ABSTRACT

KURMANJ, AGIR. Noise Ratio As a Non-Nested Model Selection Tool. (Under the direction of Dr. Atsushi Inoue and Dr. Denis Pelletier.)

In this dissertation, we analyze the noise ratio statistic of Durlauf and Hall (1989), $NR_T(\hat{\theta}_T)$, as a non-nested model selection tool using the Rivers and Vuong (2002) framework. For this purpose, we first show that, when scaled by the sample size T, $NR_T(\hat{\theta}_T)$ is distributed as a mixture of chi-square random variables. Further, we study the asymptotic distribution of functionals of this statistic for model selection purposes under different assumptions about: i) model specification and ii) the data generating processes of two non-nested Rational Expectations (RE) models whose parameter vector is estimated either by GMM (in Chapter 1) or by the continuous updating estimator (in Chapter 2).

In Chapter 3, we use Monte Carlo simulations to compute the empirical size and empirical power of tests with statistics whose limiting distributions were studied in Chapters 1 and 2. First, we use a simulation routine and compute the rejection frequency of the tests developed using these statistics, which represents empirical size under a null hypothesis and power under an alternative. Under our null hypothesis, both models are equally good from a goodness of fit perspective, that is, $H_0 : NR_*^{(1)} = NR_*^{(2)}$, where $NR_*^{(i)}$ denotes the probability limit of the noise ratio statistic evaluated at the probability limit of the GMM estimator, which we denote by $\theta_*^{(i)}$, for models $i = 1, 2$. Under the first alternative, the first model is better, that is $H_1 : NR_*^{(1)}(\theta_*^{(1)}) < NR_*^{(2)}(\theta_*^{(2)})$, and under the second alternative hypothesis, the second model is better from a goodness of fit perspective, that is, we have $H_2 : NR_*^{(1)}(\theta_*^{(1)}) > NR_*^{(2)}(\theta_*^{(2)})$. Under all scenarios covered in this chapter, we use the limit of the noise ratio statistic evaluated at the probability limit of the GMM estimator as our goodness of fit measure.

Finally, in Chapter 4, we use the model selection methodology used in Chapters 1 and 2 for comparing different formulations of the pure production smoothing model of inventories. The particular models compared are the production smoothing model of inventories and its variants, covered in Durlauf and Maccini (1995). The statistics used for comparing these models include $T\{NR_T^{(2)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}$ and B_T of Rivers and Vuong (2002). All statistics are evaluated at the GMM estimator $\hat{\theta}^{(i)}$ for the corresponding model $i = 1, 2$.

Noise Ratio As a Non-Nested Model Selection Tool

by
Agir Kurmanj

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fullfillment of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina

2009

APPROVED BY:

Dr. Douglas K. Pearce

Dr. Negash G. Medhin

Dr. Atsushi Inoue
Chair of Advisory Committee

Dr. Denis Pelletier
Co-Chair of Advisory Committee

DEDICATION

To my beloved Kurdistan

BIOGRAPHY

Agir Kurmanj was born in Ankara, Turkey, to Mehmed and Ele Kilinc in 1973. One of eight children, Agir attended high school at TED Ankara College from 1984 until 1991. He then went on to attend METU (Middle East Technical University), where he graduated in 1998 with highest possible honors. Agir came to the United States in 1998, obtaining his M.Sc. in Economics from North Carolina State University (NCSU) in 2000. He is currently pursuing his Ph.D. in Economics at NCSU.

ACKNOWLEDGMENTS

I would first like to acknowledge and sincerely thank Dr. Atsushi Inoue for his invaluable support and for sharing his exceptional knowledge in the field of Econometrics, both of which have contributed greatly to my academic growth and enrichment. Next, I would like to express my deepest gratitude to Dr. Denis Pelletier for his insight and guidance throughout this process. It has been an absolute pleasure to have worked with him. Finally, I would like to thank Dr. Negash G. Medhin and Dr. Douglas K. Pearce for their interest in my research and for their participation as committee members.

TABLE OF CONTENTS

Chapter 1

Noise Ratio and GMM

1.1 Introduction

1.1.1 Motivation

This research is motivated by the need to find tests that can be used for non-nested model selection in situations where models under consideration are possibly misspecified. A possible candidate statistic for developing such tests is the Noise Ratio Statistic (NRS) of Durlauf and Hall (1989). Durlauf and Maccini (1995) use the NRS for model selection purposes in the case of inventory models. However, their use of NRS is not based on any formal statistical theory. This dissertation aims to develop a formal theoretical foundation for using the NRS for model selection tests where models under consideration are possibly misspecified. Even when two models are misspecified, one of the models might be better from a goodness of fit perspective. The goodness of fit criteria we use here is the probability limit of the noise ratio for that model. In Chapter 1, we define our theory formally for the case where the NRS is evaluated at the GMM estimator. The special case of models covered here are known as RE models whose parameters are estimated using the GMM estimator.

1.1.2 GMM estimation of RE models

Discrete-time models of optimizing behavior of economic agents often lead to first order conditions (f.o.c.), which yield moment conditions of the following form:

$$E[g_t(\theta_*)|\Omega_t] = 0 \tag{1.1}$$

where $g_t(\theta)$ is a Rational Expectations (RE) error term at time $t+1$, $\theta_* \in \Theta$ is a $p \times 1$ parameter vector and Ω_t denotes the economic agents information set, as of time t. In this setting, the expectations are assumed to be formed rationally by basing them on all available information, hence the name RE Models.

In order to make correct inferences, we need to know the value of θ_* or be able to estimate it. In the real world, most of the time, the value of θ_* is either unknown or impractical to compute. For that reason, we have to estimate it. In Generalized Method of Moments (GMM) estimation of RE models, we begin by deriving an unconditional moment condition from Equation (1.1), using the law of iterated expectations (LIE):

$$\begin{aligned} E[z_t g_t(\theta_*)] &= E[E[z_t g_t(\theta_*)|\Omega_t]] && (1.2)\\ &= E[z_t E[g_t(\theta_*)|\Omega_t]] && (1.3)\\ &= 0 && (1.4)\end{aligned}$$

where z_t denotes the $k \times 1$ instrument vector with $z_t \in \Omega_t$.

Then, using a sample analog of this moment condition, we can compute the following first-step GMM estimator for θ_*:

$$\hat{\theta}_{T,1} = argmin_{\theta \in \Theta} Q_{T,1}(\theta) \tag{1.5}$$

$$Q_{T,1}(\theta) = \left\{T^{-1}\sum_{t=1}^{T} z_t g_t(\theta, v_t)\right\}' \left[T^{-1}\sum_{t=1}^{T} z_t z_t'\right] \left\{T^{-1}\sum_{t=1}^{T} z_t g_t(\theta, v_t)\right\} \tag{1.6}$$

where $v_t \in \Re^z$ is a set of observations on a random vector of interest.

Then, using (1.5), we can derive the following second-step GMM estimator for θ_*:

$$\hat{\theta}_T = argmin_{\theta \in \Theta} Q_T(\theta) \tag{1.7}$$

where

$$Q_T(\theta) = \left\{T^{-1}\sum_{t=1}^{T} z_t g_t(\theta, v_t)\right\}' \hat{S}_T(\hat{\theta}_{T,1})^{-1} \left\{T^{-1}\sum_{t=1}^{T} z_t g_t(\theta, v_t)\right\} \tag{1.8}$$

where

(i) $\hat{\theta}_{T,1}$ is a first-step estimator used to estimate the second-step weighting matrix,

(ii) $\hat{S}_T(\hat{\theta}_{T,1})$ is an estimate of the $lim_{T\to\infty}var(T^{-1/2}\sum_{t=1}^{T} z_t g_t(\theta_*, v_t))$, where $plim(\hat{\theta}_T) = \theta_*$,

(iii) $\hat{\theta}_T$ is such that $plim(\hat{\theta}_T^{(i)}) = \theta_*$[1].

In the next section, we introduce the Noise Ratio (NR_T) statistic of Durlauf and Hall (1989), which is a function of the GMM estimator for each model. Further, in the following sections, we study its potential for model selection purposes.

1.1.3 The NR_T statistic

The methodology behind the NR_T statistic is developed in Durlauf and Hall (1989). Durlauf and Maccini (1995) first used the NR_T statistic for comparing the goodness of fit of RE inventory models estimated by GMM. To derive the noise ratio, Durlauf and Maccini (1995) use the regression of $g_t^{(i)}(\hat{\theta}_T^{(i)})$ onto $z_t^{'}$, where $g_t^{(i)}(\hat{\theta}_T^{(i)})$ denotes an estimated Euler equation forecast error[2] derived from the underlying economic model and $z_t^{'}$ denotes the t^{th} row of the $T \times k$ instrumental variables (IV) matrix Z. Further, "$'$" denotes matrix or vector transpose.

This regression is as follows:

$$\begin{aligned} g_t^{(i)}(\hat{\theta}_T^{(i)}) &= \hat{g}_t(\hat{\theta}_T^{(i)}) + \hat{u}_t(\hat{\theta}_T^{(i)}) \\ &= z_t^{'}\left(\sum_{t=1}^{T} z_t z_t^{'}\right)^{-1} \sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)}) z_t + \hat{u}_t(\hat{\theta}_T^{(i)}) \\ &= z_t^{'}\hat{\beta}_T + \hat{u}_t(\hat{\theta}_T^{(i)}) \end{aligned} \tag{1.9}$$

where $g_t^{(i)}(\hat{\theta}_T^{(i)})$ is the estimated Euler equation forecast error, $\hat{\beta}_T$ denotes the ordinary least squares (OLS) parameter estimator in (1.9), $\hat{g}_t(\hat{\theta}_T^{(i)})$ denotes the projection of $g_t^{(i)}(\hat{\theta}_T^{(i)})$ on to the column space of Z_t and $\hat{u}_t(\hat{\theta}_T^{(i)})$ is the residual term of the same regression.

