{it} can be correlated with future values of x and x_{it} is not strictly exogenous but only predetermined in the sense that $\mathbb{E}(x_{it}v_{is}) \neq 0$ for $s < t$ and zero otherwise. Then only $x_{i1}, \cdots, x_{i(s-1)}$ are valid instruments in the differenced equation for period s so that Z_i has the form:

$$Z_i = \begin{bmatrix} y_{i1} & x'_{i1} & x'_{i2} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & y_{i1} & y_{i2} & x'_{i1} & x'_{i2} & x'_{i3} & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{i(T-2)} & x'_{i1} & \cdots & x'_{i(T-1)} \end{bmatrix},$$

which is a $(T-2)\times(T-2)[(T-1)+k(T+1)]/2$ matrix. If, further, x_{it} is endogenous instead of just predetermined, then we lose one instrument for each regressor in x for each period. That is, only $x_{i1}, \cdots, x_{i(s-2)}$ are valid instruments in the differenced equation for period s so that Z_i has the form:

$$Z_i = \begin{bmatrix} y_{i1} & x'_{i1} & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & y_{i1} & y_{i2} & x'_{i1} & x'_{i2} & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{i(T-2)} & x'_{i1} & \cdots & x'_{i(T-2)} \end{bmatrix},$$

which is a $(T-2)\times(T-2)(T-1)(1+k)/2$ matrix. Now it is pretty easy to consider the situation that q lags of x_{it} are also included in the model specification, in which case we lose q instruments for every regressor in x for each period as compared with the no lags case.

An estimator using only elements in Z_i as instruments is known as the differenced GMM estimator in the literature, while an estimator using both Z_i and Δy_{is}

as instruments is known as the system GMM estimator. Simulations by Blundell et al. (2000) show that the system GMM estimator has much better finite sample properties than the differenced GMM estimator: it not only greatly improves the precision but also greatly reduces the finite sample bias.[2] Therefore, we use the system GMM estimator. The sample moments, on which the system GMM estimator is based, can be expressed as $N^{-1}\sum_{i=1}^{N} Z_i^{+\prime} v_i^{+}$, where

$$v_i^{+} = \begin{bmatrix} \Delta v_{i3} \\ \vdots \\ \Delta v_{iT} \\ u_{i3} \\ \vdots \\ u_{iT} \end{bmatrix}_{2(T-2)\times 1} . \tag{2.4}$$

With the advent of the system GMM estimator, the imprecision of the GMM estimator is not so severe anymore, but an important decision still has to be made while using this estimation method: which instruments should be used? Perhaps the best estimator so far for dynamic panel data models is the LSDVb estimator[3] proposed by Everaert and Pozzi (2007). The bootstrap is a very effective method in bias correction for non-GMM estimators, as shown by Everaert and Pozzi (2007), but it might not be appropriate for GMM estimators due to weak instruments. However, if moments are properly selected, the system GMM estimator might not be a bad choice. In the next section, we propose a simulation-based moment selection procedure and discuss this issue in detail. We later show that the system GMM estimator is not worse than the LSDVb estimator.

[2]Interested readers may refer to Hayakawa (2007) for an analysis of why the system GMM estimator is less biased than the differenced GMM estimator even though the former uses more instruments.

[3]LSDVb is a bias-corrected version for the least-squares dummy variable estimator based on an iterative bootstrap procedure.

2.2 Simulation-Based Moment Selection

It has been seen in Section 2.1 that the number of moment conditions is of order T^3, which can be very large even if T is only moderately large. Simulations by Ziliak (1997) indicate that the bias/efficiency trade-off exists for GMM as the number of moment conditions increases, and that the downward bias in GMM is quite severe as the number of moment conditions expands, outweighing the gains in efficiency. Bun and Kiviet (2006) derive the second-order bias of the system GMM estimator and examine how the order of magnitude of bias changes when a different set of instruments is used. Andrews and Lu (2001) and Han and Phillips (2006) address the problem of GMM with many moment conditions asymptotically, and Newey and Windmeijer (2009) give a new variance estimator for generalized empirical likelihood (GEL) in this aspect. Okui (2009) derives an approximation of the mean square error (MSE) and proposes a procedure for choosing the number of instruments for the differenced GMM estimator. Doran and Schmidt (2006) use principal components of the weighting matrix to effectively drop some of the moment conditions and improve finite sample properties of GMM estimators. In this section, we consider a moment selection approach for the system GMM estimator, based on real-life data simulations. Simulation using real-life data is not new (see, e.g., Dahlberg and Johansson, 2000), but this kind of simulation for moment selection has not been studied yet. Dahlberg and Johansson (2000) use real-life data in Monte Carlo experiments to verify that testing against bootstrap critical values is superior to testing against asymptotic critical values. We use a similar method to select moments for the purpose of obtaining a better system GMM estimator.

2.2.1 Motivation

Assume x_{it} is strictly exogenous. We exploit the exogeneity of Δx_{it} and $\Delta\lambda_t$ in the differenced equation (2.2) and the instruments $\Delta y_{i(t-1)}$ for the levels equation (2.1). Then, adding a different number of lags of the dependent variable as additional instruments for equation (2.2) would have different regression results, see Table 2.2. In column (a'), only one lagged dependent variable $y_{i(t-2)}$ is used as an additional instrument. Eliminating insignificant regressors, we obtain results in (a). Similarly, if two lags ($y_{i(t-2)}, y_{i(t-3)}$) or three lags ($y_{i(t-2)}, y_{i(t-3)}, y_{i(t-4)}$) are used as additional instruments, (b) and (c) are eventually obtained respectively. For each regression in Table 2.2, the standard error is corrected based on Windmeijer (2005). Both the Arellano-Bond test and the Wald test verify that the model specification for each regression in Table 2.2 is correct[4].

The results in Table 2.2 show that as more lags of the dependent variable are added as instruments, the coefficient estimate of the lagged dependent variable becomes smaller, from .9697 to .9526, and the magnitude and significance of the estimates for other regressors are different. When all lags (from $y_{i(t-2)}$ to y_{i1}) of the dependent variable are added, the coefficient estimate of the lagged dependent variable is as small as .7111[5], which has not been reported in Table 2.2. A practical question is which regression should be used, i.e., how many instruments to employ.

A reasonable answer to the question is to choose the regression with the smallest MSE, but the MSE cannot be computed because the true values of the coefficients are unknown. Therefore, we assume that earnings are a function of tariffs and perform simulations.

[4]The null hypothesis of the Arellano-Bond test is discussed in the main text. The Wald test for model specification tests the joint significance of the independent variables, but it is not shown in the table to save space.

[5]This suggests that for some classic examples in the literature of dynamic panel data models, if fewer instruments are used, the estimate of the lagged dependent variable may approach one.

Table 2.2: GMM estimation using different instruments

Independent variables[a]	(a')	(a)	(b)	(c)
$\ln earnings_{i(t-1)}$	.9676 (.0414)	.9697 (.0394)	.9614 (.0381)	.9526 (.0377)
τ_{it}^{us}	.0274 (.0050)	.0293 (.0053)	.0268 (.0058)	.0257 (.0085)
$\tau_{i(t-1)}^{us}$	.0057 (.0078)	—	—	—
$\tau_{i(t-2)}^{us}$	.0130 (.0069)	.0117 (.0039)	.0128 (.0048)	—
$\tau_{i(t-3)}^{us}$	-.0110 (.0075)	—	—	—
$\tau_{i(t-4)}^{us}$	-.0234 (.0075)	-.0267 (.0057)	-.0288 (.0069)	-.0359 (.0116)
τ_{it}^{mex}	-.0292 (.0235)	—	-.0449 (.0194)	—
$\tau_{i(t-1)}^{mex}$	.0079 (.0171)	.0283 (.0110)	—	—
$\tau_{i(t-2)}^{mex}$	-.0087 (.0223)	—	—	—
$\tau_{i(t-3)}^{mex}$	-.0040 (.0260)	—	—	—
$\tau_{i(t-4)}^{mex}$	-.0038 (.0376)	—	—	—
Arellano-Bond test	-1.147	-1.1771	-1.1252	-1.1387
No. of observations	923	923	923	923

[a]Dependent variable is $\ln(earnings)_{it}$.

Notes:

(i) Windmeijer WC-robust estimators for standard errors are reported in parentheses;

(ii) The GMM estimates reported are all two step;

(iii) Time dummies are not shown here to save space;

(iv) The Arellano-Bond test reports the test for second-order serial correlation in the first-differenced residuals.

2.2.2 Experimental Design

Suppose earnings are a function of last year's earnings, tariffs (possibly at some lags) imposed by Canada on imports from the United States and tariffs for imports from Mexico. A time effect (λ_t), an individual fixed effect (η_i) and a white noise error term (v_{it}) are also included in the function. The basic idea for simulations is: take a set of coefficient estimates from Table 2.2 as the true parameter values and use the real data on regressors to generate values on the dependent variable, then run regressions using different instruments and calculate the MSE for each regression. Since there are multiple regressors in the model, we use the sum over the MSE calculated for the coefficient estimate on each regressor, denoted by sumMSE, as a criterion. This process is repeated many times. Note that every time, the number of instruments for the minimum sumMSE regression might not be the same. We choose the number of instruments for which sumMSE is minimized most frequently.

Three Monte Carlo experiments are conducted. The data generating process (DGP) of the three experiments are represented by equations (2.5), (2.6) and (2.7):

$$\widehat{\ln y_{i1}} = \ln y_{i1},$$

$$\widehat{\ln y_{it}} = \alpha \widehat{\ln y_{i(t-1)}} + \beta_{10}\tau_{it}^{us} + \beta_{12}\tau_{i(t-2)}^{us} + \beta_{14}\tau_{i(t-4)}^{us} + \beta_{21}\tau_{i(t-1)}^{mex} + \lambda_t + \hat{\eta}_i^a + v_{it}, \qquad (2.5)$$

$$\widehat{\ln y_{it}} = \alpha \widehat{\ln y_{i(t-1)}} + \beta_{10}\tau_{it}^{us} + \beta_{12}\tau_{i(t-2)}^{us} + \beta_{14}\tau_{i(t-4)}^{us} + \beta_{20}\tau_{it}^{mex} + \lambda_t + \hat{\eta}_i^b + v_{it}, \qquad (2.6)$$

$$\widehat{\ln y_{it}} = \alpha \widehat{\ln y_{i(t-1)}} + \beta_{10}\tau_{it}^{us} + \beta_{14}\tau_{i(t-4)}^{us} + \lambda_t + \hat{\eta}_i^c + v_{it}, \qquad (2.7)$$

$$|\alpha| < 1, \quad t = 2, \cdots, T$$

where $\hat{\eta}_i^a$, $\hat{\eta}_i^b$ and $\hat{\eta}_i^c$ are estimated from real-life data:

$$\hat{\eta}_i^a = \frac{1}{T-4}\sum_{t=5}^{T}(\ln y_{it} - \alpha \ln y_{i(t-1)} - \beta_{10}\tau_{it}^{us} - \beta_{12}\tau_{i(t-2)}^{us} - \beta_{14}\tau_{i(t-4)}^{us} - \beta_{21}\tau_{i(t-1)}^{mex} - \lambda_t),$$

$$\hat{\eta}_i^b = \frac{1}{T-4}\sum_{t=5}^{T}(\ln y_{it} - \alpha \ln y_{i(t-1)} - \beta_{10}\tau_{it}^{us} - \beta_{12}\tau_{i(t-2)}^{us} - \beta_{14}\tau_{i(t-4)}^{us} - \beta_{20}\tau_{it}^{mex} - \lambda_t),$$

$$\hat{\eta}_i^c = \frac{1}{T-4}\sum_{t=5}^{T}(\ln y_{it} - \alpha \ln y_{i(t-1)} - \beta_{10}\tau_{it}^{us} - \beta_{14}\tau_{i(t-4)}^{us} - \lambda_t).$$

Here $\ln y_{it}$, τ_{it}^{us}, τ_{it}^{mex} and λ_t are real-life data for Quebec manufacturing industries, and $\hat{\ln} y_{it}$ denotes generated data. There are 71 cross-sections denoted by i and 17 time periods denoted by t. The parameter values for equations (2.5), (2.6) and (2.7) are taken from columns (a), (b) and (c) of Table 2.2 respectively.

For each of the three experiments, the number of replications is set to be 100: 110 samples are produced from each of the equations (2.5), (2.6) and (2.7) and the first ten samples are discarded. For each sample in each experiment, regressions using different lags of the dependent variable as additional instruments are run. Finally, sumMSE for each regression is calculated and the regression with the smallest sumMSE for each sample is found. This "optimal" regression is denoted as GMMsj, where j is the number of lags of the dependent variable that are used as additional instruments.

2.2.3 Simulation Results

The results from the three experiments are displayed in Table 2.3. In Experiment I, 74 out of 100 samples (74%) have the smallest sumMSE occuring at GMMs1, i.e., using only one lagged dependent variable $y_{i(t-2)}$ as an additional instrument[6] for each period. In Experiments II and III, the occurrence increases to 81% and 85% respectively. For all three experiments, the smallest sumMSE never occurs at $j > 4$. The simulation results suggest that using lags more than four as additional instruments would never be justified in terms of sumMSE, and the system GMM using only one lagged dependent variable $y_{i(t-2)}$ as an additional instrument is statistically the best choice. In other words, adding moments beyond GMMs1 increases bias too much to make efficiency impovement useful. Therefore, the system GMM using only one lagged dependent variable $y_{i(t-2)}$ as an additional instrument for each period is appropriate for our application.

[6]in addition to $\Delta y_{i(t-1)}$.

Table 2.3: Occurrence of regression with smallest sumMSE at GMMsj : estimation

Experiment[a]:	I	II	III
GMMs1	74%	81%	85%
GMMs2	17%	14%	11%
GMMs3	7%	3%	2%
GMMs4	2%	2%	2%
Total:	100%	100%	100%

[a]Experiment I, II and III: true values of coefficients are from (a), (b) and (c) in Table 2.2 respectively.

2.2.4 System GMM vs LSDVb

Although the simulation-based moment selection procedure above is presented for the special case of Quebec manufacturing industries, its basic idea applies to any kind of empirical analysis. Obviously, it also applies to simulations with known true parameter values. Take Table 2 of Everaert and Pozzi (2007) as an example[7]. In it, the authors report bias, standard deviation and root MSE in estimating γ and β[8]. Their results suggest that the system GMM may perform worse than LSDVb. We want to find out why in some cases the system GMM estimator performs worse than LSDVb and in other cases better.

We are interested in cases with $N > T$. Specifically, we compare GMMs3 (estimated with instrument set $\{y_{i(t-2)}, y_{i(t-3)}, y_{i(t-4)}, x_{i(t-1)}, x_{it}, x_{i(t+1)}\}$ for each period) with LSDVb for five cases: (i) $T = 5$, $N = 20$; (ii) $T = 10$, $N = 20$; (iii) $T = 5$, $N = 100$; (iv) $T = 10$, $N = 100$; (v) $T = 5$, $N = 500$. The relative results of Everaert and Pozzi (2007) are reprinted in Table 2.4, from which we can say that, based on the sumMSE, the system GMM performs worse than LSDVb in cases (i), (ii) and (iv), but better in cases (iii) and (v). Table 2.5 reports the frequency of the smallest sumMSE regression occuring at GMMsj ($j = 1, \cdots, 8$)

[7]We thank Gerdie Everaert and Lorenzo Pozzi for their code.

[8]γ and β are the coefficients in their model $y_{it} = \gamma y_{i(t-1)} + \beta x_{it} + \eta_i + \varepsilon_{it}$.

for each case. Our attention at this point is focused on GMMs3, which uses three lags of the dependent variable as additional instruments. This is the number of instruments at which the relative frequency of minimum sumMSE is largest for cases (i) (60%), (iii) (67%) and (v) (65%). For cases (ii) and (iv), the frequency is 16% and 10% respectively.