The OLS estimator $\hat{\beta}_T$ is given by:

$$\hat{\beta}_T = (Z^{'}Z)^{-1}Z^{'}g(\hat{\theta}_T^{(i)}) \tag{1.10}$$

---

[1] $\hat{\theta}_T^{(i)}$ is some consistent estimator of $\theta_*^{(i)}$.

[2] The dependence of the Euler equation forecast error on $\hat{\theta}_T^{(i)}$ will play an important part in deriving the asymptotic properties of the NR_T and we will write $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ to denote our noise ratio statistic.

Now, given (1.9), NR_T is defined as follows:

$$NR_T^{(i)}(\hat{\theta}_T^{(i)}) \equiv \frac{T^{-1}\sum\limits_{t=1}^{T}\hat{g}_t(\hat{\theta}_T^{(i)})^2}{T^{-1}\sum\limits_{t=1}^{T}g_t^{(i)}(\hat{\theta}_T^{(i)})^2} \tag{1.11}$$

We can denote (1.9) more conveniently in matrix notation as follows:

$$g(\hat{\theta}_T^{(i)}) = \hat{g}(\hat{\theta}_T^{(i)}) + \hat{u}_t(\hat{\theta}_T^{(i)}) \tag{1.12}$$

$$= P_z g(\hat{\theta}_T^{(i)}) + \hat{u}(\hat{\theta}_T^{(i)}) \tag{1.13}$$

where $g(\hat{\theta}_T^{(i)})$, $\hat{g}(\hat{\theta}_T^{(i)})$ and $\hat{u}_t(\hat{\theta}_T^{(i)})$ are all $T \times 1$ vectors formed by stacking g_t, $\hat{g}_t(\hat{\theta}_T^{(i)})$ and $\hat{u}_t(\hat{\theta}_T^{(i)})$ in vectors. Also, $P_z = Z(Z'Z)^{-1}Z'$ is the $T \times T$ projection matrix which linearly projects a variable onto the column space of Z. Using this setup, we denote our noise ratio in matrix notation as follows:

$$NR_T(\hat{\theta}_T^{(i)}) = \frac{T^{-1}g(\hat{\theta}_T^{(i)})' P_z g(\hat{\theta}_T^{(i)})}{T^{-1}g(\hat{\theta}_T^{(i)})' g(\hat{\theta}_T^{(i)})} \tag{1.14}$$

$$= \frac{T^{-1}g(\hat{\theta}_T^{(i)})' Z(Z'Z)^{-1}Z' g(\hat{\theta}_T^{(i)})}{T^{-1}g(\hat{\theta}_T^{(i)})' g(\hat{\theta}_T^{(i)})} \tag{1.15}$$

$$= m_T^{(i)}(\hat{\theta}_T^{(i)})' W_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} m_T^{(i)}(\hat{\theta}_T^{(i)}) \tag{1.16}$$

where

$$m_T^{(i)}(\hat{\theta}_T^{(i)}) = T^{-1}\sum_{t=1}^{T} g_t(\hat{\theta}_T, v_t) z_t \tag{1.17}$$

$$W_T^{(i)}(\hat{\theta}_T^{(i)}) = \left(T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)'}\right)\left(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})^2\right) \tag{1.18}$$

We have indexed NR_T by T because of the dependence of NR_T on $\hat{\theta}_T$[3]. In the next section, we discuss model selection in the context of RE models estimated by GMM.

---

[3] The GMM estimator of the parameter vector, $\hat{\theta}_T$, depends on the sample size T.

1.1.4 Model selection using $NR_T^{(i)}(\hat{\theta}_T^{(i)})$

In the context of RE models estimated by GMM, model validation has traditionally been based on the well known overidentifying restrictions (OIR) test statistic which we denote by J_T.

The OIR statistic is used to test if the moment conditions are satisfied by the data. In the case of misspecified models, the OIR statistic could have a degenerate distribution. Theorem 5.2 of Hall (2005) gives the distribution of the OIR test statistic when the model under consideration is misspecified[4]. In the special case where the weighting matrix is p.s.d. (positive semi-definite) and converges to a p.d. (positive definite) matrix of constants, it is shown that J_T diverges to ∞ at rate T.

On the other hand, Hall and Pelletier (2008) studied the distribution of Rivers and Vuong (2002) statistic for models estimated by GMM, where the goodness of fit of the models is measured by their GMM minimands. Formally, the statistic Hall and Pelletier (2008) studied is given as follows:

$$N_T = \frac{T^{1/2}\{Q_T^{(1)}(\hat{\theta}_T^{(1)}) - Q_T^{(2)}(\hat{\theta}_T^{(2)})\}}{\hat{\sigma}_T} \tag{1.19}$$

where $\hat{\sigma}_T$ is a consistent estimator of the limiting standard error of the numerator of (1.19) and $Q_T^{(i)}(\hat{\theta}_T^{(i)})$ is the corresponding GMM minimand for model i, where $i = 1, 2$.

They find that the asymptotic distribution of N_T depends crucially on assumptions about model specification and the weighting matrix choice. We can summarize their findings as follows:

(i) If both models are correctly specified, then the limiting distribution of N_T, under the null that $Q_*^{(1)}(\theta_*^{(1)}) = Q_*^{(1)}(\theta_*^{(1)})$, is non-standard.

(ii) If both models are misspecified, then under the null that $Q_*^{(1)}(\theta_*^{(1)}) = Q_*^{(1)}(\theta_*^{(1)})$, N_T has a standard normal limiting distribution. However, it is also shown that this distribution depends crucially on the weighting matrix used. This finding makes it difficult to perform inference in cases where there is no a priori weighting matrix choice from the underlying economic theory.

---

[4] Where misspecification is defined as: $\nexists\, \theta_0 \in \Theta$: $E[g(\theta_0, v_t)] = 0$.

These findings motivate econometric research about other statistics that can be used when the models considered are misspecified. One such possible candidate is the NR_T statistic used by Durlauf and Maccini (1995). The NR_T is used to determine which of the two inventory RE models estimated by GMM is closer to the truth. Specifically, the noise ratio measures how well a model explains movements in data. The smaller a noise ratio, the better a model explains movements in the data. Further, the authors compare their results based on the NR_T to those obtained using the OIR. As outlined above, the usage of the GMM minimands[5] in misspecified models with non-local misspecification[6] depends on the choice of the weighting matrix. Given that NR_T is a function of the GMM estimator, we would like to know how this dependence affects the asymptotic properties of the NR_T and whether or not we can use it as an alternative statistic for model selection purposes in the case of misspecified models. Also, we would like to know if the NR_T statistic can be used as a non-nested model selection tool within the Rivers and Vuong (2002) framework.

1.1.5 Non-nested model selection methodology

Rivers and Vuong (2002) propose tests of model selection for two non-nested dynamic models, based on the difference between the goodness of fit measures for two models. In particular, one model is preferred if its goodness of fit measure is significantly smaller than the other. In the case of GMM models, they propose the GMM minimand as a measure of goodness of fit of a model where model selection is based on the statistic given in (1.19).

In this paper, we would like to evaluate whether the noise ratios for two models, $NR_T^{(1)}$ and $NR_T^{(2)}$, can be used as a basis for non-nested model selection using a similar framework to that of Rivers and Vuong (2002). Specifically, we would like to compute the limiting distribution of the Rivers and Vuong (2002) test statistic given by (1.19), where we would replace the GMM minimand with the noise ratio as the goodness of fit measure.

For this purpose, we first define the following statistic:

$$B_T = \frac{T^{1/2}\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}}{\hat{\rho}_T} \tag{1.20}$$

---

[5] The GMM minimand equals $T^{-1}J_T$.

[6] See Definition 2 of Hall and Inoue (2003).

where $\hat{\rho}_T$ is a consistent estimator of the limiting standard error of the numerator of (1.20) and $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ is the corresponding noise ratio statistic for model i, evaluated at the GMM estimator for model i for $i = 1, 2$.

In the following, we will derive the limiting distributions of the $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ and B_T statistics under different assumptions about i) the data generation process and ii) the model specification. In this setting, a model's goodness of fit criterion is that a model with a lower NR_T is preferred to one with a higher noise ratio.

Also, we would like to know how assumptions about the data generation process and model specification affect the asymptotic distribution of $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ in order to compare our results to those about the GMM minimand obtained by Hall and Pelletier (2008).

In this setting, we assume that there are two non-nested models, which we denote by M_1 and M_2, that we would like to compare. Further, we assume that each model implies a population moment condition as follows:

$$M_i \Longrightarrow E[z_t^{(i)} g_t^{(i)}(v_t, \theta_*^{(i)})] = 0 \quad \text{for a unique} \;\; \theta_*^{(i)} \in \Theta^{(i)} \tag{1.21}$$

for $i = 1, 2$ and $t = 1, 2, .., T$.

Under the null hypothesis, we assume that the two models, M_1 and M_2, are asymptotically equivalent from a goodness of fit perspective, which we can formally state as follows:

$$NR_*^{(1)}(\theta_*^{(1)}) = NR_*^{(2)}(\theta_*^{(2)}) \tag{1.22}$$

where $NR_*^{(i)}(\theta_*^{(i)})$ denotes the probability limit of the noise ratio for i=1,2.

There are two alternative hypotheses of interest. Under the first alternative hypothesis, M_1 is better than M_2, that is,

$$NR_*^{(1)}(\theta_*^{(1)}) < NR_*^{(2)}(\theta_*^{(2)}) \tag{1.23}$$

Under the second alternative hypothesis, M_2 is better than M_1, that is,

$$NR_*^{(1)}(\theta_*^{(1)}) > NR_*^{(2)}(\theta_*^{(2)}) \tag{1.24}$$

In this dissertation, we would like to analyze whether we can use $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ as a non-nested model selection tool using the Rivers and Vuong (2002) framework. For this purpose, we would like to study the distribution of NR_T and B_T under different assumptions about

i) model specification and ii) the data generating processes of two non-nested RE models whose parameter vector is estimated either by GMM, in Chapter 1, or CUE, in Chapter 2.

1.2 Notation and preliminaries

In order to derive all of the limiting distributions outlined in this dissertation, we impose the following primitive assumptions:

Assumption 1 *$\{v_t, z_t, t = 1, 2...\} \in \mathscr{V}$ is a sequence of strictly stationary and ergodic random vectors, where $\mathscr{V} \subseteq \Re^z$.*

Assumption 1 implies that all expectations of functions of v_t and z_t are independent of time.

Assumption 2 *For each $i = 1, 2$, the function $g_t^{(i)}(v_t, \theta) : \mathscr{V} \times \Theta^{(i)} \to \Re^{q_i}$, where $\Theta^{(i)} \subset \Re^{p_i}$ and $q_i < \infty$, is:*

(i) Measurable for each $\theta^{(i)} \in \Theta^{(i)}$ and $g_t^{(i)}(v, .)$ and is continuous on Θ for every $v \in \mathscr{V}$,

(ii) $E[g_t^{(i)}(v_t, \theta^{(i)}) z_t^{(i)}]$ exists and is finite for every $\theta^{(i)} \in \Theta^{(i)}$,

(iii) $E[g_t^{(i)}(v_t, \theta^{(i)}) z_t^{(i)}]$ is continuous on $\Theta^{(i)}$.

Through Assumptions 1 and 2-(i), any measurable function of the data will also be strictly stationary and ergodic.

In Chapter 1, it is assumed that parameters of both models are estimated via GMM, where the second-step GMM estimator $\hat{\theta}_T^{(i)}$ is given by (1.7). On the other hand, in Chapter 2, we derive our results when the parameters of both models are estimated via the continuous updating estimator (CUE).

Let $\hat{\theta}_T^{(i)}$ be such that $plim(\hat{\theta}_T^{(i)}) = \theta_*^{(i)}$.

Assumption 3 *$S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}$ is a positive semi-definite, symmetric matrix which converges in probability to a positive definite matrix of constants, ${S^{(i)}}^{-1}$.*

Assumption 4 *$\Theta^{(i)}$ is a compact set for $i = 1, 2$.*

Assumption 5 *$E[sup_{\theta^{(i)} \in \Theta^{(i)}} \|g_t^{(i)}(v_t, \theta^{(i)}) z_t^{(i)}\|] < \infty$ for $i = 1, 2$* [7].

Next, we impose the following assumption which will ensure unique identification of the parameter vector:

Assumption 6 *There exists $\theta_*^{(i)} \in \Theta^{(i)}$, such that $Q_*^{(i)}(\theta_*^{(i)}) < Q_*^{(i)}(\theta^{(i)})$ for all $\theta^{(i)} \in \Theta^{(i)} \backslash \{\theta_*^{(i)}\}$,*

where

$$Q_*^{(i)}(\theta^{(i)}) = E[g_t^{(i)}(v_t, \theta^{(i)}) z_t^{(i)}]'(S^{(i)})^{-1} E[g_t^{(i)}(v_t, \theta^{(i)}) z_t] \tag{1.25}$$

Before we state our next lemma, we make the following assumption:[8].

Lemma 1 *Consistency of $\hat{\theta}_T^{(i)}$*
If assumptions 1-6 hold, then $\hat{\theta}_T^{(i)} \xrightarrow{p} \theta_*^{(i)}$ for $i = 1, 2$.

Lemma 1 follows from Theorem 7.1 of Wooldridge (1994)[9]. This lemma states that the minimizer of the sample minimands converges to the unique minimizer of the population minimand. In addition to these assumptions, the following restrictions will be placed on $\theta_*^{(i)}$ and the following derivative matrices:

$$G_T^{(i)}(\theta^{(i)}) = T^{-1} \sum_{i=1}^{T} z_t^{(i)} \partial g^{(i)}(v_t, \theta^{(i)}) / \partial \theta^{(i)'} \tag{1.26}$$

$$G_*^{(i)}(\theta^{(i)}) = E[z_t^{(i)} \partial g^{(i)}(v_t, \theta^{(i)}) / \partial \theta^{(i)'}] \tag{1.27}$$

Assumption 7 *$\theta_*^{(i)}$ is an interior point of $\Theta^{(i)}$, for $i = 1, 2$.*

Assumption 8 *For $i = 1, 2$, we have the following:*

(i) $\partial g^{(i)}(v_t, \theta^{(i)}) / \partial \theta^{(i)'}$ exists and is continuous on $\Theta^{(i)}$ for each $v_t \in \mathscr{V}$,

(ii) $G_*^{(i)}(\theta_*^{(i)})$ exists, is finite and continuous on some neighborhood of $N_\epsilon^{(i)}$ of $\theta_*^{(i)}$,

(iii) $sup_{\theta^{(i)} \in N_\epsilon^{(i)}} \|G_T^{(i)}(\theta^{(i)}) - G_*^{(i)}(\theta^{(i)})\| \xrightarrow{p} 0$,

---

[7] For any $m \times n$ matrix A, the Euclidean norm is defined as $\|A\| = [tr(A'A)]^{1/2}$.

[8] This assumption will ensure that the probability limit of $\hat{\theta}_T^{(i)}$ is well defined.

[9] See Wooldridge (1994, p. 2694).

(iv) $Rank(G_*^{(i)}(\theta_*^{(i)})) = p_i$.

Assumption 9 *$g^{(i)}(v,\theta)$ is twice continuously differentiable with respect to $\theta^{(i)}$ on $int(\Theta^{(i)})$, and $\partial g_t^{(i)}(.,\theta)/\partial\theta^{(i)\prime}$ is measurable on $\mathscr{V}$ for each $\theta^{(i)} \in int(\Theta^{(i)})$ and $i = 1, 2$, where $int(\Theta^{(i)})$ denotes the interior of the set Θ.*

Given the form of $W_T^{(i)}(\hat\theta_T^{(i)})$ in (1.18), a Cholesky decomposition will exist if $W_T^{(i)}(\hat\theta_T^{(i)})$ is p.s.d. However, we also want to avoid the case where the probability limit of $W_T^{(i)}(\hat\theta_T^{(i)})$ is equal to zero.

Assumption 10 *We also assume the following:*

(i) $E[g_t^{(i)}(\hat\theta_T^{(i)})^2] > 0$,

(ii) $T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}$ *is p.s.d. and converges to M_{zz}, a p.d. matrix of constants.*

Next, we formally define a martingale difference sequence (m.d.s.) as follows[10]:

Definition 1 *Let $\{\mathcal{Y}_t, \mathcal{F}_t\}$ be an adapted stochastic sequence. $\{\mathcal{Y}_t, \mathcal{F}_t\}$ is an m.d.s. if:*

$$E[\mathcal{Y}_t|\mathcal{F}_{t-1}] = 0 \qquad \forall t \geq 1 \tag{1.28}$$

The following example from White (2001)[11] illustrates the concept of m.d.s. Let $\{\mathcal{Y}_t\}$ be an i.i.d. sequence of random variables such that $E[\mathcal{Y}_t] = 0$. If we let $\mathcal{F}_{t-1} = \sigma(..., \mathcal{Y}_{t-1}, \mathcal{Y}_t)$, the sigma-algebra generated by the history of the random variable $\mathcal{Y}_t$, then $\{\mathcal{Y}_t, \mathcal{F}_t\}$ is an m.d.s.

In our case, we take $g_t^{(i)}(\theta_*^{(i)}, v_t)$ as $\mathcal{Y}_t$ and Ω_t as $\mathcal{F}_{t-1}$. For purposes of estimation, we will use $z_t^{(i)} \in \Omega_t$ as instrumental variables for estimating the parameter vector $\theta^{(i)}$ by GMM.

Assumption 11 *$var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t^{(i)}) = \sigma^2_{git}$, where $\nexists\, \sigma^2_{gi} : \sigma^2_{git} = \sigma^2_{gi}$ for all t.*

Before we study the probability limit of $TNR_T^{(i)}(\hat\theta_T^{(i)})$, we also assume the following:

Assumption 12 *$\{\sigma^2_{git}; t = 1, 2, ...\}$ is a strictly stationary process.*

Assumption 13 *$T^{1/2} m_T^{(i)}(\theta_*^{(i)}) \xrightarrow{d} N(0, S^{(i)}(\theta_*^{(i)}))$ for $i = 1, 2$.*

---

[10] See White (2001, Definition 3.74).

[11] See White (2001, p. 58).

We also assume:

Assumption 14 $T^{1/2}[G_T^{(i)}(\theta_*^{(i)}) - G_*^{(i)}] \xrightarrow{d} N(0, S_g^{(i)})$, *where* $plim(G_T^{(i)}(\theta_*^{(i)})) = G_*^{(i)}$.

Assumption 14 will help us derive the asymptotic distributions of our statistics, in the case where models under consideration are misspecified. In the next sections, we study the asymptotic behavior of $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$ and B_T under different assumptions about $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ..., T\}$. Formally, we cover the following cases[12]:

(i) $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{gi}^2$,

(ii) $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{git}^2$, where $\nexists \sigma_{gi}^2 : \sigma_{git}^2 = \sigma_{gi}^2$ for all t,

(iii) $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is a dependent process[13].

We leave Case (i) to the Appendix.

1.3 Correct specification: GMM

In this section, we assume that both of the models considered are correctly specified and $\hat{\theta}_T^{(i)}$ represents the GMM estimator.

In order to motivate the discussion of model selection in the context of RE models estimated by GMM estimation, we first define what we mean by correct specification [14].