Table 2.4: Part of Table 2 on p. 1171, Everaert and Pozzi (2007)

T	N		Bias γ	Std γ	Rmse γ	Bias β	Std β	Rmse β
5	20	LSDVb	-0.143	0.154	0.210	-0.016	0.183	0.184
		GMMs3	-0.139	0.182	0.229	0.004	0.203	0.203
10	20	LSDVb	-0.036	0.085	0.093	-0.004	0.099	0.099
		stacked GMMs3	-0.042	0.116	0.123	0.008	0.125	0.126
5	100	LSDVb	-0.132	0.073	0.151	-0.011	0.076	0.077
		GMMs3	-0.030	0.099	0.104	-0.000	0.092	0.092
10	100	LSDVb	-0.037	0.038	0.053	-0.003	0.045	0.045
		GMMs3	-0.024	0.053	0.058	-0.001	0.059	0.059
5	500	LSDVb	-0.126	0.030	0.130	-0.010	0.035	0.036
		GMMs3	-0.003	0.037	0.037	-0.000	0.039	0.039

In cases (iii) and (v) where sample sizes are not very small, the properties of the system GMM estimator are not worse than LSDVb: the root MSE of γ is much smaller though that of β is slightly larger. In case (i) where the sample size is small, the bias of the system GMM is much smaller as compared with any bias-corrected estimators in the original table of Everaert and Pozzi (2007), but the bias is not small enough to offset the large standard deviation. If the sample size is not so small, say $N = 50$, which is a sample size often encountered in empirical applications, we run the regression using Gerdie Everaert and Lorenzo Pozzi's code and find that the root MSE of γ of the system GMM is better than LSDVb but that of β is worse than LSDVb.

In cases (ii) and (iv), however, the relative frequency of minimum sumMSE regression occuring at GMMs3 is low (16% and 10% respectively), and the system GMM performs worse than LSDVb. As shown in Table 2.5, in both cases, the relative frequency is not higher than 50% for any GMMsj.

Table 2.5: Occurrence of regression with smallest sumMSE at GMMsj : discussion

Case:	(i)	(ii)	(iii)	(iv)	(v)
GMMs1	16%	20%	10%	5%	7%
GMMs2	24%	24%	23%	10%	28%
GMMs3	60%	16%	67%	10%	65%
GMMs4	—	9%	—	10%	—
GMMs5	—	8%	—	10%	—
GMMs6	—	9%	—	18%	—
GMMs7	—	4%	—	15%	—
GMMs8	—	10%	—	22%	—
Total:	100%	100%	100%	100%	100%

2.3 Estimation Results

As suggested in Section 2.2, system GMM using one lagged dependent variable $y_{i(t-2)}$ as an additional instrument for each period seems to be appropriate in our setting. This means that the results in column (a) of Table 2.2, which are concisely reprinted in Table 2.6, are what we are looking for in this chapter.

Four cross-sections are lost in constructing lags and taking first differences, so that the number of useable observations is 923. τ_{it}^{us} and τ_{it}^{mex} are assumed to be strictly exogenous, and the Arellano and Bond test does not provide evidence to suggest that the assumption of serially uncorrelated errors is inappropriate. In general, when a tariff is lower, more foreign products would come in and the demand for domestic goods would decrease, thus decreasing earnings. The results show that a one percentage point reduction in the current tariffs on imports from the U.S. decreases Quebec manufacturing earnings by about 0.0293%[9], one percentage point reduction in tariffs two periods earlier on imports from the U.S. decreases Quebec manufacturing earnings by about 0.0117%, one percentage point reduction in tariffs one period earlier on imports from Mexico decreases Quebec manufacturing earnings by about 0.0283%, but a one percentage point reduction in tariffs four periods earlier on imports from the U.S. *increases* Quebec manufacturing earnings by about 0.0267%. One possible explanation for the *increasing* effect is the long term stable demand for domestic products because of their high competitiveness. In order to see how the effect of tariff reductions changes when exports are controlled for, Canadian exports to the U.S. and Mexico are included in the specifcation in column (d). It turns out that controlling for exports does not change the results much, and the coefficient estimates of exports are insignificant. Finally, if we define the "long-run" effect of NAFTA on earnings as the effect by one percentage point tariff reduction in each of the current year and previous four years, then our results show that the long-run effect of Canadian tariff reductions for U.S. and Mexico imports is about 0.0143% and 0.0283% respectively.

The effect of NAFTA tariff reductions on Quebec manufacturing earnings, whether it is positive or negative, is statistically significant but economically very small. In analyzing the effect of CUSFTA tariff reductions on Canadian manufacturing earnings, Gaston and Trefler (1997) and Beaulieu (2000) find no statistically significant effect, and Trefler (2004) finds slight earnings gains. Trefler (2004) says a 3% rise in earnings spread over eight years will buy you more than a cup of coffee

[9] Note that the tariffs data for regression is in percentage, but the regression results should be interpreted as the effect of each *percentage point* change in tariffs. The reason for this is that tariffs are changed gradually in reality, rather than one (i.e., 100%) at a time.

Table 2.6: GMM estimation for Quebec manufacturing earnings: 1991-2007

Independent variables[a]	(a)	(d)
$\ln earnings_{i(t-1)}$	.9697 (.0394)	.9869 (.0425)
τ_{it}^{us}	.0293 (.0053)	.0297 (.0057)
$\tau_{i(t-2)}^{us}$	.0117 (.0039)	.0098 (.0047)
$\tau_{i(t-4)}^{us}$	-.0267 (.0057)	-.0275 (.0058)
$\tau_{i(t-1)}^{mex}$	.0283 (.0110)	.0289 (.0126)
exp_{it}^{us}	—	-4.91e-06 (4.79e-06)
exp_{it}^{mex}	—	-.0003 (.0004)
Arellano-Bond test	-1.1771	-1.1932
No. of observations	923	923

[a]Dependent variable is $\ln(earnings)_{it}$. *Notes*: (i) - (iv) are the same as in Table 2.2.

but not at Starbucks. Our results indicate a reduction of earnings of even smaller magnitude.

2.4 Concluding Remarks

Using data over the period 1991-2007, this chapter presents an empirical analysis of the effect of NAFTA tariff reductions on earnings for 71 Quebec manufacturing industries. It chooses the system GMM method to estimate a two-way dynamic panel data model. In order for the system GMM estimator to have best possible properties, it proposes a simulation-based moment selection procedure, uses the procedure to select instruments, and shows that the system GMM estimator is not worse than the LSDVb estimator —the best non-GMM alternative. In other words, we use simulations to choose instruments so that the finite sample properties of GMM are optimized, then we compare this "optimal" GMM estimator with the "best" non-GMM alternative.

It is found that a one percentage point reduction in the current tariffs on imports from the U.S., in tariffs two periods earlier on imports from the U.S., and in tariffs one period earlier on imports from Mexico decreases Quebec manufacturing earnings by about 0.0293%, 0.0117%, and 0.0283% respectively, but a one percentage point reduction in tariffs four periods earlier on imports from the U.S. increases Quebec manufacturing earnings by about 0.0267%. No matter the effect is positive or negative, it is statistically significant but economically very small.

As mentioned earlier, NAFTA has become the world's largest free trade area, and Ontario and Quebec are the industrial and manufacturing centers of Canada. It is meaningful to compare the effect of NAFTA on Quebec manufacturing industries and that on Ontario's, which would be the next step of research. In addition, we found that if moments are properly selected, using system GMM method in some classic empirical examples in the dynamic panel data model literature may result in a very large estimate for the coefficient of the lagged dependent variable, which suggests that there may exist panel unit root. Existing panel unit root test methods verify that panel unit root does exist in the datasets of these examples. There is certainly a lot more to do in the future for dynamic panel data models.

Chapter 3

A Goodness-of-fit Test for Copulas

Copulas are functions that allow modeling dependence between random variables separately from their marginal distributions. Consider two continuous random variables X_1 and X_2 with cdf's F_1 and F_2 and pdf's f_1 and f_2, respectively. Suppose the joint cdf of (X_1, X_2) is H and the joint pdf is h. A copula is a function $C(u, v)$ such that $H = C(F_1, F_2)$ or, in densities, $h = c(F_1, F_2)f_1 f_2$. The marginal densities f_1 and f_2 are now "extracted" from the joint density and the copula density c captures the dependence between X_1 and X_2. Sklar (1959) showed that given H, F_1 and F_2 of continuous variables, there exists a unique C. So, given F_1 and F_2, the choice is which copula C to use.

Let C_θ denote the chosen copula family with dependence parameter(s) θ. Numerous papers have used different copula families in applications from finance (e.g., Patton, 2006; Breymann et al., 2003; Li, 2000), from risk management (e.g., Embrechts et al., 2003, 2002) and from health and labor economics (Smith, 2003; Cameron et al., 2004). Recent theoretical results on parametric and semiparametric estimation of copula-based models are contained in Genest et al. (1995); Joe (2005); Chen and Fan (2006b); Prokhorov and Schmidt (2009b); among others.

But the issue of copula specification testing – clearly relevant in any copula-based application – has not received as much attention in the literature as the estimation problem.

A copula family is correctly specified if, for some θ_o, $C_{\theta_o}(F_1, F_2) = H$. In this paper, we wish to construct a goodness of fit test for copulas using this definition. It would be desirable if such a goodness of fit test did not involve parametric specification of the marginal distributions because if it does, it essentially tests a joint hypothesis of correct copula *and* marginal specifications. It is also desirable that this test be applicable to any copula family without requiring any strategic choices and arbitrary parameters, e.g., the choice of a kernel and a bandwidth. Genest et al. (2009) call tests that have these desirable properties "blanket" goodness of fit tests.

There exist a number of copula goodness-of-fit tests (see Genest et al., 2009; Berg, 2009, for recent surveys). However, only a few are "blanket". For example, Klugman and Parsa (1999) propose tests that involve ad hoc categorization of the data; Fermanian (2005) and Scaillet (2007) propose tests that are based on kernels, weight functions and use the associated smoothing parameters; Panchenko (2005) proposes a test based on a V-statistic, whose asymptotic distribution is unknown and depends on the choice of bandwidth; Prokhorov and Schmidt (2009b) propose a conditional moment test for whether the copula-based score function has zero mean, which depends on parametric marginals and does not distinguish between the correct copula and any other copula that has a zero mean score function. All these tests do not qualify as "blanket".

Genest et al. (2009) report five testing procedures that qualify as "blanket" tests. These tests are based on empirical copula and on Kendall's and Rosenblat's probability integral transformation of the data as in, e.g., Dobric and Schmid (2007); Breymann et al. (2003); Genest and Remillard (2008). These tests are substantially more difficult computationally than the "blanket" test we propose.

Moreover, unlike our test, these tests are not asymptotically pivotal and require a parametric bootstrap procedure to obtain approximate p-values.

The test we propose is based on the information matrix equality which equates the copula Hessian and the outer-product of copula score. In essence this is the White (1982) specification test adapted to the first-step nonparametric estimation of marginal distributions. The first stage affects the asymptotic variance of the estimated Hessian and estimated outer-product in a nontrivial way. In Section 3.2 we show that our test statistic asymptotically has a χ^2 distribution and in the Appendix we provide the necessary adjustments for the first-stage rank estimation. Section 3.1 sets the stage by discussing the connection between copulas and the information matrix equality. In Section 3.3, we conduct a power study of the new test. As an illustration, Section 3.4 tests the goodness-of-fit of the Gaussian copula in a model with two stock indices. Section 3.5 concludes.

3.1 Copulas and Information Matrix Equivalence

Consider an N-dimensional copula $C(u_1, \ldots, u_N)$ and N univariate marginals $F_n(x_n), n = 1, \ldots, N$. Then, by Sklar's theorem, the joint distribution of $(X_1, \ldots, X_N)$ is given by

$$H(x_1, \ldots, x_N) = C(F_1(x_1), \ldots, F_N(x_N)).$$

Assume F_n is continuous, $n = 1, \ldots, N$, so $C(u_1, \ldots, u_n)$ is unique. The joint density of $(X_1, \ldots, X_N)$ is

$$\begin{aligned} h(x_1, \ldots, x_N) &= \left. \frac{\partial^N C(u_1, \ldots, u_N)}{\partial u_1 \ldots \partial u_N} \right|_{u_n = F_n(x_n), n=1,\ldots,N} \prod_{n=1}^{N} f_n(x_n) \\ &= c(F_1(x_1), \ldots, F_N(x_N)) \prod_{n=1}^{N} f_n(x_n), \end{aligned}$$

where $c(u_1, \ldots, u_N)$ is the copula density.

We are interested in goodness-of-fit testing of parametric copula families, so our copulas are parametric. For example, the N-variate Gaussian copula with $\frac{N(N-1)}{2}$ parameters can be written as follows

$$\Phi_N(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_N); R),$$

where Φ_N is the joint distribution function of N standard normal covariates with a given correlation matrix R and Φ^{-1} is the inverse of the standard normal cdf. For Gaussian copulas, the copula parameters are the distinct elements of R. (See Nelsen, 2006; Joe, 1997, for examples of other copula families).

Let subscript θ denote the dependence parameter vector of a copula function and let p denote its dimension. It is well known that if there exists a value θ_o such that $H(x_1, \ldots, x_N) = C_{\theta_o}(F_1(x_1), \ldots, F_N(x))$ then we have a correctly specified likelihood model and, under regularity conditions, the MLE is consistent for θ_o. Moreover, in this case White's (1982) information matrix equivalence theorem holds: the Fisher information matrix can be equivalently calculated as minus the expected Hessian or as the expected outer product of the score function.

We wish to apply the information matrix equivalence theorem to copulas. Assume that the copula-based likelihood is (three times) continuously differentiable and the relevant expectations exist. Let $\mathbb{H}(\theta)$ denote the expected Hessian matrix of $\ln c_\theta$ and let $\mathbb{C}(\theta)$ denote the expected outer product of the corresponding score function. Then,

$$\begin{aligned}
\mathbb{H}(\theta) &= \mathbb{E}\nabla_\theta^2 \ln c_\theta(F_1(x_1), \ldots, F_N(x_N)) \\
\mathbb{C}(\theta) &= \mathbb{E}\nabla_\theta \ln c_\theta(F_1(x_1), \ldots, F_N(x_N))\nabla_\theta' \ln c_\theta(F_1(x_1), \ldots, F_N(x_N)),
\end{aligned}$$

where "∇_θ" denotes derivatives with respect to θ and expectations are with respect to the true distribution H.

White's (1982) information matrix equivalence theorem essentially says that, under correct copula specification,

$$-\mathbb{H}(\theta_o) = \mathbb{C}(\theta_o).$$

Our copula misspecification test uses this equality.

3.2 Test

In practice, θ_o is not observed. Moreover, the matrices $\mathbb{H}(\theta)$ and $\mathbb{C}(\theta)$ contain the marginals F_n which are usually unknown. However, these quantities are easily estimated. In particular, it is common to use the empirical distribution function $\hat{F}_n$ in place of F_n, a consistent estimate $\hat{\theta}$ in place of θ_o, the sample averages $\bar{\mathbb{H}}$ and $\bar{\mathbb{C}}$ in place of the expectations $\mathbb{H}$ and $\mathbb{C}$.