Definition 2 *(Correctly Specified Models).*

The model, $M^{(i)}$*, is said to be correctly specified if there exists a unique value* $\theta_*^{(i)}$ *in* $\Theta^{(i)} \subset \mathscr{R}^{p_i}$*, such that* $E[g_t^{(i)}(v_t, \theta_*^{(i)})z_t^{(i)}] = 0$ *for* $i = 1, 2$*, where* v_t *is a vector of observed random variables,* $g_t^{(i)}$ *is the RE error term and* $\hat{\theta}_T^{(i)} \xrightarrow{p} \theta_*^{(i)}$.

---

[12] When $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{gi}^2$, the distribution of $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$ is identical to that of the GMM minimand, which is covered in Hall and Pelletier (2008).

[13] When the models under consideration are misspecified in the non-local sense, one can no longer assume an m.d.s. However, the data generating process for $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ can still be serially uncorrelated for the case of non-local misspecification.

[14] See Definition 1 of Hall and Inoue (2003).

1.3.1 The limiting distribution of $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$: GMM

In the Appendix, we show that the probability limit of $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ is degenerate. In this section, we study the limiting distribution of $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$. Also, we will use some of the equations in this section for computing the limiting distribution of B_T in the next section.

1.3.2 $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{git}^2$

In this section, we analyze the probability limit of TNR_T when Assumptions 1-12 hold. We begin studying the asymptotic distribution of NR_T, by a mean value expansion of $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ [15] around $\theta_*^{(i)}$.

$$NR_T^{(i)}(\hat{\theta}_T^{(i)}) = NR_T^{(i)}(\theta_*^{(i)}) + \left.\frac{\partial NR_T^{(i)}}{\partial \theta^{(i)\prime}}\right|_{\theta^{(i)}=\bar{\theta}^{(i)}} (\hat{\theta}_T^{(i)} - \theta_*) \tag{1.29}$$

where $\bar{\theta}^{(i)} = \lambda_T \theta_*^{(i)} + (1-\lambda_T)\hat{\theta}_T^{(i)}$ for some $0 < \lambda_T < 1$ and $\hat{\theta}_T^{(i)}$ denotes the second-step GMM estimator for the i^{th} model.

Multiplying through by T yields:

$$TNR_T^{(i)}(\hat{\theta}_T^{(i)}) = TNR_T^{(i)}(\theta_*^{(i)}) + T^{1/2} \left.\frac{\partial NR_T^{(i)}}{\partial \theta^{(i)\prime}}\right|_{\theta^{(i)}=\bar{\theta}^{(i)}} T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{1.30}$$

$$\begin{aligned}
\frac{\partial NR_T^{(i)}}{\partial {\theta^{(i)}}'} &= 2\left(\frac{\partial NR_T^{(i)}}{\partial {m_T^{(i)}}'}\right)'\left(\frac{\partial m_T^{(i)}}{\partial \theta^{(i)\prime}}\right) \\
&\quad + \left(\frac{\partial NR_T^{(i)}}{\partial vec(W_T^{(i)-1})}\right)\left(\frac{\partial vec(W_T^{(i)-1})}{\partial vec(W_T^{(i)})'}\right)\left(\frac{\partial vec(W_T^{(i)})}{\partial \theta^{(i)\prime}}\right) && (1.31) \\
&= 2m_T^{(i)}(\theta^{(i)})' W_T^{(i)}(\theta^{(i)})^{-1} G_t^{(i)}(\theta^{(i)}) \\
&\quad + \left(m_T^{(i)}(\theta^{(i)})' \otimes m_T^{(i)}(\theta^{(i)})'\right)\left(\frac{\partial vec(W_T^{(i)-1})}{\partial vec(W_T^{(i)})'}\right)\left(\frac{\partial vec(W_T^{(i)})}{\partial \theta^{(i)\prime}}\right) && (1.32) \\
&= 2m_T^{(i)}(\theta^{(i)})' W_T^{(i)}(\theta^{(i)})^{-1} G_t^{(i)}(\theta^{(i)}) \\
&\quad - \left(m_T^{(i)}(\theta^{(i)})' \otimes m_T^{(i)}(\theta^{(i)})'\right)\left(W_T^{(i)}(\theta^{(i)})^{-1} \otimes W_T^{(i)}(\theta^{(i)})^{-1}\right)\left(\frac{\partial vec(W_T^{(i)})}{\partial \theta^{(i)\prime}}\right) && (1.33)
\end{aligned}$$

---

[15] Here, we proceed in a similar fashion to the proof of Rivers and Vuong (2002, Theorem 1, p. 27).

where the first line follows from the fatc that $NR_T^{(i)}(\theta^{(i)}) = m_T^{(i)}(\theta^{(i)})' W_T^{(i)}(\theta^{(i)})^{-1} m_T^{(i)}(\theta^{(i)})$. Further, $G_t^{(i)}(\theta^{(i)})$ is $k \times p$. $G_t^{(i)}(\theta_*^{(i)})$ is $O_p(1)$ by Assumption (8) and $W_T^{(i)}(\theta_*^{(i)})^{-1} \otimes W_T^{(i)}(\theta_*^{(i)})^{-1}$ is $O_p(1)$ by Assumption (10). Further,

$$\frac{\partial vec(W_T^{(i)})}{\partial \theta^{(i)\prime}} = 2\left[vec\left(T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}\right)\right]\left[T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)\left(\frac{\partial g_t^{(i)}(\theta_*^{(i)}, v_t)}{\partial \theta^{(i)\prime}}\right)\right] \tag{1.34}$$

Under our assumptions, we can show that $\partial vec(W_T^{(i)})/\partial\theta^{(i)\prime} = O_p(1)$ as follows:

$$\begin{aligned}
\left\|\frac{\partial vec(W_T^{(i)})}{\partial \theta^{(i)\prime}}\right\|_{\theta^{(i)}=\theta_*^{(i)}} &= 2\left\|\left[vec\left(T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}\right)\right]\left[T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)\left(\frac{\partial g_t^{(i)}(\theta_*^{(i)}, v_t)}{\partial \theta^{(i)\prime}}\right)\right]\right\| \\
&\le 2\left\|T^{-1}\sum_{t=1}^{T} vec\left(z_t z_t^{'}\right)\right\|\left\|T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)\left(\frac{\partial g_t^{(i)}(\theta_*^{(i)}, v_t)}{\partial \theta^{(i)\prime}}\right)\right\| \\
&\le 2T^{-1}\sum_{t=1}^{T}\|z_t^{(i)} \otimes z_t^{(i)}\| T^{-1}\sum_{t=1}^{T}\left\|g_t^{(i)}(\theta_*^{(i)}, v_t)\left(\frac{\partial g_t^{(i)}(\theta_*^{(i)}, v_t)}{\partial \theta^{(i)\prime}}\right)\right\| \\
&\le 2T^{-1}\sum_{t=1}^{T}\|z_t^{(i)} \otimes z_t^{(i)}\| T^{-1}\sum_{t=1}^{T}\left\|g_t^{(i)}(\theta_*^{(i)}, v_t)\right\|\left\|\frac{\partial g_t^{(i)}(\theta_*^{(i)}, v_t)}{\partial \theta^{(i)\prime}}\right\| \\
&\le 2T^{-1}\sum_{t=1}^{T}\|z_t^{(i)} \otimes z_t^{(i)}\|\left(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\left\|\frac{\partial g_t^{(i)}(\theta_*^{(i)}, v_t)}{\partial \theta^{(i)\prime}}\right\|^2\right)^{1/2} \qquad (1.35) \\
&= O_p(1)O_p(1)O_p(1) \\
&= O_p(1) \qquad (1.36)
\end{aligned}$$

where the second line follows from the first, via the Cauchy-Schwartz inequality for matrices and the fact that *vec* is a linear function. The third line follows from the second, via the triangle inequality and the fact that $vec(z_t z_t^{'}) = z_t^{(i)} \otimes z_t^{(i)}$. Equation (1.35) follows by applying the Cauchy-Schwartz inequality to the next line consecutively and recognizing that $g_t^{(i)}(\theta_*^{(i)}, v_t)$ is a scalar. Finally, the right-hand side of (1.35) is $O_p(1)$, by the Ergodic Theorem applied to each individual component under our assumptions.

Equations (1.33) and (1.36) yield:

$$\begin{aligned}
T^{1/2}\left[\left.\frac{\partial NR_T^{(i)}}{\partial\theta^{(i)\prime}}\right|_{\theta^{(i)}=\bar{\theta}^{(i)}}\right] &= 2T^{1/2}m_T^{(i)}(\theta_*^{(i)})'W_T^{(i)}(\theta_*^{(i)})^{-1}g_t^{(i)}(\theta_*^{(i)},v_t) \\
&- \left(T^{1/2}m_T^{(i)}(\theta_*^{(i)})'\otimes m_T^{(i)}(\theta_*^{(i)})'\right)\left(W_T^{(i)}(\theta_*^{(i)})^{-1}\otimes W_T^{(i)}(\theta_*^{(i)})^{-1}\right) \\
&\times\left(\frac{\partial vec(W_T^{(i)})}{\partial\theta^{(i)\prime}}\right) \qquad (1.37) \\
&= 2T^{1/2}m_T^{(i)}(\theta_*^{(i)})'W_T^{(i)}(\theta_*^{(i)})^{-1}g_t^{(i)}(\theta_*^{(i)},v_t) \\
&- \left(T^{1/2}m_T^{(i)}(\theta_*^{(i)})'\otimes m_T^{(i)}(\theta_*^{(i)})'\right)\left(W_T^{(i)}(\theta_*^{(i)})^{-1}\otimes W_T^{(i)}(\theta_*^{(i)})^{-1}\right) \\
&\times\left(\frac{\partial vec(W_T^{(i)})}{\partial\theta^{(i)\prime}}\right) \qquad (1.38) \\
&= 2T^{1/2}m_T^{(i)}(\theta_*^{(i)})'W_T^{(i)}(\theta_*^{(i)})^{-1}g_t^{(i)}(\theta_*^{(i)},v_t)+o_p(1) \qquad (1.39)
\end{aligned}$$

where (1.39) follows from Assumptions 8, 10 and 13, under the fact that for correctly specified models $E[g_t^{(i)}(\theta_*^{(i)},v_t)z_t^{(i)}]=0$.
Further, $\hat{\theta}_T^{(i)}$ represents the GMM estimator, hence we have:

$$\begin{aligned}
T^{1/2}(\hat{\theta}_T^{(i)}-\theta_*^{(i)}) &= -[G_T^{(i)}(\hat{\theta}_T^{(i)})'\hat{S}_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}G_T^{(i)}(\bar{\theta}^{(i)})]^{-1}G_T^{(i)}(\hat{\theta}_T^{(i)})'\hat{S}_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}T^{1/2}m_T^{(i)}(\theta_*^{(i)}) \qquad (1.40) \\
&= -[G_t^{(i)}(\theta_*^{(i)})'\hat{S}_T^{(i)}(\theta_*^{(i)})^{-1}G_T^{(i)}(\bar{\theta}^{(i)})]^{-1}G_t^{(i)}(\theta_*^{(i)})'S_T^{(i)}(\theta_*^{(i)})^{-1}T^{1/2}m_T^{(i)}(\theta_*^{(i)})+o_p(1) \qquad (1.41)
\end{aligned}$$

First, Equation (1.41) implies that $T^{1/2}(\hat{\theta}_T^{(i)}-\theta_*^{(i)})=O_p(1)$, under our assumptions. Further, substituting Equations (1.39) and (1.41) into Equation (1.30) yields:

$$\begin{aligned}
TNR_T^{(i)}(\hat{\theta}_T^{(i)}) &= TNR_T^{(i)}(\theta_*^{(i)})-2\big[T^{1/2}m_T^{(i)}(\theta_*^{(i)})'W_T^{(i)}(\theta_*^{(i)})^{-1}G_T^{(i)}(\theta_*^{(i)}) \\
&\times[G_T^{(i)}(\theta_*^{(i)})'S_T^{(i)}(\theta_*^{(i)})^{-1}G_T^{(i)}(\bar{\theta}^{(i)})]^{-1}G_T^{(i)}(\theta_*^{(i)})'S_T^{(i)}(\theta_*^{(i)})^{-1}T^{1/2}m_T^{(i)}(\theta_*^{(i)})\big]+o_p(1) \qquad (1.42) \\
&= T^{1/2}m_T^{(i)}(\theta_*^{(i)})'\big(W_T^{(i)}(\theta_*^{(i)})^{-1}[I-P_T^{(i)}]\big)T^{1/2}m_T^{(i)}(\theta_*^{(i)})+o_p(1) \qquad (1.43) \\
&= T^{1/2}m_T^{(i)}(\theta_*^{(i)})'F_T^{(i)}T^{1/2}m_T^{(i)}(\theta_*^{(i)})+o_p(1) \qquad (1.44)
\end{aligned}$$

where

$$F_T^{(i)} = W_T^{(i)}(\theta_*^{(i)})^{-1}[I - P_T^{(i)}] \tag{1.45}$$

and

$$\begin{aligned} P_T^{(i)} &= G_T^{(i)}(\theta_*^{(i)})[G_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} G_T^{(i)}(\bar{\theta}^{(i)})]^{-1} G_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} && (1.46) \\ &= G_T^{(i)}(\theta_*^{(i)})[G_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} G_T(\theta_*)]^{-1} G_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} + o_p(1) && (1.47) \end{aligned}$$

We continue studying the asymptotic distribution of $T^{1/2} m_T^{(i)}(\theta_*^{(i)})' F_T^{(i)} T^{1/2} m_T^{(i)}(\theta_*^{(i)})$. Using Assumptions 10 and 13, and taking $S^{(i)} \equiv S(\theta_*^{(i)})$ we can write the following:

$$\begin{aligned} TNR_T^{(i)}(\hat{\theta}_T^{(i)}) &= T^{1/2} m_T^{(i)}(\theta_*^{(i)})' F_T^{(i)} T^{1/2} m_T^{(i)}(\theta_*^{(i)}) + o_p(1) && (1.48) \\ &\xrightarrow{d} n_{k_i}' S^{(i)1/2} F^{(i)} S^{(i)1/2} n_{k_i} && (1.49) \end{aligned}$$

where $F^{(i)} = W_*^{(i)}(\theta_*^{(i)})^{-1}[I_{k_i} - P_*^{(i)}]$, where $W_*^{(i)} = plim(W_T^{(i)})$, $P_*^{(i)}$ is the plim of the right-hand side of (1.47), $n_{k_i} \sim N(0, I_{k_i})$, $S^{(i)1/2} n_{k_i} \sim N(0, S^{(i)})$, hence:

$$\begin{aligned} S^{(i)1/2\,\prime} F_T^{(i)} S^{(i)1/2} &\xrightarrow{p} S^{(i)1/2\,\prime} F^{(i)} S^{(i)1/2} && (1.50) \\ &= C^{(i)\prime} \Lambda^{(i)} C^{(i)} && (1.51) \end{aligned}$$

where $C^{(i)}$ is an orthogonal matrix of eigenvectors of $S^{(i)1/2\,\prime} F^{(i)} S^{(i)1/2}$ and $\Lambda^{(i)} = diag(\lambda_1^{(i)}, \lambda_2^{(i)}, ..., \lambda_k^{(i)})$, with corresponding eigenvalues $\lambda_1^{(i)}, \lambda_2^{(i)}, ..., \lambda_k^{(i)}$. The orthogonality of $C^{(i)}$ implies that $C^{(i)} n_{k_i}$ has the same asymptotic distribution as n_{k_i}. Therefore, we have:

$$TNR_T^{(i)}(\hat{\theta}_T^{(i)}) \xrightarrow{d} n_{k_i}' C^{(i)\prime} \Lambda^{(i)} C^{(i)} n_{k_i} \tag{1.52}$$

In other words, $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$ is distributed as a mixture of χ^2 random variables, which we state formally in the following theorem:[16].

Theorem 1 *When Assumptions 1-13 hold, we have:*

$$TNR_T^{(i)}(\hat{\theta}_T^{(i)}) \xrightarrow{d} \sum_{j=1}^{k} \lambda_j n_{k_i j}^2 \tag{1.53}$$

---

[16] Given $n_{qi} \sim N(0,1)$, it follows that $n_{k_i}^2 \sim \chi_1^2$.

1.3.3 Distribution of $\mathscr{K}_T$ under correct specification: GMM

In the case of correctly specified models, we know from the previous section that when $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{git}^2$, $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$ is distributed as a mixture of χ^2 random variables. Therefore, we will use T as the correct rate of convergence to derive the asymptotic distribution of:

$$\mathscr{K}_T = T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\} \tag{1.54}$$

In this section, we study the asymptotic distribution of $T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}$, under the case when:[17],

(i) $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{git}^2$,

(ii) The model under consideration is correctly specified, that is, $E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}] = 0$.

First, the following assumption is imposed:

Assumption 15 $T^{1/2}m_T^{(i)}(\theta_*^{(i)}) \xrightarrow{d} N(0, S(\theta_*^{(i)}))$ *and*

(i) $S(\theta_*^{(i)})$ *is a positive definite matrix of constants,*

(ii) $rank(G_*^{(i)}(\theta_*^{(i)})) = p_i$,

(iii) $\hat{S}(\hat{\theta}^{(i)}) \xrightarrow{p} S^{(i)}(\theta_*^{(i)})$.

where $\theta_* = [\theta_*^{(1)'}, \theta_*^{(2)'}]'$, $m_T(\theta_*)$ is $(k_1 + k_2) \times 1$, which is given as:

$$m_T(\theta_*) = \begin{bmatrix} m_T^{(1)}(\theta_*^{(1)}) \\ m_T^{(2)}(\theta_*^{(2)}) \end{bmatrix} \tag{1.55}$$

$$S_T(\theta_*) = \begin{bmatrix} S^{(1)}(\theta_*^{(1)}) & S^{(1,2)}(\theta_*^{(i)}) \\ S^{(1,2)}(\theta_*^{(i)}) & S^{(2)}(\theta_*^{(2)}) \end{bmatrix} \tag{1.56}$$

where $S_T(\theta_*)$ is $(k_1 + k_2) \times (k_1 + k_2)$.

Next, we write Equation (1.44) for model i as follows:

$$TNR_T^{(i)}(\hat{\theta}_T^{(i)}) = T^{1/2}m_T^{(i)}(\theta_*^{(i)})F_T^{(i)}(\theta_*^{(i)})T^{1/2}m_T(\theta_*^{(i)}) + o_p(1) \tag{1.57}$$

---

[17] The derivations in this section are very similar to the proof of Theorem 1 of Hall and Pelletier (2008).

Subtracting (1.57) for $i = 1, 2$, we have:

$$\begin{aligned} T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\} &= T^{1/2} m_T^{(1)}(\theta_*^{(1)}) F_T^{(1)}(\theta_*^{(1)}) T^{1/2} m_T^{(1)}(\theta_*^{(1)}) \\ &\quad - T^{1/2} m_T^{(2)}(\theta_*^{(2)}) F_T^{(2)}(\theta_*^{(2)}) T^{1/2} m_T^{(2)}(\theta_*^{(2)}) \\ &\quad + o_p(1) \qquad (1.58) \\ &= T^{1/2} m_T(\theta_*)' U_T(\theta_*) T^{1/2} m_T(\theta_*) + o_p(1) \qquad (1.59) \end{aligned}$$

where matrix $U_T(\theta_*)$ is given as follows:

$$\begin{aligned} U_T(\theta_*) &= \begin{bmatrix} F_T^{(1)}(\theta_*^{(1)}) & 0 \\ 0 & -F_T^{(2)}(\theta_*^{(2)}) \end{bmatrix} \qquad (1.60) \\ &= \begin{bmatrix} W_T^{(1)}(\theta_*^{(1)})^{-1}[I_{k_1} - P_T^{(1)}(\theta_*^{(1)})] & 0 \\ 0 & -W_T^{(2)}(\theta_*^{(2)})^{-1}[I_{k_2} - P_T^{(2)}(\theta_*^{(2)})] \end{bmatrix} \qquad (1.61) \end{aligned}$$

where $P_T^{(i)}(\theta_*^{(i)})$ is given by Equation (1.47).