Given T observations $(x_1, \ldots, x_N)$, the empirical distribution function is given by

$$\hat{F}_n(s) = T^{-1} \sum_{t=1}^{T} I\{x_{nt} \leq s\},$$

where $I\{\cdot\}$ is the indicator function and s takes values in the observed set of x_n. Then, $\hat{\theta}$ – a consistent estimator of θ_o sometimes called the Canonical Maximum Likelihood estimator (CMLE) – is the solution to

$$\max_{\theta} \sum_{t=1}^{T} \ln c_\theta(\hat{F}_1(x_{1t}), \ldots, \hat{F}_N(x_{Nt})).$$

The following new notation is used for sample counterparts of $\mathbb{H}$ and $\mathbb{C}$. Let

$$\begin{aligned} \hat{\mathbb{H}}_t(\theta) &= \nabla^2_\theta \ln c_\theta(\hat{F}_1(x_{1t}), \ldots, \hat{F}_N(x_{Nt})), \\ \hat{\mathbb{C}}_t(\theta) &= \nabla_\theta \ln c_\theta(\hat{F}_1(x_{1t}), \ldots, \hat{F}_N(x_{Nt})) \nabla'_\theta \ln c_\theta(\hat{F}_1(x_{1t}), \ldots, \hat{F}_N(x_{Nt})). \end{aligned}$$

Then, the sample equivalents of $\mathbb{H}(\theta)$ and $\mathbb{C}(\theta)$ for arbitrary θ are

$$\begin{aligned} \bar{\mathbb{H}}(\theta) &= T^{-1} \sum_{t=1}^{T} \hat{\mathbb{H}}_t(\theta), \\ \bar{\mathbb{C}}(\theta) &= T^{-1} \sum_{t=1}^{T} \hat{\mathbb{C}}_t(\theta). \end{aligned}$$

The test we propose is based on distinct elements of the testing matrix $\bar{\mathbb{H}}(\hat{\theta}) + \bar{\mathbb{C}}(\hat{\theta})$. Given that the dimension of θ is p, there are $p(p+1)/2$ such elements. Under correct copula specification, these are all zero. So our test is in essence a variant of the likelihood misspecification test of White (1982). What distinguishes our test is that we deal with a semiparametric likelihood specification – a parametric copula and nonparametric marginals – while White (1982) deals with a full but possibly incorrect parametric log-density. Correspondingly, the elements of the White (1982) testing matrix (he calls them "indicators") do not contain empirical marginal distributions as arguments and this precludes direct application of his test statistic in our setting.

White (1982) points out that it is sometimes appropriate to drop some of the indicators because they are identically zero or represent a linear combination of the others. When $p = 1$ – the case of bivariate one-parameter copula – this problem does not arise. Whether it arises in higher dimensional models is a copula-specific question that we do not address in this paper. Assume that no indicators need be dropped.

Following White (1982), define

$$d_t(\theta) = vech(\mathbb{H}_t(\theta) + \mathbb{C}_t(\theta))$$

and

$$\hat{d}_t(\theta) = vech(\hat{\mathbb{H}}_t(\theta) + \hat{\mathbb{C}}_t(\theta))$$

Note that, in our setting, $d_t(\theta)$ depends on the unknown marginals while $\hat{d}_t(\theta)$ uses their empirical counterparts $\hat{F}_n, n = 1, \ldots, N$. Define the indicators of interest

$$\bar{D}_\theta \equiv \bar{D}(\theta) \equiv T^{-1} \sum_{t=1}^{T} \hat{d}_t(\theta).$$

Let $\bar{D}_{\hat{\theta}} = \bar{D}(\hat{\theta})$ and $D_\theta = Ed_t(\theta)$. Note that, under correct specification, $D_{\theta_o} \equiv Ed_t(\theta_o) = 0$.

What is different in the present setting from White (1982) is that nonparametric estimates of the marginals are used to construct the joint density. It is well known that the empirical distribution converges to the true distribution at the rate $\sqrt{T}$ so the CMLE estimate $\hat{\theta}$ that uses empirical distributions $\hat{F}_n$ is still $\sqrt{T}$-consistent. However, the asymptotic variance matrix of $\sqrt{T}\hat{\theta}$ will be affected by the nonparametric estimation of marginals. Therefore, the asymptotic variance of $\sqrt{T}\bar{D}_{\hat{\theta}}$ will also be affected. To derive the proper adjustments to the variance matrix we use the results on semiparametric estimation of Newey (1994) and Chen and Fan (2006b). Specifically, Chen and Fan (2006b) derive the distribution of $\hat{\theta}$ given the empirical estimates $\hat{F}_n, n = 1, \ldots, N$. Our setting is complicated by the fact that the test statistic is a function of both $\hat{\theta}$ and $\hat{F}_n, n = 1, \ldots, N$. The main result is given in the following proposition while the derivation of the asymptotic distribution is deferred to the Appendix.

Proposition 3.1. *Under correct copula specification, the information matrix test statistic*

$$\mathscr{I} = T\bar{D}'_{\hat{\theta}} V_{\theta_o}^{-1} \bar{D}_{\hat{\theta}},$$

where V_{θ_o} *is given in (A.61) in Appendix, is distributed asymptotically as* $\chi^2_{p(p+1)/2}$.

In practice, a consistent estimate of V_{θ_o} will be used. Under correct copula specification, such an estimate can be obtained by replacing θ_o and F_{nt} in (A.61) by their consistent estimates $\hat{\theta}$ and $\hat{F}_{nt}$.

Unlike available alternatives, this test statistic is simple, easy to compute and has a standard asymptotically pivotal distribution. It involves no strategic choices such as the choice of a kernel and associated smoothing parameters or any arbitrary categorization of the data. Essentially this is White's information equivalence test with the complication of a first-step empirical distribution estimation. However, as such, it also inherits a number of drawbacks. One complication is the need to evaluate the third derivative of the log-copula density function. Lancaster (1984) and Chesher (1983) show how to construct simplified versions of the test statistic, which are asymptotically equivalent to White's original statistic but do not use the third order derivatives. Probably the simplest form of the test is TR^2, where R^2 comes from the regression of a vector of ones on

$$\nabla_{\theta_j} \ln c_\theta(\hat{F}_1(x_{1t}), \ldots, \hat{F}_N(x_{Nt})), \quad j = 1, \ldots, p$$

and

$$\nabla^2_{\theta_j \theta_k} \ln c_\theta(\hat{F}_1(x_{1t}), \ldots, \hat{F}_N(x_{Nt}))$$
$$+ \nabla_{\theta_j} \ln c_\theta(\hat{F}_1(x_{1t}), \ldots, \hat{F}_N(x_{Nt})) \nabla_{\theta_k} \ln c_\theta(\hat{F}_1(x_{1t}), \ldots, \hat{F}_N(x_{Nt})),$$

$$j = 1, \ldots, p, \quad k = 1, \ldots, p,$$

evaluated at $\hat{\theta}$.

An important problem is the well-documented poor finite sample properties of the test, especially of the TR^2 form (see, e.g., Taylor, 1987; Hall, 1989; Chesher and Spady, 1991; Davidson and MacKinnon, 1992). Horowitz (1994), for example, points out to large deviations of the finite-sample size of various forms of the White test from their nominal size based on asymptotic critical values and suggests using bootstrapped critical values instead. Of course our test will inherit this problem.

3.3 Power Study

In this section, we study the size and power properties of the test statistic we derived in Proposition 3.1. We remark on how this test compares with other copula goodness-of-fit tests but we do not compare here the various alternative forms of the test statistic derived in the proposition. We start by plotting size-power curves under various copula families (see, e.g., Davidson and MacKinnon, 1998, for a comparison of this and other graphical ways of studying test properties). We generate K realizations of the test statistic $\mathscr{I}$ using a data-generating process (DGP). Denote these simulated values by $\mathscr{I}_j$, $j = 1, \cdots, K$. Our size-power curves are based on the empirical distribution function (EDF) of the bootstrap p-value of $\mathscr{I}_j$, $p_j \equiv p(\mathscr{I}_j)$, i.e. the probability that $\mathscr{I}$ is greater than or equal to $\mathscr{I}_j$ according to its bootstrap distribution. At any point y in the (0, 1) interval, the EDF of the p-values is defined by

$$\hat{F}(y) \equiv \frac{1}{K} \sum_{j=1}^{K} I(p_j \leq y).$$

We choose the following values for y_i, $i = 1, \cdots, m$:

$$y_i = 0.001,\ 0.002,\ \cdots,\ 0.010,\ 0.015,\ \cdots,\ 0.990,\ 0.991,\ \cdots,\ 0.999 \quad (m = 215),$$

where we follow Davidson and MacKinnon (1998) and use a smaller grid near 0 and 1 in order to study the tail behavior more closely.

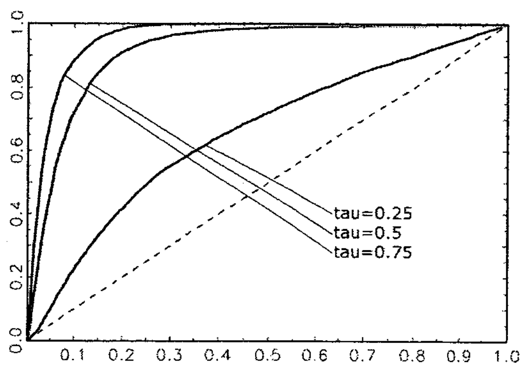
The point of drawing size-power curves is to plot power against true, rather than nominal, size. Given the well-documented poor finite sample size property of the information matrix test, this is useful because we can display the test power in situations when the nominal size is definitely incorrect. Two values of the test statistic are computed: one under the null DGP (H_0) and the other under the alternative DGP (H_1). Let $F(y)$ and $F^*(y)$ be the probabilities of getting a p-value less than y under the null and the alternative, respectively, and let $\hat{F}(y)$ and $\hat{F}^*(y)$ be their empirical counterparts. Given the sample size T, the number of bootstrap replications K and the grid of size m, a size-power curve is the set of points $(\hat{F}(y_i), \hat{F}^*(y_i))$, $i = 1, \cdots, m$, on the unit square where the horizontal axis measures size and the vertical axis measures power. We keep the grid the same, set $K = 10,000$, and vary the sample size T and the strength of dependence in the various null and alternative DGPs we consider. The various null and alternative copula families are selected from the list used by Genest et al. (2009) in a large scale

Monte Carlo study and, as usual, the dependence strength is measured by Kendall's τ. To preserve space we report curves for $T = 200, 300$ and $\tau = 0.25, 0.33, 0.5, 0.75$ only.

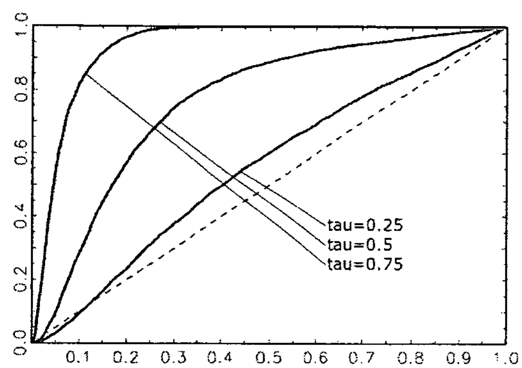
Figure 3.1 shows what happens as we change the strength of dependence holding T fixed at 300. Panel (a) displays the size-power curves under H_0: $C \in$ Normal copula and H_1: $C \in$ Clayton copula, panel (b) displays the curves for H_0: $C \in$ Normal and H_1: $C \in$ Frank, panel (c) is for the test of H_0: $C \in$ Clayton against H_1: $C \in$ Normal, and panel (d) is for H_0: $C \in$ Clayton against H_1: $C \in$ Frank. We can clearly see from the figure that as the strength of dependence increases, the power of the test becomes larger. This agrees with similar observations by Genest et al. (2009) made for other copula goodness-of-fit tests. Interestingly, there are areas on the plots where the test actually has power less than its size. This happens at small enough sizes to make this observation important but the same thing occasionally happens with other "blanket" tests under weak dependence (for $\tau = 0.25$, see, e.g., Genest et al., 2009, Table 1).

Figure 3.2 displays the size-power curves for different null and alternative DGPs holding both T and τ fixed. The set of nulls and alternatives we report includes H_0: Normal vs H_1: Clayton, H_0: Normal vs H_1: Frank, H_0: Clayton vs H_1: Normal, H_0: Clayton vs H_1: Frank. An interesting observation is that the size-adjusted power of the test varies greatly for the different nulls and alternatives – something that has been noted for other tests as well. If we further allow τ to increase holding sample size fixed, the variation in power becomes much smaller. It is interesting to observe that for the tests that involve the Clayton copula under H_0, the test has much more power than for the other models we consider. Again, this interesting observation coincides with results of Genest et al. (2009) obtained for other available "blanket" tests (see their Tables 1-3). Note that the ranking of power of the various tests changes as we change strength of dependence, but the two tests involving the Clayton null remain more powerful than the others.

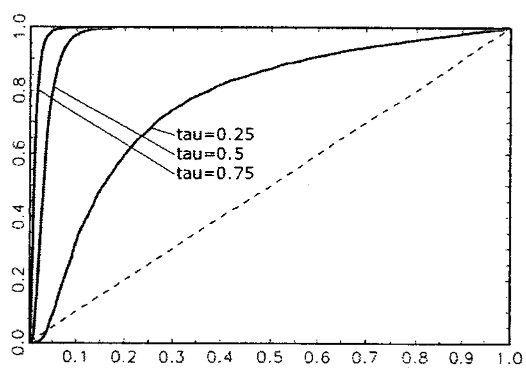
Figure 3.3 shows how the size-power curves shift as the sample size changes from $N = 200$ to $N = 300$. The test in each panel is the same as in Figure 3.1. Not surprisingly, the power increases as the sample size grows. Plots for larger samples (not reported here) illustrate that as the sample size becomes larger, H_0 is rejected with probability approaching one whenever H_1 is true, i.e. these tests are consistent.



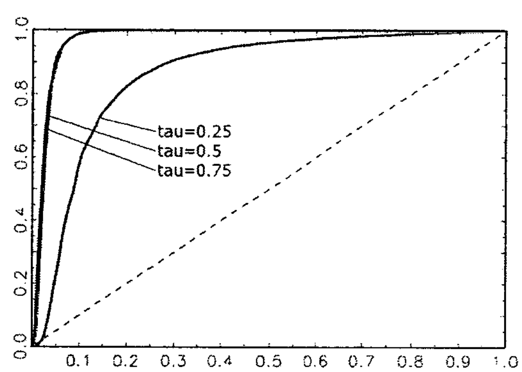
(a) H_0: Normal; H_1: Clayton



(b) H_0: Normal; H_1: Frank copula

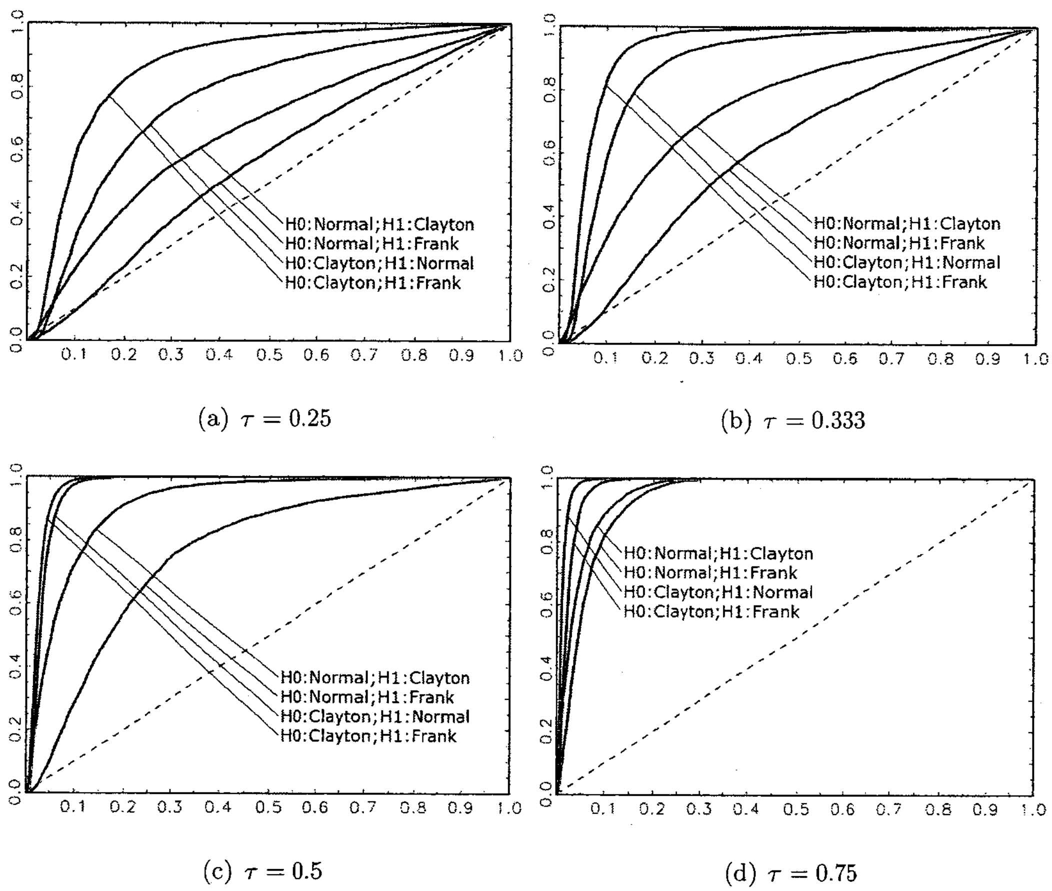


(c) H_0: Clayton; H_1: Normal



(d) H_0: Clayton; H_1: Frank

Figure 3.1: Size-power curves for different levels of dependence: Kendall's $\tau = 0.25, 0.5$ and 0.75. Sample size is $T = 300$.



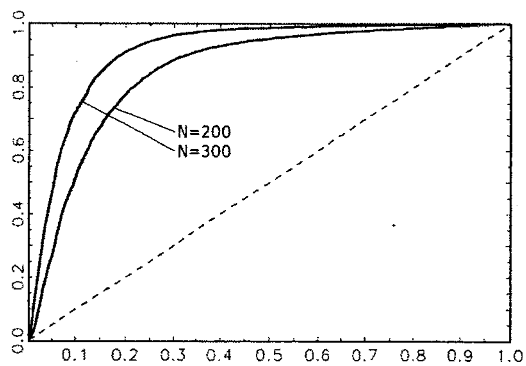
(a) $\tau = 0.25$

(b) $\tau = 0.333$

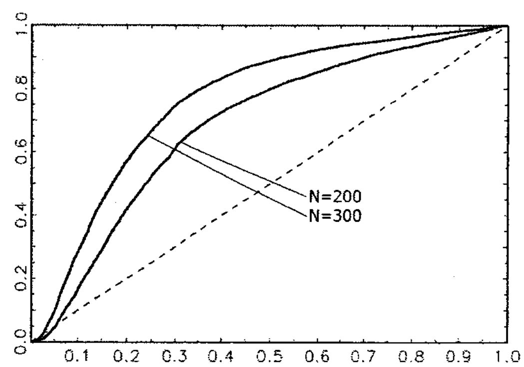
(c) $\tau = 0.5$

(d) $\tau = 0.75$

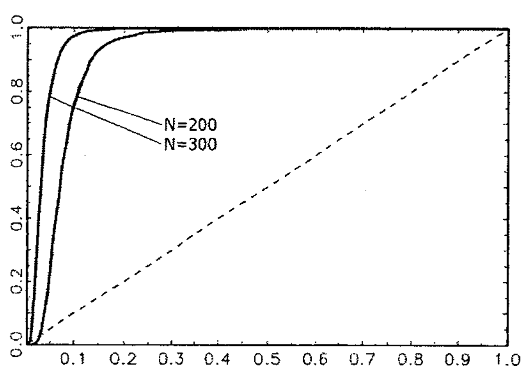
Figure 3.2: Size-power curves for selected copulas. Sample size is $T = 300$.



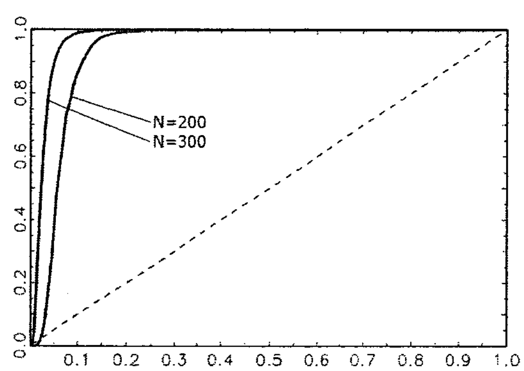
(a) H_0: Normal; H_1: Clayton



(b) H_0: Normal; H_1: Frank



(c) H_0: Clayton; H_1: Normal



(d) H_0: Clayton; H_1: Frank

Figure 3.3: Size-power curves for different sample sizes: $T = 200$ and $T = 300$. Kendall's $\tau = 0.5$.

To compare our test with other "blanket" tests in more detail, we construct a size and power table similar to those reported by Genest et al. (2009). Tables 3.1 and 3.2 report size and power of our test at the 5% significance level for $T = 200$ and $T = 1,000$. As before we also vary Kendall's τ from 0.25 to 0.75. In each row, we report the percentage of rejections of H_0 associated with different tests. For example, when testing for the Normal copula against Clayton at $T = 200$ and $\tau = 0.75$, the chance of the test rejecting the incorrect null is approximately 34.6%.

Table 3.1: Power(Size) for T=200 at nominal size 5%

Copula	True	$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.75$	
under H_0	Copula	Boot.	Asy.	Boot.	Asy.	Boot.	Asy.
Normal	Clayton	4.9(5)	7.7(7)	21.7(5)	34.8(8)	34.6(5)	62.9(10)
	Frank	2.5(5)	4.0(7)	3.8(5)	7.8(7)	16.5(5)	42.0(9)
	Gumbel	6.8(5)	9.6(6)	9.2(5)	18.3(8)	9.1(5)	26.7(10)
Clayton	Normal	1.3(5)	12.2(10)	29.8(5)	85.06(11)	86.1(5)	99.2(11)
	Frank	4.2(5)	26.4(10)	41.6(5)	93.2(11)	64.2(5)	94.6(11)
	Gumbel	8.6(5)	36.5(10)	60.4(5)	96.5(12)	86.4(5)	98.4(10)
Frank	Normal	6.5(5)	8.2(6)	9.2(5)	14.6(9)	3.1(5)	8.0(10)
	Clayton	4.0(5)	5.3(6)	1.5(5)	5.7(9)	2.7(5)	22.4(10)
	Gumbel	4.8(5)	5.8(6)	1.8(5)	5.4(9)	1.0(5)	8.7(10)
Gumbel	Normal	2.9(5)	5.1(8)	1.3(5)	5.2(9)	1.0(5)	9.9(10)
	Clayton	16.9(5)	30.4(8)	37.5(5)	80.0(10)	79.1(5)	97.2(10)
	Frank	3.5(5)	8.0(8)	6.3(5)	31.2(9)	32.7(5)	80.2(10)

Similar to analogous entries for other "blanket" tests, the frequencies reported in Tables 3.1 and 3.2 show that for these sample sizes the test generally holds its nominal size. Compared to equivalent entries in Tables 1 to 3 of Genest et al. (2009), the power of our test statistic is generally lower than that of the more complicated "blanket" tests available in the literature. However, at the sample size equal to 1,000, our test power is usually reasonably high. Similar to other "blanket" tests, the performance of our test varies greatly with the DGPs. For

Table 3.2: Power(Size) for T=1000 at nominal size 5%

Copula under H_0	True Copula	$\tau = 0.25$ Boot.	$\tau = 0.25$ Asy.	$\tau = 0.50$ Boot.	$\tau = 0.50$ Asy.	$\tau = 0.75$ Boot.	$\tau = 0.75$ Asy.
Normal	Clayton	44.0(5)	5(6)	96.9(5)	98.8(7)	93.2(5)	99.0(12)
	Frank	10.7(5)	16.2(7)	65.2(5)	80.0(8)	90.3(5)	98.1(12)
	Gumbel	58.0(5)	63.4(6)	83.4(5)	92.3(8)	78.8(5)	94.7(11)
Clayton	Normal	83.5(5)	87.8(6)	100(5)	100(7)	100(5)	100(7)
	Frank	98.6(5)	99.3(7)	100(5)	100(7)	100(5)	100(7)
	Gumbel	99.6(5)	99.8(6)	100(5)	100(7)	100(5)	100(7)
Frank	Normal	10.1(5)	10.7(5)	21.2(5)	24.0(6)	4.3(5)	5.4(6)
	Clayton	8.5(5)	9.6(6)	17.2(5)	19.9(6)	93.5(5)	95.8(6)
	Gumbel	20.2(5)	21.4(5)	14.9(5)	17.3(6)	53.8(5)	64.4(7)
Gumbel	Normal	8.3(5)	9.4(6)	20.9(5)	25.7(6)	68.3(5)	72.8(6)
	Clayton	98.2(5)	98.6(6)	100(5)	100(6)	100(5)	100(6)
	Frank	50.8(5)	53.7(6)	99.2(5)	99.5(6)	100(5)	100(6)

some combinations of copulas under the null hypothesis and the alternative, the test's power is remarkably low. For example, if the null hypothesis is Frank and the true copula is Normal, the power of our test at $T = 1,000$ is as low as 4-6% even for $\tau = 0.75$. Interestingly, the power of other "blanket" tests is not very high for some combinations either, and for some combinations of copulas and some sample sizes, Genest et al. (2009) report even lower percentages of rejection. In such cases, the results of more than one "blanket" test should probably be considered together.

3.4 Application

To demonstrate how the test procedure in Section 3.2 can be applied in practice, in this section we test whether the bivariate Gaussian copula is appropriate for modeling dependence between an American and an European stock index. The power study demonstrated that the proposed test of the null of Normal copula has

Table 3.3: Summary statistics of returns series

	FTSE	DJIA
mean	0.0001	-.0001
st.d.	0.107	0.103
m_3	0.104	0.020
m_4	6.101	6.590
$Q(20)$	52.97	33.30

power against commonly used alternatives such as the Clayton, Frank and Gumbel copulas.

The two time series we use are FTSE100 and DJIA closing quotes from June 26, 2000 to June 23, 2008. There are 1972 pairs of returns once holidays are eliminated. Table 3.3 contains descriptive statistics of the returns. The statistics we use are third (m_3) and fourth (m_4) central sample moments and the Ljung-Box Q test statistics for testing autocorrelation of up to 20 lags in returns [$Q(20)$] and squared returns [$Q^2(20)$]. Both return series display excess kurtosis and FTSE returns are a bit more skewed than DJIA.

We first apply an AR-GARCH filter to the return data. As shown in Table 3.4, this accounts for most of observed autocorrelation in returns and squared returns. The preferred AR-GARCH models contain up to one lag in the conditional mean equation and a GARCH (1,1) in the conditional variance with Normal innovations (allowing for Student-t innovations resulted in a relatively high estimate of the degrees of freedom (over 9) and did not improve the fit substantially). Table 3.4 reports the results of the AR-GARCH modeling.

The test statistic is reported in Table 3.5. It is based on the residuals from the AR-GARCH models. In principle, this prefiltering should affect the second step estimation and an adjustment should be required to account for that. However,

Table 3.4: AR-GARCH estimates and standard errors

	FTSE	DJIA
μ	-0.0006(0.0004)	-0.0007(0.0004)
AR(1)	-0.0711(0.0230)	-0.0455(0.0221)
ω	0.0000(0.0000)	0.0000(0.0000)
α	0.1158(0.0183)	0.0737(0.0167)
β	0.8742(0.0207)	0.9192(0.0196)
ll	6397.001	6438.337
m_3	-0.144	-0.096
m_4	3.349	3.724
$Q(20)$	22.388	21.442
$Q^2(20)$	16.095	31.774

Table 3.5: Testing the Gaussian copula

$\hat{\theta}$	0.4830(0.0188)
$\mathscr{I}$	3.8800
$p-value$ for $\mathscr{I}$	0.0489

Chen and Fan (2006a) show that the limiting distribution of the copula parameter is not affected by the estimation of dynamic parameters, although as before it is affected by the nonparametric estimation of marginal distributions. So, in this case, the prefiltering is innocuous.

For the bivariate Gaussian copula, the estimated parameter θ is simply the sample correlation between the margins of the bivariate normal distribution used to construct the copula. The parameter estimate is large, positive and significant. The test statistic reported in Table 3.5 is quite large. At the 5% level, we reject the hypothesis that the Gaussian copula is appropriate to model dependence between the two time series. This is a weak rejection (we would not reject at the 1% level, for example), which is surprising given other empirical evidence against using the Gaussian copula in financial applications due to the restrictions this places on symmetry and tail dependence in the data.

3.5 Concluding remarks

We have proposed a new goodness-of-fit test for copulas and have shown that it has reasonable properties in samples as small as 1,000. The main advantage of the test is its simplicity. Basically, it is the well-studied White specification test adapted to a two-step semiparametric estimation. As such, it inherits White test's benefits and costs. The most costly feature of the test is its poor behavior in samples smaller than 1,000. Besides the benefit of simplicity, White's test has many asymptotically equivalent forms, some of which are derived specifically to make finite sample behavior more appealing (see, e.g., Davidson and MacKinnon, 1992). Overall, the balance of costs and benefits speaks, we believe, in favor of this copula goodness-of-fit test, especially in large sample settings of a financial application, similar to the one we considered.

References

ABOWD, J. M. AND D. CARD (1987): "Intertemporal Labor Supply and Long-Term Employment Contracts," *The American Economic Review*, 77, 50–68.

——— (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57, 411–445.

ANDREWS, D. W. K. AND B. LU (2001): "Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models," *Journal of Econometrics*, 101, 123–164.

ARELLANO, M. (2003): *Panel Data Econometrics*, Oxford ; New York : Oxford University Press.

ARELLANO, M. AND S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies*, 58, 277–297.

ARELLANO, M. AND O. BOVER (1995): "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29–51.

BAKER, M. (1997): "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings," *Journal of Labor Economics*, 15, 338–375.

BAKER, M. AND G. SOLON (2003): "Earnings Dynamics and Inequality among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records," *Journal of Labor Economics*, 21, 289–321.

BALTAGI, B. H. (2005): *Econometric Analysis of Panel Data*, Chichester ; Hoboken, NJ : J. Wiley & Sons, 3th ed.

Barndorff-Nielsen, O. and D. R. Cox (1979): "Edgeworth and Saddle-Point Approximations with Statistical Applications," *Journal of the Royal Statistical Society. Series B (Methodological)*, 41, 279–312.

Bartlett, M. S. (1937): "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 160, 268–282.

Beaulieu, E. (2000): "The Canada-U.S. Free Trade Agreement and Labour Market Adjustment in Canada," *The Canadian Journal of Economics*, 33, 540–563.

Berg, D. (2009): "Copula goodness-of-fit testing: an overview and power comparison," *The European Journal of Finance*, 15, 675 – 701.

Blundell, R. and S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

Blundell, R., S. Bond, and F. Windmeijer (2000): "Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator," in *Advances in Econometrics*, vol. 15, 53–91.

Breymann, W., A. Dias, and P. Embrechts (2003): "Dependence structures for multivariate high-frequency data in finance," *Quantitative Finance*, 3, 1–14.

Browne, M. W. (1984): "Asymptotically Distribution-free Methods for the Analysis of Covariance Structures," *British Journal of Mathematical and Statistical Psychology*, 37, 62–83.

Bun, M. J. G. and J. F. Kiviet (2006): "The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models," *Journal of Econometrics*, 132, 409–444.

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 (2006a): "Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification." *Journal of Econometrics*, 135, 125–154.

——— (2006b): "Estimation of copula-based semiparametric time series models," *Journal of Econometrics*, 130, 307–335.

CHESHER, A. (1983): "The information matrix test : Simplified calculation via a score test interpretation," *Economics Letters*, 13, 45–48.

CHESHER, A. AND R. SPADY (1991): "Asymptotic Expansions of the Information Matrix Test Statistic," *Econometrica*, 59, 787–815.

CHOU, C.-P., P. M. BENTLER, AND A. SATORRA (1991): "Scaled Test Statistics and Robust Standard Errors for Non-normal Data in Covariance Structure Analysis: A Monte Carlo Study," *British Journal of Mathematical and Statistical Psychology*, 44, 347–357.

CLARK, T. E. (1996): "Small-Sample Properties of Estimators of Nonlinear Models of Covariance Structure," *Journal of Business & Economic Statistics*, 14, 367–373.

CLAUSING, K. A. (2001): "Trade Creation and Trade Diversion in the Canada - United States Free Trade Agreement," *The Canadian Journal of Economics*, 34, 677–696.