Let $\eta_{k_1+k_2} \sim N(0, I_{k_1+k_2})$.

Then, under our assumptions, it follows that:

$$T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\} \xrightarrow{d} \eta_{k_1+k_2}' S(\theta_*)^{1/2} U_* S(\theta_*)^{1/2} \eta_{k_1+k_2} \qquad (1.62)$$

where $U_* = plim(U_T(\theta_*))$.

Given Equation (1.62), the asymptotic distribution of $T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}$, under the current case, is given by the following theorem:

Theorem 2 *Under Assumptions 1-13 and 15 we have:*

$$T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\} \xrightarrow{d} \eta_{k_1+k_2}' S(\theta_*)^{1/2} U_* S(\theta_*)^{1/2} \eta_{k_1+k_2} \qquad (1.63)$$

1.4 Distribution of B_T under misspecification: GMM

In this section, we derive the asymptotic distribution of the Rivers and Vuong's test statistic, B_T, given by Equation (1.20), where the goodness of fit criterion is $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ and both models under consideration have non-local misspecification.

For that purpose, we first define what we mean by a misspecified model[18].

Definition 3 *Misspecified Models The model, $M^{(i)}$ is said to be misspecified if, for $E[g^{(i)}(v_t, \theta^{(i)})z_t^{(i)}] = \mu^{(i)}(\theta^{(i)})$, we have*

$$\|\mu^{(i)}(\theta^{(i)})\| \neq 0 \qquad \forall\, \theta^{(i)} \in \Theta^{(i)} \tag{1.64}$$

We begin our analysis by multiplying Equation (1.29) through by $T^{1/2}$ as follows:

$$T^{1/2}NR_T^{(i)}(\hat{\theta}_T^{(i)}) = T^{1/2}NR_T^{(i)}(\theta_*^{(i)}) + \left.\frac{\partial NR_T^{(i)}}{\partial \theta^{(i)\prime}}\right|_{\theta^{(i)}=\bar{\theta}^{(i)}} T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{1.65}$$

Subtracting Equation (1.65) for models $i = 1, 2$ and recognizing that under H_0, we have $NR_*^{(1)}(\theta_*^{(1)}) = NR_*^{(2)}(\theta_*^{(2)})$, we can write the following:

$$\begin{aligned} T^{1/2}\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\} &= T^{1/2}\{NR_T^{(1)}(\theta_*^{(1)}) - NR_T^{(2)}(\theta_*^{(2)})\} \\ &+ \left.\frac{\partial NR_T^{(1)}}{\partial \theta^{(1)\prime}}\right|_{\theta^{(1)}=\bar{\theta}^{(1)}} T^{1/2}(\hat{\theta}_T^{(1)} - \theta_*^{(1)}) \\ &- \left.\frac{\partial NR_T^{(2)}}{\partial \theta^{(2)\prime}}\right|_{\theta^{(2)}=\bar{\theta}^{(2)}} T^{1/2}(\hat{\theta}_T^{(2)} - \theta_*^{(2)}) \qquad (1.66) \\ &= T^{1/2}\{NR_T^{(1)}(\theta_*^{(1)}) - NR_*^{(1)}(\theta_*^{(1)})\} \\ &- T^{1/2}\{NR_T^{(2)}(\theta_*^{(2)}) - NR_*^{(2)}(\theta_*^{(2)})\} \\ &+ \Pi^{(1)}(\bar{\theta}^{(1)})T^{1/2}(\hat{\theta}_T^{(1)} - \theta_*^{(1)}) \\ &- \Pi^{(2)}(\bar{\theta}^{(2)})T^{1/2}(\hat{\theta}_T^{(2)} - \theta_*^{(2)}) \qquad (1.67) \end{aligned}$$

where $\Pi^{(i)} = \partial NR_T^{(i)} / \partial \theta^{(i)^{\prime}}$.

Next, using the fact that $E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}] = \mu_*^{(i)}$, the derivative matrix in (1.65) converges

---

[18] This form of misspecification, as defined in Hall and Inoue (2003), is called non-local misspecification.

as follows:

$$\begin{aligned}
\left.\frac{\partial NR_T^{(i)}}{\partial \theta^{(i)\prime}}\right|_{\theta^{(i)}=\bar{\theta}^{(i)}} &= \left.\left(\frac{\partial NR_T^{(i)}}{\partial m_T^{(i)\prime}}\right)^{\prime}\left(\frac{\partial m_T^{(i)}}{\partial \theta^{(i)\prime}}\right)\right|_{\theta^{(i)}=\bar{\theta}^{(i)}} \\
&+ \left.\left(\frac{\partial NR_T^{(i)}}{\partial vec(W_T^{(i)-1})}\right)\left(\frac{\partial vec(W_T^{(i)-1})}{\partial vec(W_T^{(i)})^{\prime}}\right)\left(\frac{\partial vec(W_T^{(i)})}{\partial \theta^{(i)\prime}}\right)\right|_{\theta^{(i)}=\bar{\theta}^{(i)}} \qquad (1.68) \\
&= 2m_T^{(i)}(\bar{\theta}^{(i)})^{\prime}W_T^{(i)}(\bar{\theta}^{(i)})^{-1}G_T^{(i)}(\bar{\theta}^{(i)}) \\
&+ \left(m_T^{(i)}(\bar{\theta}^{(i)})^{\prime}\otimes m_T^{(i)}(\bar{\theta}^{(i)})^{\prime}\right)\left(\frac{\partial vec(W_T^{(i)-1})}{\partial vec(W_T^{(i)})^{\prime}}\right)\left(\left.\frac{\partial vec(W_T^{(i)})}{\partial \theta^{(i)^{\prime}}}\right|_{\theta^{(i)}=\bar{\theta}^{(i)}}\right) \\
&\qquad (1.69) \\
&= 2m_T^{(i)}(\theta_*^{(i)})^{\prime}W_T^{(i)}(\theta_*^{(i)})^{-1}G_T^{(i)}(\theta_*^{(i)}) \\
&- \left(m_T^{(i)}(\theta_*^{(i)})^{\prime}\otimes m_T^{(i)}(\theta_*^{(i)})^{\prime}\right)\left(W_T^{(i)}(\theta_*^{(i)})^{-1}\otimes W_T^{(i)}(\theta_*^{(i)})^{-1}\right)\Sigma_T^{(i)}(\theta_*^{(i)}) \\
&+ o_p(1) \qquad (1.70) \\
&= 2\mu_*^{(i)\prime}W^{(i)}(\theta_*^{(i)})^{-1}G_*^{(i)}(\theta_*^{(i)}) \\
&+ (\mu_*^{(i)\prime}\otimes\mu_*^{(i)\prime})(W^{(i)}(\theta_*^{(i)})^{-1}\otimes W^{(i)}(\theta_*^{(i)})^{-1})\Sigma_*^{(i)} + o_p(1) \qquad (1.71) \\
&= \Pi_*^{(i)} + o_p(1) \qquad (1.72)
\end{aligned}$$

where $\Sigma_T^{(i)}(\theta^{(i)}) = \partial vec(W_T^{(i)})/\partial\theta^{(i)^{\prime}}$, $plim(\Sigma_T^{(i)}(\theta_*^{(i)})) = \Sigma_*^{(i)}$.
First of all,

$$\begin{aligned}
NR_T^{(i)}(\theta_*^{(i)}) - NR_*^{(i)}(\theta_*^{(i)}) &= m_T^{(i)}(\theta_*^{(i)})W_T^{(i)}(\theta_*^{(i)})^{-1}m_T^{(i)}(\theta_*^{(i)}) \\
&- m_*^{(i)}(\theta_*^{(i)})W_*^{(i)}(\theta_*^{(i)})^{-1}m_*^{(i)}(\theta_*^{(i)}) \qquad (1.73)
\end{aligned}$$

From Equation (1.73), we see that $NR_T^{(1)}(\theta_*^{(1)}) - NR_*^{(1)}(\theta_*^{(1)})$ is in the form of $\hat{h}^{(i)\prime}\hat{W}^{(i)}\hat{h}^{(i)} - h^{(i)\prime}W^{(i)}h^{(i)}$, which we can write as:[19]

$$\hat{h}^{(i)\prime}\hat{W}^{(i)}\hat{h}^{(i)} - h^{(i)\prime}W^{(i)}h^{(i)} = \hat{h}^{(i)\prime}\hat{W}^{(i)}(\hat{h}^{(i)} - h^{(i)}) + \hat{h}^{(i)\prime}(\hat{W}^{(i)} - W^{(i)})h^{(i)} + (\hat{h}^{(i)} - h^{(i)})W^{(i)}h^{(i)} \qquad (1.74)$$

---

[19] See Hall and Pelletier (2008)[p.35].