CRIBARI-NETO, F. AND G. M. CORDEIRO (1996): "On Bartlett and Bartlett-Type Corrections," *Econometric Reviews*, 15, 339–367.

DAHLBERG, M. AND E. JOHANSSON (2000): "An Examination of the Dynamic Behaviour of Local Governments Using GMM Bootstrapping Methods," *Journal of Applied Econometrics*, 15, 401–416.

DAVIDSON, R. AND J. G. MACKINNON (1992): "A New Form of the Information Matrix Test," *Econometrica*, 60, 145–157.

——— (1998): "Graphical Methods for Investigating the Size and Power of Hypothesis Tests," *The Manchester School*, 66, 1–26.

DOBRIC, J. AND F. SCHMID (2007): "A goodness of fit test for copulas based on Rosenblatt's transformation," *Computational Statistics and Data Analysis*, 51, 4633–4642.

DORAN, H. E. AND P. SCHMIDT (2006): "GMM Estimators with Improved Finite Sample Properties Using Principal Components of the Weighting Matrix, with an Application to the Dynamic Panel Data Model." *Journal of Econometrics*, 133, 387–409.

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.

EVERAERT, G. AND L. POZZI (2007): "Bootstrap-Based Bias Correction for Dynamic Panels," *Journal of Economic Dynamics and Control*, 31, 1160–1184.

FAITC (2001): "NAFTA at Seven: Building on a North American Partnership," Foreign Affairs and International Trade Canada, http://www.international.gc.ca/trade-agreements-accords-commerciaux/agr-acc/nafta-alena/nafta_7.aspx?lang=en.

FERMANIAN, J.-D. (2005): "Goodness-of-fit tests for copulas," *Journal of Multivariate Analysis*, 95, 119–152, doi: DOI: 10.1016/j.jmva.2004.07.004.

FUKAO, K., T. OKUBO, AND R. M. STERN (2003): "An Econometric Analysis of Trade Diversion Under NAFTA," *The North American Journal of Economics and Finance*, 14, 3–24.

GASTON, N. AND D. TREFLER (1997): "The Labour Market Consequences of the Canada-U.S. Free Trade Agreement," *The Canadian Journal of Economics*, 30, 18–41.

GENEST, C., K. GHOUDI, AND L.-P. RIVEST (1995): "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrika*, 82, 543–552.

GENEST, C. AND B. REMILLARD (2008): "Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models," *Annales de l'Institut Henri Poincare - Probabilites et Statistiques*, 44, 10961127.

GENEST, C., B. REMILLARD, AND D. BEAUDOIN (2009): "Goodness-of-fit tests for copulas: A review and a power study," *Insurance: Mathematics and Economics*, 44, 199–213.

HALL, A. (1989): "On the calculation of the information matrix test in the normal linear regression model," *Economics Letters*, 29, 31–35.

HALL, P. (1992): *The Bootstrap and the Edgeworth Expansion*, New York: Springer-Verlag.

HAN, C. AND P. C. B. PHILLIPS (2006): "GMM with Many Moment Conditions," *Econometrica*, 74, 147–192.

Hansen, B. E. (2006): "Edgeworth Expansions for the Wald and GMM Statistics for Nonlinear Restrictions," in *Econometric theory and practice: frontiers of analysis and applied research*, ed. by D. Corbae, S. N. Durlauf, and B. E. Hansen, Cambridge; New York: Cambridge University Press, 9–35.

Hausman, J. and G. Kuersteiner (2008): "Difference In Difference Meets Generalized Least Squares: Higher Order Properties of Hypotheses Tests," *Journal of Econometrics*, 144, 371–391.

Hayakawa, K. (2007): "Small Sample Bias Properties of the System GMM Estimator in Dynamic Panel Data Models," *Economics Letters*, 95, 32–38.

Head, K. and J. Ries (1999a): "Can Small-Country Manufacturing Survive Trade Liberalization? Evidence from the Canada-U.S. Free Trade Agreement," Perspectives on North American Free Research Publication No. 1, Industry Canada.

——— (1999b): "Rationalization Effects of Tariff Reductions," *Journal of International Economics*, 47, 295–320.

Herzog, W., A. Boomsma, and S. Reinecke (2007): "The Model-Size Effect on Traditional and Modified Tests of Covariance Structures," *Structural Equation Modeling*, 14, 361–390.

Holtz-Eakin, D. (1988): "Testing for Individual Effects in Autoregressive Models," *Journal of Econometrics*, 39, 297–307.

Hoogland, J. J. and A. Boomsma (1998): "Robustness Studies in Covariance Structure Modeling," *Sociological Methods & Research*, 26, 329–367.

Horowitz, J. L. (1994): "Bootstrap-based critical values for the information matrix test," *Journal of Econometrics*, 61, 395–411.

Joe, H. (1997): *Multivariate models and dependence concepts*, vol. 73 of *Monographs on Statistics and Applied Probability*, Chapman and Hall.

——— (2005): "Asymptotic efficiency of the two-stage estimation method for copula-based models," *Journal of Multivariate Analysis*, 94, 401–419.

Kallenberg, W. C. M. (1993): "Interpretation and Manipulation of Edgeworth Expansions," *Annals of the Institute of Statistical Mathematics*, 45, 341–351.

KENNY, D. A. AND D. B. MCCOACH (2003): "Effect of the Number of Variables on Measures of Fit in Structural Equation Modeling," *Structural Equation Modeling*, 10, 333–351.

KLUGMAN, S. A. AND R. PARSA (1999): "Fitting bivariate loss distributions with copulas," *Insurance: Mathematics and Economics*, 24, 139–148.

KOLLO, T. AND D. V. ROSEN (2005): *Advanced Multivariate Statistics with Matrices*, Dordrecht : Springer.

KORIN, B. P. (1968): "On the Distribution of a Statistic Used for Testing a Covariance Matrix," *Biometrika*, 55, 171–178.

LANCASTER, T. (1984): "The Covariance Matrix of the Information Matrix Test," *Econometrica*, 52, 1051–1053.

LI, D. X. (2000): "On Default Correlation: A Copula Function Approach," *Journal of Fixed Income*, 9, 43–54.

LINTON, O. (2002): "Edgeworth Approximations for Semiparametric Instrumental Variable Estimators and Test Statistics," *Journal of Econometrics*, 106, 325–368.

MACURDY, T. E. (1982): "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, 18, 83–114.

MAGEE, L. (1989): "An Edgeworth Test Size Correction for the Linear Model with AR(1) Errors," *Econometrica*, 57, 661–674.

MUTHEN, B. AND D. KAPLAN (1991): "A Comparison of Some Methodologies for the Factor Analysis of Non-normal Likert Variables: A Note on the Size of the Model," *British Journal of Mathematical and Statistical Psychology*, 45, 19–30.

NAGARSENKER, B. N. AND K. C. S. PILLAI (1973): "Distribution of the Likelihood Ratio Criterion for Testing a Hypothesis Specifying a Covariance Matrix," *Biometrika*, 60, 359–364.

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

NEWEY, W. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

NEWEY, W. K. AND D. MCFADDEN (1994): "Chapter 36 Large Sample Estimation and Hypothesis Testing," Elsevier, vol. 4 of *Handbook of Econometrics*, 2111–2245.

NEWEY, W. K. AND K. D. WEST (1987): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28, 777–787.

NEWEY, W. K. AND F. WINDMEIJER (2009): "Generalized Method of Moments With Many Weak Moment Conditions," *Econometrica*, 77, 687–719.

OKUI, R. (2009): "The Optimal Choice of Moments in Dynamic Panel Data Models," *Journal of Econometrics*, 151, 1–16.

PAGAN, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221–247.

PANCHENKO, V. (2005): "Goodness-of-fit test for copulas," *Physica A: Statistical Mechanics and its Applications*, 355, 176–182.

PATTON, A. J. (2006): "Modelling asymmetric exchange rate dependence," *International Economic Review*, 47, 527–556.

PESARAN, M. H. (2007): "A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence," *Journal of Applied Econometrics*, 22, 265–312.

PHILLIPS, P. AND J. Y. PARK (1988): "On The Formulation of Wald Tests of Nonlinear Restrictions," *Econometrica*, 56, 1065–1083.

PHILLIPS, P. C. B. (1977): "A General Theorem in the Theory of Asymptotic Expansions as Approximations to the Finite Sample Distributions of Econometric Estimators," *Econometrica*, 45, 1517–1534.

——— (1978): "Edgeworth and Saddlepoint Approximations in the First-Order Noncircular Autoregression," *Biometrika*, 65, 91–98.

——— (1983): "ERA's: A New Approach to Small Sample Theory," *Econometrica*, 51, 1505–1525.

POSTREL, V. (2005): "Economic Scene: What Happened When Two Rich Countries Liberalized Trade? Pain, and then Gain." *The New York Times*, January 27.

PROKHOROV, A. AND P. SCHMIDT (2009a): "GMM Redundancy Results for General Missing Data Problems," *Journal of Econometrics*, 151, 47–55.

——— (2009b): "Likelihood-based estimation in a panel setting: robustness, redundancy and validity of copulas," *Journal of Econometrics*, 153, 93–104.

REID, N. (1991): "Approximations and Asymptotics," in *Statistical Theory and Modeling: In Honour of Sir David Cox, FRS*, ed. by D. Hinkley, N. Reid, and E. Snell, London: Chapman and Hall.

ROMALIS, J. (2007): "NAFTA's and CUSFTA's Impact on International Trade," *Review of Economics and Statistics*, 89, 416–435.

ROTHENBERG, T. J. (1984): "Chapter 15 Approximating the Distributions of Econometric Estimators and Test Statistics," Elsevier, vol. 2 of *Handbook of Econometrics*, 881–935.

SARGAN, J. D. AND S. E. SATCHELL (1986): "A Theorem of Validity for Edgeworth Expansions," *Econometrica*, 54, 189–213.

SATORRA, A. AND P. BENTLER (2001): "A scaled Difference Chi-square Test Statistic for Moment Structure Analysis," *Psychometrika*, 66, 507–514.

SCAILLET, O. (2007): "Kernel-based goodness-of-fit tests for copulas with fixed smoothing parameters," *Journal of Multivariate Analysis*, 98, 533–543.

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 Archimedean copulas," *The Econometrics Journal*, 6, 99–123.

SUGIURA, N. (1969): "Asymptotic Expansions of the Distributions of the Likelihood Ratio Criteria for Covariance Matrix," *The Annals of Mathematical Statistics*, 40, 2051–2063.

TAYLOR, L. W. (1987): "The size bias of White's information matrix test," *Economics Letters*, 24, 63–67.

TOPEL, R. H. AND M. P. WARD (1992): "Job Mobility and the Careers of Young Men," *The Quarterly Journal of Economics*, 107, 439–479.

TREFLER, D. (2004): "The Long and Short of the Canada-U.S. Free Trade Agreement," *The American Economic Review*, 94, 870–895.

WHITE, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1–26.

WINDMEIJER, F. (2005): "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators," *Journal of Econometrics*, 126, 25–51.

WOOLDRIDGE, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Mass. : MIT Press.

YANAGIHARA, H., C. MATSUMOTO, AND T. TONDA (2004): "Asymptotic Expansion of the Null Distribution of the Modified normal likelihood ratio criterion for testing $\Sigma = \Sigma_0$ under Nonnormality," *Hiroshima Mathematical Journal*, 34, 81–100.

YUAN, K.-H. AND P. M. BENTLER (1997): "Mean and Covariance Structure Analysis: Theoretical and Practical Improvements," *Journal of the American Statistical Association*, 92, 767–774.

——— (1998): "Normal Theory Based Test Statistics in Structural Equation Modeling," *British Journal of Mathematical and Statistical Psychology*, 51, 289–309.

——— (1999): "F Tests for Mean and Covariance Structure Analysis," *Journal of Educational and Behavioral Statistics*, 24, 225–243.

ZILIAK, J. P. (1997): "Efficient Estimation with Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators," *Journal of Business & Economic Statistics*, 15, 419–431.

Appendix A

Proofs and Data Descriptions

Theorem 3.1.1 of Kollo and Rosen (2005)

Let $\{x_n\}$ and $\{\varepsilon_n\}$ be sequences of random p-vectors and positive numbers, respectively, and let $x_n - x_0 = o_p(\varepsilon_n)$, where $\varepsilon_n \to 0$ as $n \to \infty$. If the function $f(x)$ from $\mathbb{R}^p$ to $\mathbb{R}^s$ has continuous partial derivatives up to the order $(\mathcal{M}+1)$ in a neighborhood $\mathcal{D}$ of a point x_0, then the function $f(x)$ can be expanded at the point x_0 into the Taylor series

$$f(x) = f(x_0) + \sum_{k=1}^{\mathcal{M}} \frac{1}{k!} \left(I_s \otimes (x - x_0)^{\otimes(k-1)}\right)' \left(\frac{d^k f(x)}{dx^k}\right)'_{x=x_0} (x - x_0) + o(\rho^{\mathcal{M}}(x, x_0)),$$

where the Kroneckerian power $A^{\otimes k}$ for any matrix A is given by $A^{\otimes k} = \underbrace{A \otimes \cdots \otimes A}_{\text{k times}}$ with $A^{\otimes 0} = 1$, $\rho(.,.)$ is the Euclidean distance in $\mathbb{R}^p$, and the matrix derivative for any matrices Y and X is given by $\frac{d^kY}{dX^k} = \frac{d}{dX}\left(\frac{d^{k-1}Y}{dX^{k-1}}\right)$ with $\frac{dY}{dX} \equiv \frac{dvec'Y}{dvecX}$; and

$$f(x_n) = f(x_0) + \sum_{k=1}^{\mathcal{M}} \frac{1}{k!} \left(I_s \otimes (x_n - x_0)^{\otimes(k-1)}\right)' \left(\frac{d^k f(x_n)}{dx_n^k}\right)'_{x_n=x_0} (x_n - x_0) + o_p(\varepsilon_n^{\mathcal{M}}).$$

Proofs of Chapter 1

Proof of Theorem 1.1: Write (1.9) as

$$DM \cong 1_{DM} + 2_{DM} + 3_{DM} + 4_{DM}, \tag{A.1}$$

where,

$$\begin{aligned}
1_{DM} &= vec'G_N(\hat{\theta}_N)M_1 vecG_N(\hat{\theta}_N),\\
2_{DM} &= \mathbb{M}'_N(\hat{\theta}_N)M_2 vecD_N(\hat{\theta}_N),\\
3_{DM} &= -N^{-1/2}vec'G_N(\hat{\theta}_N)M_3 vecD_N(\hat{\theta}_N),\\
4_{DM} &= N^{-1}\frac{1}{4}vec'D_N(\hat{\theta}_N)M_4 vecD_N(\hat{\theta}_N).
\end{aligned}$$

Taking Taylor expansions of $\mathbb{M}_N(\hat{\theta}_N)$, $vec\,G_N(\hat{\theta}_N)$ and $vec\,D_N(\hat{\theta}_N)$ about θ_0 and using (1.5) and (1.6), we have

$$\begin{aligned}
\underset{k\times 1}{\mathbb{M}_N(\hat{\theta}_N)} &= \mathbb{M}_N(\theta_0) + G'(\hat{\theta}_N-\theta_0) + \frac{1}{2}[I_k\otimes(\hat{\theta}_N-\theta_0)']D'(\hat{\theta}_N-\theta_0) + o_p(N^{-1})\\
&= -N^{-1/2}\bar{q} + N^{-1/2}G'B^{-1}G\bar{q} + N^{-1}\frac{1}{2}(I_k\otimes\bar{q}'G'B^{-1})D'B^{-1}G\bar{q} + o_p(N^{-1}),\\
\underset{pk\times 1}{vec\,G_N(\hat{\theta}_N)} &= vecG + D'(\hat{\theta}_N-\theta_0) + \frac{1}{2}[I_{pk}\otimes(\hat{\theta}_N-\theta_0)']C'(\hat{\theta}_N-\theta_0) + o_p(N^{-1})\\
&= vecG + N^{-1/2}D'B^{-1}G\bar{q} + N^{-1}\frac{1}{2}(I_{pk}\otimes\bar{q}'G'B^{-1})C'B^{-1}G\bar{q} + o_p(N^{-1}),\\
\underset{p^2k\times 1}{vec\,D_N(\hat{\theta}_N)} &= vecD + C'(\hat{\theta}_N-\theta_0) + o_p(N^{-1/2})\\
&= vecD + N^{-1/2}C'B^{-1}G\bar{q} + o_p(N^{-1/2}).
\end{aligned}$$

Note that we do not need to expand $vecD_N(\hat{\theta}_N)$ further for our purpose. Substituting these expressions into the terms of (A.1) gives:

$$\begin{aligned}
1_{DM} &= vec'G_N(\hat{\theta}_N)M_1 vecG_N(\hat{\theta}_N)\\
&= vec'GM_1vecG + N^{-1/2}2\bar{q}'G'B^{-1}DM_1vecG\\
&\quad + N^{-1}[\bar{q}'G'B^{-1}DM_1D'B^{-1}G\bar{q} + \bar{q}'G'B^{-1}C(I_{pk}\otimes B^{-1}G\bar{q})M_1vecG] \qquad (A.2)\\
&\quad + o_p(N^{-1})\\
&= \bar{q}'P\bar{q} + N^{-1/2}u_1(\bar{q}) + N^{-1}v_1(\bar{q}) + o_p(N^{-1}),
\end{aligned}$$

where

$$\underset{k\times k}{P} \equiv G'\mathbb{H}G$$

is a projection matrix, and

$$u_1(\bar{q}) = 2\bar{q}'G'B^{-1}DM_1 vecG,$$

$$v_1(\bar{q}) = \bar{q}'G'B^{-1}DM_1D'B^{-1}G\bar{q} + \bar{q}'G'B^{-1}C(I_{pk} \otimes B^{-1}G\bar{q})M_1 vecG;$$

$$\begin{aligned}
2_{DM} &= \mathbb{M}'_N(\hat{\theta}_N)M_2 vecD_N(\hat{\theta}_N) \\
&= -N^{-1/2}\bar{q}'M_2 vecD - N^{-1}\bar{q}'M_2C'B^{-1}G\bar{q} \\
&\quad + N^{-1/2}\bar{q}'G'B^{-1}GM_2 vecD + N^{-1}\bar{q}'G'B^{-1}GM_2C'B^{-1}G\bar{q} \\
&\quad + N^{-1}\frac{1}{2}\bar{q}'G'B^{-1}D(I_k \otimes B^{-1}G\bar{q})M_2 vecD + o_p(N^{-1}) \\
&= N^{-1/2}(\bar{q}'G'B^{-1}M_2 vecD - \bar{q}'M_2 vecD) \\
&\quad + N^{-1}[\bar{q}'G'B^{-1}GM_2C'B^{-1}G\bar{q} - \bar{q}'M_2C'B^{-1}G\bar{q} \\
&\qquad + \frac{1}{2}\bar{q}'G'B^{-1}D(I_k \otimes B^{-1}G\bar{q})M_2 vecD] + o_p(N^{-1}) \\
&= N^{-1/2}u_2(\bar{q}) + N^{-1}v_2(\bar{q}) + o_p(N^{-1}),
\end{aligned} \tag{A.3}$$

where

$$\begin{aligned}
u_2(\bar{q}) &= \bar{q}'G'B^{-1}GM_2 vecD - \bar{q}'M_2 vecD \\
&= \bar{q}'(G'B^{-1}G - I_k)M_2 vecD, \\
v_2(\bar{q}) &= \bar{q}'G'B^{-1}GM_2C'B^{-1}G\bar{q} - \bar{q}'M_2C'B^{-1}G\bar{q} + \frac{1}{2}\bar{q}'G'B^{-1}D(I_k \otimes B^{-1}G\bar{q})M_2 vecD \\
&= \bar{q}'(G'B^{-1}G - I_k)M_2C'B^{-1}G\bar{q} + \frac{1}{2}\bar{q}'G'B^{-1}D(I_k \otimes B^{-1}G\bar{q})M_2 vecD;
\end{aligned}$$

$$\begin{aligned}
3_{DM} &= -N^{-1/2}vec'G_N(\hat{\theta}_N)M_3 vecD_N(\hat{\theta}_N) \\
&= -N^{-1/2}vec'GM_3 vecD - N^{-1}vec'GM_3C'B^{-1}G\bar{q} - N^{-1}\bar{q}'G'B^{-1}DM_3 vecD + o_p(N^{-1}) \\
&= N^{-1/2}u_3(\bar{q}) + N^{-1}v_3(\bar{q}) + o_p(N^{-1}),
\end{aligned} \tag{A.4}$$

where

$$\begin{aligned}
u_3(\bar{q}) &= -vec'GM_3 vecD, \\
v_3(\bar{q}) &= -vec'GM_3C'B^{-1}G\bar{q} - \bar{q}'G'B^{-1}DM_3 vecD \\
&= -\bar{q}'G'B^{-1}CM'_3 vecG - \bar{q}'G'B^{-1}DM_3 vecD;
\end{aligned}$$

$$\begin{aligned}
4_{DM} &= N^{-1}\frac{1}{4}vec'D_N(\hat{\theta}_N)M_4 vecD_N(\hat{\theta}_N) \\
&= N^{-1}\frac{1}{4}vec'DM_4 vecD + o_p(N^{-1}) \\
&= N^{-1}v_4(\bar{q}) + o_p(N^{-1}),
\end{aligned} \tag{A.5}$$

where

$$v_4(\bar{q}) = \frac{1}{4} vec' D M_4 vec D.$$

Finally, collecting the terms (A.2)-(A.5) gives equation (1.10). □

Proof of Lemma 1.1: From Theorem 1.1, if $u_i(\bar{q})$ $(i = 1, 2, 3)$ and $v_i(\bar{q})$ $(i = 1, 2, 3, 4)$ could be rewritten as

$$u_i(\bar{q}) = vec' J_i(\bar{q} \otimes \bar{q} \otimes \bar{q}), \tag{A.6}$$

$$v_i(\bar{q}) = tr[L_i(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')], \tag{A.7}$$

then,

$$u(\bar{q}) = vec' J(\bar{q} \otimes \bar{q} \otimes \bar{q}),$$

$$v(\bar{q}) = tr[L(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')],$$

where

$$vec J = vec J_1 + vec J_2 + vec J_3,$$

and

$$L = L_1 + L_2 + L_3 + L_4.$$

Therefore, the proof is reduced to showing (A.6) and (A.7).

Using

$$(A \otimes C)(B \otimes D) = (AB) \otimes (CD),$$

$$K_{p,q} vec A = vec(A'),$$

$$A \otimes B = K_{p,r}(B \otimes A) K_{s,q},$$

for $A : p \times q$ and $B : r \times s$ where K is the commutation matrix, we can rewrite (1.12):

$$\begin{aligned} u_1(\bar{q}) &= 2\bar{q}' G' B^{-1} D (I_k \otimes \mathbb{H} G \bar{q}) vec(\bar{q}' G' \mathbb{H} G) \\ &= 2\bar{q}' G' \mathbb{H} G (I_k \otimes \bar{q}' G' \mathbb{H})(\bar{q}' G' B^{-1} \otimes I_{pk}) vec(D') \\ &= 2\bar{q}' G' \mathbb{H} G (I_k \otimes \bar{q}' G' \mathbb{H})(I_{pk} \otimes \bar{q}' G' B^{-1}) vec D \\ &= 2(\bar{q}' G' \mathbb{H} G \otimes \bar{q}' G' \mathbb{H} \otimes \bar{q}' G' B^{-1}) vec D \\ &= 2(\bar{q}' \otimes \bar{q}' \otimes \bar{q}')(G' \mathbb{H} G \otimes G' \mathbb{H} \otimes G' B^{-1}) vec D \\ &= vec' J_1 (\bar{q} \otimes \bar{q} \otimes \bar{q}), \end{aligned} \tag{A.8}$$

where

$$vecJ_1 = 2(G'\mathbb{H}G \otimes G'\mathbb{H} \otimes G'B^{-1})vecD. \tag{A.9}$$

Let

$$R_1 = (\mathbb{H}G \otimes B^{-1}G)(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')(G'\mathbb{H} \otimes G'B^{-1}), \tag{A.10}$$

partition $vecD$ as

$$\underset{p^2k\times 1}{vecD} = \begin{bmatrix} V_{D1} \\ V_{D2} \\ \vdots \\ V_{Dk} \end{bmatrix} \tag{A.11}$$

where each subvector V_{Di} is $p^2 \times 1$, and let

$$V_D = V_{D1}V_{D1}' + V_{D2}V_{D2}' + \cdots + V_{Dk}V_{Dk}'. \tag{A.12}$$

Then, since

$$\begin{aligned}(I_k \otimes \bar{q}'G'\mathbb{H})D'B^{-1}G\bar{q} &= (I_k \otimes \bar{q}'G'\mathbb{H})(\bar{q}'G'B^{-1} \otimes I_{pk})vec(D') \\ &= (I_k \otimes \bar{q}'G'\mathbb{H})(I_{pk} \otimes \bar{q}'G'B^{-1})vecD \\ &= (I_k \otimes \bar{q}'G'\mathbb{H} \otimes \bar{q}'G'B^{-1})vecD,\end{aligned}$$

the first term of $v_1(\bar{q})$ in (1.15) becomes

$$\begin{aligned}&\bar{q}'G'B^{-1}D(I_k \otimes \mathbb{H}G\bar{q})(I_k \otimes \bar{q}'G'\mathbb{H})D'B^{-1}G\bar{q} \\ &\quad = vec'D(I_k \otimes \mathbb{H}G\bar{q} \otimes B^{-1}G\bar{q})(I_k \otimes \bar{q}'G'\mathbb{H} \otimes \bar{q}'G'B^{-1})vecD \\ &\quad = vec'D(I_k \otimes R_1)vecD \\ &\quad = \begin{bmatrix} V_{D1}' & V_{D2}' & \cdots & V_{Dk}' \end{bmatrix} \begin{bmatrix} R_1 & & & 0 \\ & R_1 & & \\ & & \ddots & \\ 0 & & & R_1 \end{bmatrix} \begin{bmatrix} V_{D1} \\ V_{D2} \\ \vdots \\ V_{Dk} \end{bmatrix} \\ &\quad = V_{D1}'R_1V_{D1} + V_{D2}'R_1V_{D2} + \cdots + V_{Dk}'R_1V_{Dk} \\ &\quad = tr[(V_{D1}V_{D1}' + V_{D2}V_{D2}' + \cdots + V_{Dk}V_{Dk}')R_1] \\ &\quad = tr[V_D(\mathbb{H}G \otimes B^{-1}G)(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')(G'\mathbb{H} \otimes G'B^{-1})] \\ &\quad = tr[(G'\mathbb{H} \otimes G'B^{-1})V_D(\mathbb{H}G \otimes B^{-1}G)(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')].\end{aligned} \tag{A.13}$$

Similarly, let

$$R_2 = (\mathbb{H}G \otimes B^{-1}G)(\bar{q} \otimes \bar{q}), \tag{A.14}$$

$$R_3 = \bar{q}'G'\mathbb{H}, \tag{A.15}$$

partition $G'B^{-1}C$ and $vecG$ as

$$\underset{k\times p^2k}{G'B^{-1}C} = \begin{bmatrix} M_{GC1} & M_{GC2} & \cdots & M_{GCk} \end{bmatrix}, \tag{A.16}$$

$$\underset{pk\times 1}{vecG} = \begin{bmatrix} V_{G1} \\ V_{G2} \\ \vdots \\ V_{Gk} \end{bmatrix}, \tag{A.17}$$

where M_{GCi} and V_{Gi} are $k \times p^2$ and $p \times 1$ respectively, and let

$$M_V = M'_{GC1} \otimes V'_{G1} + M'_{GC2} \otimes V'_{G2} + \cdots + M'_{GCk} \otimes V'_{Gk}. \tag{A.18}$$

Then, since

$$\begin{aligned} \bar{q}'m\bar{q}'M(\bar{q} \otimes \bar{q}) &= m'\bar{q}\bar{q}'M(\bar{q} \otimes \bar{q}) \\ &= [(\bar{q} \otimes \bar{q})'M' \otimes m']vec(\bar{q}\bar{q}') \\ &= (\bar{q} \otimes \bar{q})'(M' \otimes m')(\bar{q} \otimes \bar{q}) \\ &= tr[(M' \otimes m')(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')] \end{aligned}$$

for some vector m and matrix M of appropriate sizes, the second term of $v_1(\bar{q})$ in (1.15) becomes

$$
\begin{aligned}
&\bar{q}'G'B^{-1}C(I_{pk}\otimes B^{-1}G\bar{q})(I_k\otimes \mathbb{H}G\bar{q})(I_k\otimes \bar{q}'G'\mathbb{H})vecG \\
&\quad = \bar{q}'G'B^{-1}C(I_k\otimes R_2)(I_k\otimes R_3)vecG \\
&\quad = \bar{q}'\begin{bmatrix} M_{GC1} & M_{GC2} & \cdots & M_{GCk}\end{bmatrix}\begin{bmatrix} R_2 & & & 0 \\ & R_2 & & \\ & & \ddots & \\ 0 & & & R_2\end{bmatrix}\begin{bmatrix} R_3 & & & 0 \\ & R_3 & & \\ & & \ddots & \\ 0 & & & R_3\end{bmatrix}\begin{bmatrix} V_{G1} \\ V_{G2} \\ \vdots \\ V_{Gk}\end{bmatrix} \\
&\quad = \sum_{i=1}^{k}(\bar{q}'M_{GCi}R_2R_3V_{Gi}) \\
&\quad = tr\sum_{i=1}^{k}[\bar{q}'M_{GCi}(\mathbb{H}G\otimes B^{-1}G)(\bar{q}\otimes\bar{q})\bar{q}'G'\mathbb{H}V_{Gi}] \\
&\quad = tr\sum_{i=1}^{k}[\bar{q}'G'\mathbb{H}V_{Gi}\bar{q}'M_{GCi}(\mathbb{H}G\otimes B^{-1}G)(\bar{q}\otimes\bar{q})] \\
&\quad = tr\sum_{i=1}^{k}\{\{[(G'\mathbb{H}\otimes G'B^{-1})M'_{GCi}]\otimes V'_{Gi}\mathbb{H}G\}(\bar{q}\bar{q}'\otimes\bar{q}\bar{q}')\} \\
&\quad = tr\sum_{i=1}^{k}[(G'\mathbb{H}\otimes G'B^{-1})(M'_{GCi}\otimes V'_{Gi})(I_k\otimes \mathbb{H}G)(\bar{q}\bar{q}'\otimes\bar{q}\bar{q}')] \\
&\quad = tr[(G'\mathbb{H}\otimes G'B^{-1})M_V(I_k\otimes\mathbb{H}G)(\bar{q}\bar{q}'\otimes\bar{q}\bar{q}')].
\end{aligned}
\tag{A.19}
$$

From (A.13) and (A.19), (1.15) can be rewritten as

$$v_1(\bar{q}) = tr[L_1(\bar{q}\bar{q}'\otimes\bar{q}\bar{q}')], \tag{A.20}$$

where

$$L_1 = (G'\mathbb{H}\otimes G'B^{-1})V_D(\mathbb{H}G\otimes B^{-1}G) + (G'\mathbb{H}\otimes G'B^{-1})M_V(I_k\otimes\mathbb{H}G). \tag{A.21}$$

Similar to $u_1(\bar{q})$, $u_2(\bar{q})$ in (1.13) can be rewritten as

$$
\begin{aligned}
u_2(\bar{q}) &= \bar{q}'(G'B^{-1}G - I_k)(I_k\otimes\bar{q}'G'\mathbb{H}\otimes\bar{q}'G'\mathbb{H})vecD \\
&= (\bar{q}'\otimes\bar{q}'\otimes\bar{q}')[(G'B^{-1}G - I_k)\otimes G'\mathbb{H}\otimes G'\mathbb{H}]vecD \\
&= vec'J_2(\bar{q}\otimes\bar{q}\otimes\bar{q}),
\end{aligned}
\tag{A.22}
$$

where

$$vecJ_2 = [(G'B^{-1}G - I_k)\otimes G'\mathbb{H}\otimes G'\mathbb{H}]vecD. \tag{A.23}$$