Applying the expansion in (1.74) to (1.73), with $\hat{h}^{(i)} = m_T^{(i)}(\theta_*^{(i)})$, we have the following:[20],

$$\begin{aligned}
T^{1/2}\{NR_T^{(i)}(\theta_*^{(i)}) - NR_*^{(i)}(\theta_*^{(i)})\} &= 2\mu^{(i)}(\theta_*^{(i)})' W^{(i)}(\theta_*^{(i)})^{-1} T^{-1/2} \sum_{t=1}^{T} [z_t^{(i)} g_t^{(i)}(\theta_*^{(i)}) - \mu_*^{(i)}] \\
&+ \mu_*^{(i)\prime} T^{-1/2}\{W_T^{(i)}(\theta_*^{(i)})^{-1} - W^{(i)}(\theta_*^{(i)})^{-1}\}\mu_*^{(i)} \\
&+ o_p(1) \qquad (1.75)
\end{aligned}$$

Also, we can write the following analog to Equation (41) of Hall and Pelletier (2008):

$$\begin{aligned}
\mu_*^{(i)\prime} T^{-1/2}\{W_T^{(i)}(\theta_*^{(i)})^{-1} - W^{(i)}(\theta_*^{(i)})^{-1}\}\mu_*^{(i)} &= \mu_*^{(i)\prime} W^{(i)}(\theta_*^{(i)})^{-1} T^{-1/2}\{W_T^{(i)}(\theta_*^{(i)}) - W^{(i)}(\theta_*^{(i)})\} \\
&\times W_T^{(i)}(\theta_T^{(i)})^{-1}\mu_*^{(i)} \\
&= \left[\mu_*^{(i)\prime} \otimes \mu_*^{(i)\prime}\right]\left[W^{(i)}(\theta_*^{(i)})^{-1} \otimes W^{(i)}(\theta_*^{(i)})^{-1}\right] B^{(i)} \\
&\times T^{-1/2} \sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)})^2 vech[z_t^{(i)} z_t^{(i)\prime} - E[z_t^{(i)} z_t^{(i)\prime}]] \\
&+ o_p(1) \qquad (1.76)
\end{aligned}$$

where B is the permutation matrix such that $vec(A) = Bvech(A)$, for some conformable matrix A.

Using Equation (1.76), we can re-write Equation (1.75) as follows:

$$\begin{aligned}
T^{1/2}\{NR_T^{(i)}(\theta_*^{(i)}) - NR_*^{(i)}(\theta_*^{(i)})\} &= 2\mu^{(i)}(\theta_*^{(i)})' W^{(i)}(\theta_*^{(i)})^{-1} T^{-1/2} \sum_{t=1}^{T} [z_t^{(i)} g_t^{(i)}(\theta_*^{(i)}) - \mu_*^{(i)}] \\
&- \left[\mu_*^{(i)\prime} \otimes \mu_*^{(i)\prime}\right]\left[W^{(i)}(\theta_*^{(i)})^{-1} \otimes W^{(i)}(\theta_*^{(i)})^{-1}\right] \\
&\times B^{(i)} T^{-1/2} \sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)})^2 vech[z_t^{(i)} z_t^{(i)\prime} - E[z_t^{(i)} z_t^{(i)\prime}]] \\
&+ o_p(1) \qquad (1.77)
\end{aligned}$$

Hall and Inoue (2003) show that when models under consideration are subject to non-local misspecification, given in Definition 3, the distribution of the GMM estimator depends on the choice of the weighting matrix and we can no longer use Equation (1.41). Instead, we

---

[20] See Equations (39)-(41) of Hall and Pelletier (2008).

have the following:[21]

$$T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) = \left[I_{p_i} - H_{0,T}^{(i)} L_T^{(i)}\right]^{-1} H_{0,T}^{(i)} \{H_{1,T}^{(i)} + H_{2,T}^{(i)} + H_{3,T}^{(i)}\} \tag{1.78}$$

$$= \Omega_T^{(i)} \{H_{1,T}^{(i)} + H_{2,T}^{(i)} + H_{3,T}^{(i)}\} \tag{1.79}$$

where

$$H_{0,T}^{(i)} = -[G_T^{(i)\prime}(\hat{\theta}_T^{(i)}) S_T^{(i)-1} G_T^{(i)}(\hat{\theta}_T^{(i)}, \theta_*^{(i)}, \lambda_T)]^{-1} \tag{1.80}$$

$$H_{1,T}^{(i)} = G_T^{(i)\prime}(\hat{\theta}_T^{(i)}) S_T^{(i)-1} T^{-1/2} \sum_{t=1}^{T} [g_t^{(i)} z_t^{(i)} - \mu_*^{(i)}] \tag{1.81}$$

$$H_{2,T}^{(i)} = T^{1/2}[G_T^{(i)}(\theta_*^{(i)}) - G_*^{(i)}]' S_T^{(i)} \mu_*^{(i)} \tag{1.82}$$

$$H_{3,T}^{(i)} = G_*^{(i)\prime} T^{1/2} [S_T^{(i)-1} - S^{(i)-1}] \mu_*^{(i)} \tag{1.83}$$

$$L_T^{(i)} = (\mu_*^{(i)\prime} S_T^{(i)-1} \otimes I_{p_i}) G_T^{(i)2}(\bar{\theta}_T^{(i)}) \tag{1.84}$$

$$G_T^{(i)2}(\bar{\theta}_T^{(i)}) = (\partial / \partial \theta^{(i)\prime}) vec(G_T^{(i)}(\theta)) \tag{1.85}$$

$$\Omega_T^{(i)} = [I_{p_i} - H_{0,T}^{(i)} L_T^{(i)}]^{-1} H_{0,T}^{(i)} \tag{1.86}$$

where $plim(G_T^{(i)}(\theta_*^{(i)})) = G_*^{(i)}$.

In our case, the choice of weighting matrix is $S_T^{(i)} = T^{-1} \sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}$, hence:

$$\begin{aligned} G_*^{(i)}(\theta_*^{(i)})' T^{1/2} [S_T^{(i)-1} - S^{(i)-1}] \mu_*^{(i)} &= -G_*^{(i)}(\theta_*^{(i)})' M_{zz}^{(i)-1} T^{-1/2} \left[\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime} - M_{zz}^{(i)} \right] \\ &\quad \times \left[T^{-1} \sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime} \right]^{-1} \mu_*^{(i)} + o_p(1) \qquad (1.87) \\ &= -(\mu_*^{(i)\prime} \otimes G_*^{(i)}(\theta_*^{(i)})')(M_{zz}^{(i)-1} \otimes M_{zz}^{(i)-1}) \\ &\quad \times V^{(i)} T^{-1/2} \sum_{t=1}^{T} vech[z_t^{(i)} z_t^{(i)\prime} - E[z_t^{(i)} z_t^{(i)\prime}]] + o_p(1) \end{aligned} \tag{1.88}$$

where $M_{zz}^{(i)} = E[z_t^{(i)} z_t^{(i)\prime}]$ and V is the permutation matrix such that $vec(A) = V vech(A)$ for some conformable matrix A.

---

[21] See Equation (9) of Hall and Inoue (2003)[p.368].

Let $\hat{D}^{(i)}_{zz,t} = \partial g^{(i)}_t(\theta)/\partial\theta^{(i)\prime}]' z^{(i)\prime}_t$ and $D^{(i)}_{zz} = E[z^{(i)}_t(\partial g^{(i)}_t(\theta)/\partial\theta^{(i)\prime})]'$. A similar analysis yields the following:

$$\begin{aligned} T^{1/2}[G^{(i)}_T(\theta^{(i)}_*) - G^{(i)}_*]' S^{(i)}_T \mu^{(i)}_* &= ([\mu^{(i)\prime}_* S^{(i)}_T] \otimes I_{p_i}) T^{1/2} vec([G^{(i)}_T(\theta^{(i)}_*) - G^{(i)}_*]') && (1.89) \\ &= ([\mu^{(i)\prime}_* M^{(i)-1}_{zz}] \otimes I_{p_i}) T^{-1/2} \sum_{t=1}^{T} vec\big(\hat{D}^{(i)}_{zz,t} - D^{(i)}_{zz}\big) && (1.90) \end{aligned}$$

Using (1.72), (1.77), (1.78) and (1.88), we can rearrange Equation (1.67) as follows:

$$\begin{aligned} T^{1/2}\big\{NR^{(1)}_T(\hat{\theta}^{(1)}_T) - NR^{(2)}_T(\hat{\theta}^{(2)}_T)\big\} &= NR^{(1)}_{T,1}(\theta^{(1)}_*) + NR^{(1)}_{T,2}(\theta^{(1)}_*) \\ &\quad + NR^{(2)}_{T,1}(\theta^{(2)}_*) + NR^{(2)}_{T,2}(\theta^{(2)}_*) && (1.91) \\ &= L'_T T^{1/2} D_T + o_p(1) && (1.92) \end{aligned}$$