The first term of $v_2(\bar{q})$ in (1.16) can be written as

$$\bar{q}'G'B^{-1}C(I_k \otimes \mathbb{H}G\bar{q} \otimes \mathbb{H}G\bar{q})(G'B^{-1}G - I_k)\bar{q}.$$

Since

$$\begin{aligned}(G'B^{-1}G - I_k)\bar{q} &= vec[\bar{q}'(G'B^{-1}G - I_k)] \\ &= (I_k \otimes \bar{q}')vec(G'B^{-1}G - I_k),\end{aligned}$$

and $vec(G'B^{-1}G - I_k)$ can be partitioned as

$$vec(G'B^{-1}G - I_k) = \begin{bmatrix} V_{GI1} \\ V_{GI2} \\ \vdots \\ V_{GIk} \end{bmatrix} \tag{A.24}$$

where V_{GIi} is $k \times 1$, we may mimic the second term of $v_1(\bar{q})$ and rewrite the first term of $v_2(\bar{q})$ further as

$$\begin{aligned} tr\sum_{i=1}^{k}[\bar{q}'M_{GCi}(\mathbb{H}G \otimes \mathbb{H}G)(\bar{q} \otimes \bar{q})\bar{q}'V_{GIi}] \\ = tr[(G'\mathbb{H} \otimes G'\mathbb{H})M_{VI}(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')],\end{aligned} \tag{A.25}$$

where

$$M_{VI} = M'_{GC1} \otimes V'_{GI1} + M'_{GC2} \otimes V'_{GI2} + \cdots + M'_{GCk} \otimes V'_{GIk}. \tag{A.26}$$

Similar to the first term of $v_1(\bar{q})$, since

$$\bar{q}'G'B^{-1}D = vec'(\bar{q}'G'B^{-1}D) = vec'D(I_{pk} \otimes B^{-1}G\bar{q}),$$

the second term of $v_2(\bar{q})$ in (1.16) can be rewritten as

$$\begin{aligned}&\frac{1}{2}vec'D(I_k \otimes B^{-1}G\bar{q} \otimes B^{-1}G\bar{q})(I_k \otimes \bar{q}'G'\mathbb{H} \otimes \bar{q}'G'\mathbb{H})vecD \\ &= \frac{1}{2}tr[V_D(B^{-1}G \otimes B^{-1}G)(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')(G'\mathbb{H} \otimes G'\mathbb{H})] \\ &= tr[\frac{1}{2}(G'\mathbb{H} \otimes G'\mathbb{H})V_D(B^{-1}G \otimes B^{-1}G)(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}'))].\end{aligned} \tag{A.27}$$

From (A.25) and (A.27), we have

$$v_2(\bar{q}) = tr[L_2(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')], \tag{A.28}$$

where

$$L_2 = (G'\mathbb{H} \otimes G'\mathbb{H})M_{VI} + \frac{1}{2}(G'\mathbb{H} \otimes G'\mathbb{H})V_D(B^{-1}G \otimes B^{-1}G). \tag{A.29}$$

Since

$$vec'G(I_k \otimes \mathbb{H}G\bar{q})$$
$$= [(I_k \otimes \bar{q}'G'\mathbb{H})vecG]'$$
$$= \bar{q}'G'\mathbb{H}G,$$

(1.14) becomes

$$\begin{aligned} u_3(\bar{q}) &= -\bar{q}'G'\mathbb{H}G(I_k \otimes \bar{q}'G'\mathbb{H} \otimes \bar{q}'G'\mathbb{H})vecD \\ &= -(\bar{q}' \otimes \bar{q}' \otimes \bar{q}')(G'\mathbb{H}G \otimes G'\mathbb{H} \otimes G'\mathbb{H})vecD \\ &= vec'J_3(\bar{q} \otimes \bar{q} \otimes \bar{q}), \end{aligned} \tag{A.30}$$

where

$$vecJ_3 = -(G'\mathbb{H}G \otimes G'\mathbb{H} \otimes G'\mathbb{H})vecD. \tag{A.31}$$

Similar to the second term of $v_1(\bar{q})$, the first term of $v_3(\bar{q})$ in (1.17) can be rewritten as

$$\begin{aligned} &-\bar{q}'G'B^{-1}C(I_k \otimes \mathbb{H}G\bar{q} \otimes \mathbb{H}G\bar{q})(I_k \otimes \bar{q}'G'\mathbb{H})vecG \\ &= tr\sum_{i=1}^{k}[-\bar{q}'M_{GCi}(\mathbb{H}G \otimes \mathbb{H}G)(\bar{q} \otimes \bar{q})\bar{q}'G'\mathbb{H}V_{Gi}] \\ &= tr[-(G'\mathbb{H} \otimes G'\mathbb{H})M_V(I_k \otimes \mathbb{H}G)(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')]. \end{aligned} \tag{A.32}$$

Similar to the second term of $v_2(\bar{q})$, the second term of $v_3(\bar{q})$ in (1.17) can be rewritten as

$$\begin{aligned} &-\bar{q}'G'B^{-1}D(I_k \otimes \mathbb{H}G\bar{q})(I_k \otimes \bar{q}'G'\mathbb{H} \otimes \bar{q}'G'\mathbb{H})vecD \\ &= -vec'D(I_{pk} \otimes B^{-1}G\bar{q})(I_k \otimes \mathbb{H}G\bar{q})(I_k \otimes \bar{q}'G'\mathbb{H} \otimes \bar{q}'G'\mathbb{H})vecD \\ &= -vec'D(I_k \otimes \mathbb{H}G\bar{q} \otimes B^{-1}G\bar{q})(I_k \otimes \bar{q}'G'\mathbb{H} \otimes \bar{q}'G'\mathbb{H})vecD \\ &= tr[-V_D(\mathbb{H}G \otimes B^{-1}G)(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')(G'\mathbb{H} \otimes G'\mathbb{H})] \\ &= tr[-(G'\mathbb{H} \otimes G'\mathbb{H})V_D(\mathbb{H}G \otimes B^{-1}G)(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')]. \end{aligned} \tag{A.33}$$

From (A.32) and (A.33), we have

$$v_3(\bar{q}) = tr[L_3(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')], \tag{A.34}$$

where

$$L_3 = -(G'\mathbb{H} \otimes G'\mathbb{H})M_V(I_k \otimes \mathbb{H}G) - (G'\mathbb{H} \otimes G'\mathbb{H})V_D(\mathbb{H}G \otimes B^{-1}G). \tag{A.35}$$

Similar to the first term of $v_1(\bar{q})$, $v_4(\bar{q})$ in (1.18) can be easily rewritten as

$$\begin{aligned} v_4(\bar{q}) &= \frac{1}{4}tr[V_D(\mathbb{H}G \otimes \mathbb{H}G)(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')(G'\mathbb{H} \otimes G'\mathbb{H})] \\ &= tr[\frac{1}{4}(G'\mathbb{H} \otimes G'\mathbb{H})V_D(\mathbb{H}G \otimes \mathbb{H}G)(\bar{q}\bar{q}' \otimes \bar{q}\bar{q}')] \\ &= tr[L_4(\bar{q}\bar{q}' \otimes \bar{q}\bar{q})], \end{aligned} \tag{A.36}$$

where

$$L_4 = \frac{1}{4}(G'\mathbb{H} \otimes G'\mathbb{H})V_D(\mathbb{H}G \otimes \mathbb{H}G). \tag{A.37}$$

By (A.8), (A.20), (A.22), (A.28), (A.30), (A.34) and (A.36), (A.6) and (A.7) are obtained, thus finishing the proof. □

Proof of Theorem 1.2: First, a_i and b_i are defined (Phillips and Park, 1988) as

$$a_i = tr(A_i) \quad (i = 0, 1, 2), \tag{A.38}$$

where

$$A_0 = L[(I + K_{k,k})(\bar{P} \otimes \bar{P}) + vec\bar{P}vec'\bar{P}],$$

$$A_1 = L[(I + K_{k,k})(\bar{P} \otimes P + P \otimes \bar{P}) + vec\bar{P}vec'P + vecPvec'\bar{P}],$$

$$A_2 = L[(I + K_{k,k})(P \otimes P) + vecPvec'P];$$

$$b_i = vec'JB_i vecJ \quad (i = 1, 2, 3), \tag{A.39}$$

where

$$\begin{aligned}
B_0 &= H(\bar{P} \otimes \bar{P} \otimes \bar{P}) + H(\bar{P} \otimes vec\bar{P}vec'\bar{P})H \\
&\quad + \bar{P} \otimes K_{k,k}(\bar{P} \otimes \bar{P}) + K_{k,k}(\bar{P} \otimes \bar{P}) \otimes \bar{P} \\
&\quad + K_{k,k^2}[\bar{P} \otimes K_{k,k}(\bar{P} \otimes \bar{P})]K_{k^2,k} = C_0(\bar{P}), \ say, \\
B_1 &= H(P \otimes \bar{P} \otimes \bar{P})H \\
&\quad + H(P \otimes vec\bar{P}vec'\bar{P} + \bar{P} \otimes vecPvec'\bar{P} + \bar{P} \otimes vec\bar{P}vec'P)H \\
&\quad + P \otimes K_{k,k}(\bar{P} \otimes \bar{P}) + \bar{P} \otimes K_{k,k}(P \otimes \bar{P}) \\
&\quad + \bar{P} \otimes K_{k,k}(\bar{P} \otimes P) + K_{k,k}(P \otimes \bar{P}) \otimes \bar{P} \\
&\quad + K_{k,k}(\bar{P} \otimes \bar{P}) \otimes \bar{P} + K_{k,k}(\bar{P} \otimes \bar{P}) \otimes P \\
&\quad + K_{k,k^2}\{[P \otimes K_{k,k}(\bar{P} \otimes \bar{P})] + [\bar{P} \otimes K_{k,k}(P \otimes \bar{P})] \\
&\qquad + [\bar{P} \otimes K_{k,k}(\bar{P} \otimes P)]\}K_{k^2,k} = C_1(\bar{P}, P), \ say, \\
B_2 &= C_1(P, \bar{P}), \\
B_3 &= C_0(P),
\end{aligned}$$

with

$$H = I + K_{k,k^2} + K_{k^2,k},$$

$$\bar{P} \equiv I - P.$$

Secondly, from (A.38),

$$a_0 = tr(A_0) = tr\{L[(I + K_{k,k})(\bar{P} \otimes \bar{P}) + vec\bar{P}vec'\bar{P}]\}$$
$$= tr[(\bar{P} \otimes \bar{P})L(I + K_{k,k}) + vec'\bar{P}Lvec\bar{P}] \quad \text{(A.40)}$$
$$= tr[(\bar{P} \otimes \bar{P})L(I + K_{k,k})] + tr(vec'\bar{P}Lvec\bar{P}).$$

Using (1.11) and $\bar{P} \equiv I - P$, we have

$$(A'B^{-1}G)\bar{P} = 0, \quad \text{(A.41)}$$

$$\bar{P}(G'B^{-1}A) = 0. \quad \text{(A.42)}$$

Therefore, by (1.23)-(1.27),

$$(\bar{P} \otimes \bar{P})L = 0, \quad \text{(A.43)}$$

and

$$(\mathbb{H}G \otimes B^{-1}G)vec\bar{P} = vec(B^{-1}G\bar{P}G\mathbb{H}) = 0, \quad \text{(A.44)}$$

$$(I_k \otimes \mathbb{H}G)vec\bar{P} = vec(\mathbb{H}G\bar{P}) = 0. \quad \text{(A.45)}$$

Combining (A.44) and (A.45) with (1.24) yields

$$L_1 vec\bar{P} = 0. \quad \text{(A.46)}$$

Similarly,

$$L_3 vec\bar{P} = 0, \quad \text{(A.47)}$$

$$L_4 vec\bar{P} = 0, \quad \text{(A.48)}$$

and

$$vec'\bar{P}L_2 = (L_2' vec\bar{P})' = 0. \quad \text{(A.49)}$$

From (A.46)-(A.49),

$$tr(vec'\bar{P}Lvec\bar{P}) = 0. \quad \text{(A.50)}$$

Substituting (A.43) and (A.50) into (A.40) gives

$$a_0 = 0. \quad \text{(A.51)}$$

Also, from (A.39),

$$
\begin{aligned}
b_1 &= vec'JB_1vecJ \\
&= vec'JH(P\otimes\bar{P}\otimes\bar{P})HvecJ \\
&\quad + vec'JH(P\otimes vec\bar{P}vec'\bar{P}+\bar{P}\otimes vecPvec'\bar{P}+\bar{P}\otimes vec\bar{P}vec'P)HvecJ \\
&\quad + vec'J[P\otimes K_{k,k}(\bar{P}\otimes\bar{P})+\bar{P}\otimes K_{k,k}(P\otimes\bar{P})]vecJ \\
&\quad + vec'J[\bar{P}\otimes K_{k,k}(\bar{P}\otimes P)+K_{k,k}(P\otimes\bar{P})\otimes\bar{P}]vecJ \\
&\quad + vec'J[K_{k,k}(\bar{P}\otimes\bar{P})\otimes\bar{P}+K_{k,k}(\bar{P}\otimes\bar{P})\otimes P]vecJ \\
&\quad + vec'JK_{k,k^2}\{[P\otimes K_{k,k}(\bar{P}\otimes\bar{P})]+[\bar{P}\otimes K_{k,k}(P\otimes\bar{P})] \\
&\qquad\qquad + [\bar{P}\otimes K_{k,k}(\bar{P}\otimes P)]\}K_{k^2,k}vecJ.
\end{aligned}
\tag{A.52}
$$

Using

$$K_{p,q}vecA = vec(A'),$$

$$A\otimes B = K_{p,r}(B\otimes A)K_{s,q},$$

for $A : p\times q$ and $B : r\times s$ where K is the commutation matrix, the following equations are obtained:

$$K_{k,k^2}vecJ_1 = 2(G'B^{-1}\otimes G'\mathbb{H}G\otimes G'\mathbb{H})vec(D'), \tag{A.53}$$

$$K_{k,k^2}vecJ_2 = [G'\mathbb{H}\otimes(G'B^{-1}G-I_k)\otimes G'\mathbb{H}]vec(D'), \tag{A.54}$$

$$K_{k,k^2}vecJ_3 = -(G'\mathbb{H}\otimes G'\mathbb{H}G\otimes G'\mathbb{H})vec(D'); \tag{A.55}$$

$$K_{k^2,k}vecJ_1 = 2(G'\mathbb{H}\otimes G'B^{-1}\otimes G'\mathbb{H}G)K_{p^2,k}vecD, \tag{A.56}$$

$$K_{k^2,k}vecJ_2 = [G'\mathbb{H}\otimes G'\mathbb{H}\otimes(G'B^{-1}G-I_k)]K_{p^2,k}vecD, \tag{A.57}$$

$$K_{k^2,k}vecJ_3 = -(G'\mathbb{H}\otimes G'\mathbb{H}\otimes G'\mathbb{H}G)K_{p^2,k}vecD. \tag{A.58}$$

Then, substituting (A.53)-(A.58) into (A.52), and using

$$vec(ABC) = (C'\otimes A)vecB,$$

$$(A\otimes B)' = A'\otimes B',$$

$$(A\otimes C)(B\otimes D) = (AB)\otimes(CD),$$

together with (A.41) and (A.42) yield

$$b_1 = 0. \tag{A.59}$$

Combining (A.51), (A.59) and the proof of Theorem 2.4 in Phillips and Park (1988) gives Theorem 1.2 in the current paper. □

Data Description of Chapter 1

The earnings data used are drawn from the Panel Study of Income Dynamics (PSID), available at *http : //psidonline.isr.umich.edu/*. The sample consists of men who were heads of household in every year from 1969 to 1974, who were between the ages of 21 (not inclusive) and 64 (not inclusive) in each year, and who reported positive earnings in each year. Individuals with average hourly earnings greater than $100 or reported annual hours greater than 4680 were excluded.

Some variables such as V7492, V7490, V0313, V0794, V7460, V7476, V7491 listed on p.443 of Abowd and Card (1989) are not available now on the PSID website. The variables for sex listed on that page are not consistent with those on the PSID website. The following are the PSID variables used in the present paper:

ANNUAL EARNINGS: V1196, V1897, V2498, V3051, V3463, V3863;

ANNUAL HOURS: V1138, V1839, V2439, V3027, V3423, V3823;

SEX: ER32000;

AGE: ER30046.

Data Description of Chapter 2

The 71 Quebec manufacturing industries, listed in Table A, are selected from the 4-digit manufacturing industries classified using the North American Industry Classification System (NAICS). The selection is based on the data availability of the variables used in the present chapter. The earnings data are average weekly earnings (Table 281-0027, Statistics Canada) adjusted by consumer price index (CPI) (Table 326-0021, Statistics Canada). The exports data are from Industry Canada, in 1,000,000 Canadian dollars and adjusted by CPI. The tariffs data are calculated according to both the tariff schedules from the UNCTAD-TRAINS database and the actual import duty raw data from the DLI database of Statistics Canada[1]. In calculation of the manufacturing tariff rates, the Concordance between the Customs Tariff of Canada (CT) and the 2002 NAICS is used, with the Concordance between NAICS Canada 2007 and NAICS Canada 2002 as a reference.

In studying the effect of free trade agreement, Trefler (2004, p. 888) considers tariffs of only non-zero imports, and Romalis (2007, p. 424) takes into account tariffs of both non-zero and zero imports. There might be two reasons that cause a product not to be imported from a foreign country: one is that this foreign country does not have comparative advantage, and the other is that the tariff is too high. By investigating both the tariff schedules and actual import duties for Quebec manufacturing industries around the starting years of NAFTA, it is found that the latter cause should not be ignored. The tariffs data used in this chapter are calculated using the following equation:

$$\tau_i = \frac{1}{2}\Big(\sum_{j \in J:\, m_j \neq 0} \tau_j m_j + \frac{1}{J_2} \sum_{j \in J:\, m_j = 0} \tau_j\Big),$$

where i denotes an industry, j denotes an HS10 (for actual import duties) or HS8 (for tariff schedules) item feeding into industry i, J denotes the set of HS10 or HS8 items feeding into industry i, m_j denotes the share of industry i's imports accounted for by item j, J_2 denotes the set of items with zero imports, and τ denotes tariff rate. We divide all products in industry i into two parts: those with non-zero imports and those with zero imports. For the non-zero imports part, import-weighted average of tariffs over items is calculated; for the zero imports part, simple average of tariffs is calculated. Then, simple average over the resulting tariffs in the two parts is calculated to obtain the final tariffs data. The raw data for products with non-zero imports are from the DLI database of Statistics Canada, and the raw data for products with zero imports are from the UNCTAD-TRAINS database. At the time of calculation, the 2004 import duty raw

[1] The results or views expressed are those of the authors.

data in the DLI database were partly missing, which were added later by Statistics Canada but some data are still not available, so we replace the 2004 tariffs data for the non-zero imports part using linear interpolation[2].

Table A: The 71 Quebec manufacturing industries

NAICS	Industry Description
3111	Animal food manufacturing
3112	Grain and oilseed milling
3114	Fruit and vegetable preserving and specialty food manufacturing
3115	Dairy product manufacturing
3116	Meat product manufacturing
3118	Bakeries and tortilla manufacturing
3119	Other food manufacturing
3131	Fibre, yarn and thread mills
3132	Fabric mills
3133	Textile and fabric finishing and fabric coating
3141	Textile furnishings mills
3149	Other textile product mills
3151	Clothing knitting mills
3152	Cut and sew clothing manufacturing
3159	Clothing accessories and other clothing manufacturing
3161	Leather and hide tanning and finishing
3162	Footwear manufacturing
3169	Other leather and allied product manufacturing
3211	Sawmills and wood preservation
3212	Veneer, plywood and engineered wood product manufacturing
3219	Other wood product manufacturing
3221	Pulp, paper and paperboard mills
3222	Converted paper product manufacturing
3231	Printing and related support activities
3241	Petroleum and coal product manufacturing
3251	Basic chemical manufacturing
3252	Resin, synthetic rubber, and artificial and synthetic fibres and filaments manufacturing
3254	Pharmaceutical and medicine manufacturing
3255	Paint, coating and adhesive manufacturing
3256	Soap, cleaning compound and toilet preparation manufacturing
3259	Other chemical product manufacturing
3261	Plastic product manufacturing
3262	Rubber product manufacturing

continued on the next page

[2]Fukao et al. (2003) use a similar way for the missing tariffs data.

Table A: The 71 Quebec manufacturing industries *(continued)*

NAICS	Industry Description
3271	Clay product and refractory manufacturing
3272	Glass and glass product manufacturing
3273	Cement and concrete product manufacturing
3279	Other non-metallic mineral product manufacturing
3311	Iron and steel mills and ferro-alloy manufacturing
3312	Steel product manufacturing from purchased steel
3313	Alumina and aluminum production and processing
3314	Non-ferrous metal (except aluminum) production and processing
3315	Foundries
3321	Forging and stamping
3322	Cutlery and hand tool manufacturing
3323	Architectural and structural metals manufacturing
3325	Hardware manufacturing
3326	Spring and wire product manufacturing
3327	Machine shops, turned product, and screw, nut and bolt manufacturing
3329	Other fabricated metal product manufacturing
3331	Agricultural, construction and mining machinery manufacturing
3332	Industrial machinery manufacturing
3333	Commercial and service industry machinery manufacturing
3334	Ventilation, heating, air-conditioning and commercial refrigeration equipment manufacturing
3335	Metalworking machinery manufacturing
3336	Engine, turbine and power transmission equipment manufacturing
3339	Other general-purpose machinery manufacturing
3342	Communications equipment manufacturing
3343	Audio and video equipment manufacturing
3344	Semiconductor and other electronic component manufacturing
3345	Navigational, measuring, medical and control instruments manufacturing
3346	Manufacturing and reproducing magnetic and optical media
3351	Electric lighting equipment manufacturing
3352	Household appliance manufacturing
3353	Electrical equipment manufacturing
3359	Other electrical equipment and component manufacturing
3364	Aerospace product and parts manufacturing
3369	Other transportation equipment manufacturing
3371	Household and institutional furniture and kitchen cabinet manufacturing
3372	Office furniture (including fixtures) manufacturing
3391	Medical equipment and supplies manufacturing
3399	Other miscellaneous manufacturing

Proof of Chapter 3

Proof of Proposition 3.1: We start with $N = 2$ for simplicity and later give the formulas for any N. Let $\hat{F}_{nt} = \hat{F}_n(x_{nt})$, $n = 1, 2$, $t = 1, \ldots, T$, be the empirical cdf's. Then,

$$\hat{d}_t(\theta) = vech[\nabla^2_\theta \ln c(\hat{F}_{1t}, \hat{F}_{2t}; \theta) + \nabla_\theta \ln c(\hat{F}_{1t}, \hat{F}_{2t}; \theta)\nabla'_\theta \ln c(\hat{F}_{1t}, \hat{F}_{2t}; \theta)].$$

Provided that the derivatives and expectation exist, let

$$\nabla D_\theta = E\nabla_\theta d_t(\theta)$$

and

$$\nabla \bar{D}_\theta = T^{-1}\sum_{t=1}^{T} \nabla_\theta \hat{d}_t(\theta).$$

First, expand $\sqrt{T}\bar{D}_{\hat{\theta}}$ with respect to θ:

$$\sqrt{T}\bar{D}_{\hat{\theta}} = \sqrt{T}\bar{D}_{\theta_o} + \nabla D_{\theta_o}\sqrt{T}(\hat{\theta} - \theta_o) + o_p(1).$$

Chen and Fan (2006b) show that

$$\sqrt{T}(\hat{\theta} - \theta_o) \to N(0, B^{-1}\Sigma B^{-1}),$$

where

$$B = -\mathbb{H}(\theta_0),$$

$$\Sigma = \lim_{T\to\infty} Var(\sqrt{T}A^*_T),$$

$$A^*_T = \frac{1}{T}\sum_{t=1}^{T}(\nabla_\theta \ln c(F_{1t}, F_{2t}; \theta_0) + W_1(F_{1t}) + W_2(F_{2t})).$$

Here terms $W_1(F_{1t})$ and $W_2(F_{2t})$ are the adjustments needed to account for the empirical distributions used in place of the true distributions. These terms are calculated as follows:

$$W_1(F_{1t}) = \int_0^1\int_0^1 [I\{F_{1t} \le u\} - u]\nabla^2_{\theta,u} \ln c(u, v; \theta_0)\, c(u, v; \theta_0)dudv,$$

$$W_2(F_{2t}) = \int_0^1\int_0^1 [I\{F_{2t} \le v\} - v]\nabla^2_{\theta,v} \ln c(u, v; \theta_0)\, c(u, v; \theta_0)dudv.$$

So,

$$\sqrt{T}(\hat{\theta} - \theta_o) = B^{-1}\sqrt{T}A^*_T + o_p(1).$$

Second, expand $\sqrt{T}\bar{D}_{\theta_0}$ with respect to F_{1t} and F_{1t}:

$$\sqrt{T}\bar{D}_{\theta_0} \simeq \frac{1}{\sqrt{T}}\sum_{t=1}^{T} d_t(\theta_0) + \frac{1}{T}\sum_{t=1}^{T} \nabla_{F_{1t}} d_t(\theta_0)\sqrt{T}(\hat{F}_{1t} - F_{1t}) + \frac{1}{T}\sum_{t=1}^{T} \nabla_{F_{2t}} d_t(\theta_0)\sqrt{T}(\hat{F}_{2t} - F_{2t}).$$

(A.60)

Under suitable regularity conditions,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\nabla_{F_{1t}}d_t(\theta_0)\sqrt{T}(\hat{F}_{1t}-F_{1t})\\
&\simeq \int_0^1\int_0^1 \nabla_u vech[\nabla_\theta^2 \ln c(u,v;\theta_0)+\nabla_\theta \ln c(u,v;\theta_0)\nabla'_\theta \ln c(u,v;\theta_0)]\\
&\qquad \sqrt{T}(\hat{F}_1(F_1^{-1}(u))-u)c(u,v;\theta_0)dudv\\
&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\int_0^1\int_0^1 [I\{F_{1t}\le u\}-u]\\
&\qquad \nabla_u vech[\nabla_\theta^2 \ln c(u,v;\theta_0)+\nabla_\theta \ln c(u,v;\theta_0)\nabla'_\theta \ln c(u,v;\theta_0)]c(u,v;\theta_0)dudv.
\end{aligned}$$

Denote

$$\begin{aligned}
M_1(F_{1t}) = &\int_0^1\int_0^1 [I\{F_{1t}\le u\}-u]\\
&\nabla_u vech[\nabla_\theta^2 \ln c(u,v;\theta_0)+\nabla_\theta \ln c(u,v;\theta_0)\nabla'_\theta \ln c(u,v;\theta_0)]c(u,v;\theta_0)dudv,
\end{aligned}$$

then

$$\frac{1}{T}\sum_{t=1}^{T}\nabla_{F_{1t}}d_t(\theta_0)\sqrt{T}(\hat{F}_{1t}-F_{1t}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}M_1(F_{1t}).$$

Similarly, denote

$$\begin{aligned}
M_2(F_{2t}) = &\int_0^1\int_0^1 [I\{F_{2t}\le v\}-v]\\
&\nabla_v vech[\nabla_\theta^2 \ln c(u,v;\theta_0)+\nabla_\theta \ln c(u,v;\theta_0)\nabla'_\theta \ln c(u,v;\theta_0)]c(u,v;\theta_0)dudv,
\end{aligned}$$

then

$$\frac{1}{T}\sum_{t=1}^{T}\nabla_{F_{2t}}d_t(\theta_0)\sqrt{T}(\hat{F}_{2t}-F_{2t}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}M_2(F_{2t}).$$

Therefore, equation (A.60) can be rewritten as

$$\sqrt{T}\bar{D}_{\theta_0} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}d(\theta_0)+\sqrt{T}B_T^*+o_p(1),$$

where

$$B_T^* = \frac{1}{T}\sum_{t=1}^{T}[M_1(F_{1t})+M_2(F_{2t})].$$

Finally, combining the expansions gives

$$\sqrt{T}\bar{D}_{\hat{\theta}} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}d(\theta_0)+\sqrt{T}B_T^*+\nabla D_{0_0}B^{-1}\sqrt{T}A_T^*+o_p(1).$$

So

$$\sqrt{T}\bar{D}_{\hat{\theta}} \to N(0, V_{\theta_0}),$$

or, equivalently,

$$T\bar{D}'_{\hat{\theta}} V_{\theta_0}^{-1} \bar{D}_{\hat{\theta}} \to \chi^2_{p(p+1)/2},$$

where

$$\begin{aligned} V_{\theta_0} = E\{&d_t(\theta_0) + M_1(F_{1t}) + M_2(F_{2t}) \\ &+\nabla D_{\theta_0} B^{-1}[\nabla_\theta \ln c(F_{1t}, F_{2t}; \theta_0) + W_1(F_{1t}) + W_2(F_{2t})]\} \\ &\times \{d_t(\theta_0) + M_1(F_{1t}) + M_2(F_{2t}) \\ &+\nabla D_{\theta_0} B^{-1}[\nabla_\theta \ln c(F_{1t}, F_{2t}; \theta_0) + W_1(F_{1t}) + W_2(F_{2t}))]\}'. \end{aligned}$$

Extension to $N \geq 2$ is straightforward. Now

$$d_t(\theta) = vech[\nabla^2_\theta \ln c(F_{1t}, F_{2t}, \cdots, F_{Nt}; \theta) + \nabla_\theta \ln c(F_{1t}, F_{2t}, \cdots, F_{Nt}; \theta)\nabla'_\theta \ln c(F_{1t}, F_{2t}, \cdots, F_{Nt}; \theta)],$$

and the asymptotic variance matrix becomes

$$\begin{aligned} V_{\theta_0} = E&\left\{d_t(\theta_0) + \nabla D_{\theta_0} B^{-1}\left[\nabla_\theta \ln c(F_{1t}, F_{2t}, \cdots, F_{Nt}; \theta_0) + \sum_{n=1}^{N} W_n(F_{nt})\right] + \sum_{n=1}^{N} M_n(F_{nt})\right\} \\ &\times \left\{d_t(\theta_0) + \nabla D_{\theta_0} B^{-1}\left[\nabla_\theta \ln c(F_{1t}, F_{2t}, \cdots, F_{Nt}; \theta_0) + \sum_{n=1}^{N} W_n(F_{nt})\right] + \sum_{n=1}^{N} M_n(F_{nt})\right\}', \end{aligned} \tag{A.61}$$

where, for $n = 1, 2, \cdots, N$,

$$\begin{aligned} W_n(F_{nt}) = \int_0^1 \int_0^1 \cdots \int_0^1 &[I\{F_{nt} \leq u_n\} - u_n]\nabla^2_{\theta, u_n} \ln c(u_1, u_2, \cdots, u_N; \theta_0) \\ &c(u_1, u_2, \cdots, u_N; \theta_0) du_1 du_2 \cdots du_N, \end{aligned}$$

and

$$\begin{aligned} M_n(F_{nt}) = \int_0^1 \int_0^1 \cdots \int_0^1 &[I\{F_{nt} \leq u_n\} - u_n]\nabla_{u_n} vech[\nabla^2_\theta \ln c(u_1, u_2, \cdots, u_N; \theta_0) \\ &+ \nabla_\theta \ln c(u_1, u_2, \cdots, u_N; \theta_0)\nabla'_\theta \ln c(u_1, u_2, \cdots, u_N; \theta_0)] \\ &c(u_1, u_2, \cdots, u_N; \theta_0) du_1 du_2 \cdots du_N. \end{aligned}$$

□