where

$$\begin{aligned} NR^{(1)}_{T,1}(\theta^{(1)}_*) &= 2\mu^{(1)}(\theta^{(1)}_*)' W^{(1)}(\theta^{(1)}_*)^{-1} T^{-1/2} \sum_{t=1}^{T} [g^{(1)}_t(\theta^{(1)}_*) z^{(1)}_t - \mu^{(1)}_*] \\ &\quad + \big[\mu^{(1)\prime}_* \otimes \mu^{(1)\prime}_*\big]\big[W^{(1)}(\theta^{(1)}_*)^{-1} \otimes W^{(1)}(\theta^{(1)}_*)^{-1}\big] \\ &\quad \times B^{(1)} T^{-1/2} \sum_{t=1}^{T} g^{(1)}_t(\theta^{(1)}_*)^2 vech[z^{(1)}_t z^{(1)\prime}_t - E[z^{(1)}_t z^{(1)\prime}_t]] && (1.93) \\ NR^{(2)}_{T,1}(\theta^{(2)}_*) &= 2\mu^{(2)}(\theta^{(2)}_*)' W^{(2)}(\theta^{(2)}_*)^{-1} T^{-1/2} \sum_{t=1}^{T} [z^{(2)}_t g^{(2)}_t(\theta^{(2)}_*) - \mu^{(2)}_*] \\ &\quad + \big[\mu^{(2)\prime}_* \otimes \mu^{(2)\prime}_*\big]\big[W^{(2)}(\theta^{(2)}_*)^{-1} \otimes W^{(2)}(\theta^{(2)}_*)^{-1}\big] \\ &\quad \times B^{(2)} T^{-1/2} \sum_{t=1}^{T} g^{(2)}_t(\theta^{(2)}_*)^2 vech[z^{(2)}_t z^{(2)\prime}_t - E[z^{(2)}_t z^{(2)\prime}_t]] && (1.94) \\ NR^{(1)}_{T,2}(\theta^{(2)}_*) &= \Pi^{(1)}_* \Omega^{(1)}_* \Bigg[G^{(1)\prime}_T(\theta^{(1)}_*) M^{(i)-1}_{zz} T^{-1/2} \sum_{t=1}^{T} [g^{(1)}_t(\theta^{(1)}_*) z^{(1)}_t - \mu^{(1)}_*] \\ &\quad + ([\mu^{(1)\prime}_* M^{(1)-1}_{zz}] \otimes I_{p_1}) T^{-1/2} \sum_{t=1}^{T} vec\big(\hat{D}^{(1)}_{zz,t} - D^{(1)}_{zz}\big) \\ &\quad - (\mu^{(1)\prime}_* \otimes G^{(1)}_*(\theta^{(1)}_*)')(M^{(1)-1}_{zz} \otimes M^{(1)-1}_{zz}) \\ &\quad \times V^{(1)} T^{-1/2} \sum_{t=1}^{T} vech[z^{(1)}_t z^{(1)\prime}_t - E[z^{(1)}_t z^{(1)\prime}_t]] \Bigg] && (1.95) \end{aligned}$$

$$
\begin{aligned}
NR^{(2)}_{T,2}(\theta^{(2)}_*) = \Pi^{(2)}_*\Omega^{(2)}_*\Bigg[& G^{(2)\prime}_T(\theta^{(2)}_*)M^{(2)-1}_{zz}T^{-1/2}\sum_{t=1}^{T}[g^{(2)}_t(\theta^{(2)}_*)z^{(2)}_t - \mu^{(2)}_*] \\
& + ([\mu^{(2)\prime}_* M^{(2)-1}_{zz}] \otimes I_{p_2})T^{-1/2}\sum_{t=1}^{T} vec\big(\hat{D}^{(2)}_{zz,t} - D^{(2)}_{zz}\big) \\
& - (\mu^{(2)\prime}_* \otimes G^{(2)}_*(\theta^{(2)}_*)')(M^{(2)-1}_{zz} \otimes M^{(2)-1}_{zz}) \\
& \times V^{(2)}T^{-1/2}\sum_{t=1}^{T} vech[z^{(2)}_t z^{(2)\prime}_t - E[z^{(2)}_t z^{(2)\prime}_t]]\Bigg]
\end{aligned}
\tag{1.96}
$$

where

$$
L_T(\theta^{(i)}_*)' = \begin{bmatrix}
2\mu^{(1)}(\theta^{(1)}_*)'W^{(1)}(\theta^{(1)}_*)^{-1} + \Pi^{(1)}_*\Omega^{(1)}_*G^{(1)\prime}_T(\theta^{(1)}_*)M^{(1)-1}_{zz} \\
\big[\mu^{(1)\prime}_* \otimes \mu^{(1)\prime}_*\big]\big[W^{(1)}(\theta^{(1)}_*)^{-1} \otimes W^{(1)}(\theta^{(1)}_*)^{-1}\big]B^{(1)} \\
\Pi^{(1)}_*\Omega^{(1)}_*\big([\mu^{(1)\prime}_* M^{(1)-1}_{zz}] \otimes I_{p_1}\big) \\
\Pi^{(1)}_*\Omega^{(1)}_*\big(\mu^{(1)\prime}_* \otimes G^{(1)}_*(\theta^{(1)}_*)'\big)\big(M^{(1)-1}_{zz} \otimes M^{(1)-1}_{zz}\big)V^{(1)} \\
2\mu^{(2)}(\theta^{(2)}_*)'W^{(2)}(\theta^{(2)}_*)^{-1} + \Pi^{(2)}_*\Omega^{(2)}_*G^{(2)\prime}_T(\theta^{(2)}_*)M^{(2)-1}_{zz} \\
\big[\mu^{(2)\prime}_* \otimes \mu^{(2)\prime}_*\big]\big[W^{(2)}(\theta^{(2)}_*)^{-1} \otimes W^{(2)}(\theta^{(2)}_*)^{-1}\big]B^{(2)} \\
\Pi^{(2)}_*\Omega^{(2)}_*\big([\mu^{(2)\prime}_* M^{(2)-1}_{zz}] \otimes I_{p_2}\big) \\
\Pi^{(2)}_*\Omega^{(2)}_*\big(\mu^{(2)\prime}_* \otimes G^{(2)}_*(\theta^{(2)}_*)'\big)\big(M^{(2)-1}_{zz} \otimes M^{(2)-1}_{zz}\big)V^{(2)}
\end{bmatrix}
\tag{1.97}
$$

and

$$
D_T = T^{-1}\sum_{t=1}^{T}\xi_t \tag{1.98}
$$

where

$$
\xi_t = \begin{bmatrix}
z^{(1)}_t g^{(1)}_t(\theta^{(1)}_*) - \mu^{(1)}_* \\
g^{(1)}_t(\theta^{(1)}_*)^2 vech[z^{(1)}_t z^{(1)\prime}_t - E[z^{(1)}_t z^{(1)\prime}_t]] \\
vec\big(\hat{D}^{(1)}_{zz,t} - D^{(1)}_{zz}\big) \\
vech[z^{(1)}_t z^{(1)\prime}_t - E[z^{(1)}_t z^{(1)\prime}_t]] \\
z^{(2)}_t g^{(2)}_t(\theta^{(2)}_*) - \mu^{(2)}_* \\
g^{(2)}_t(\theta^{(2)}_*)^2 vech[z^{(2)}_t z^{(2)\prime}_t - E[z^{(2)}_t z^{(2)\prime}_t]] \\
vec\big(\hat{D}^{(2)}_{zz,t} - D^{(2)}_{zz}\big) \\
vech[z^{(2)}_t z^{(2)\prime}_t - E[z^{(2)}_t z^{(2)\prime}_t]]
\end{bmatrix}
\tag{1.99}
$$

Next, the following assumption is imposed:

Assumption 16 *To derive our asymptotic results, we assume:*

(i) $L_T^{'}T^{-1/2}\sum_{t=1}^{T}\xi_t \xrightarrow{d} N(0, L_*^{'}V_*L_*)$,

(ii) $Rank(G_*^{(i)}(\theta_*^{(i)})) = p_i$,

(iii) $\hat{V}_T \xrightarrow{p} V_*$,

(iv) $[L_*^{'}V_*L_*]^{-1}$ *exists.*

where $V_* = lim_{T\to\infty}T^{-1}var(\sum_{t=1}^{T}\xi_t(\theta_*^{(i)}))$.

From Equation (1.41), it follows that $T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) = O_p(1)$. From Equations (1.92)-(1.99) and Assumption 16, we have the following limiting distribution for the numerator of B_T:

$$L_*^{'}V^{1/2}\eta_k \tag{1.100}$$

where $\eta_k \sim N(0, I_{k_i})$ and $k = 4(k_1 + k_2)$ and $L_T(\theta_*^{(i)})$ is given by (1.97).

Next, we study the limiting distribution of the denominator of B_T. For that purpose, we define the following:

$$V_T = var(T^{1/2}D_T) \tag{1.101}$$

$$= var(T^{-1/2}\sum_{t=1}^{T}\xi_t) \tag{1.102}$$

where ξ_t is given by (1.99).

Then, the limiting distribution of the denominator of B_T is given as:

$$\rho_T \xrightarrow{p} [L_*{'}V_*L_*]^{1/2} \tag{1.103}$$

where $plim(L_T) = L_*$ and $V_* = var(T^{-1/2}\sum_{t=1}^{T}\xi_t)$, ξ_t is given by (1.99) and $L_T(\theta_*^{(i)})$ is given by (1.97).

Given Equations (1.91)-(1.103), we have the following limiting distribution of B_T:

Theorem 3 *When Assumptions 1-14, 15 and 16 hold, we have:*

$$B_T \xrightarrow{d} [L_*^{'}V_*L_*]^{-1/2}L_*^{'}V_*^{1/2}\eta_k \tag{1.104}$$

$$= N(0,1) \tag{1.105}$$

1.4.1 $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is a serially correlated process

In this section, we study the limiting distribution of B_T in the case of misspecified models where $g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is a serially correlated process. The dependence of the B_T statistic on the GMM estimator, through D_T, makes it natural to study the limiting distribution of B_T under the current case in parallel to Hall and Inoue (2003), where the weighting matrix, $S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}$, used in the two step GMM estimation, is the inverse of a centered HACC matrix estimator.

Formally, the weighting matrix used in the second-stage of GMM estimation takes the following form:

$$S_T^{(i)}(\hat{\theta}_T^{(i)}) = \sum_{i=-T+1}^{T-1} \omega(i/b_T)\tilde{\Gamma}_i \tag{1.106}$$

where

$$\begin{aligned}\tilde{\Gamma}_i &= T^{-1}\sum_{t=i+1}^{T}[g_t^{(i)}(v_t, \hat{\theta}_T^{(i)}(1))z_t^{(i)} - m_T(\hat{\theta}_{T,1}^{(i)})] \\ &\times [g_t^{(i)}(v_{t-i}, \hat{\theta}_T^{(i)}(1))z_t^{(i)} - m_T^{(i)}(\hat{\theta}_{T,1}^{(i)})]' \quad \text{for } i \geq 0 \qquad (1.107)\\ &= T^{-1}\sum_{t=-i+1}^{T}[g_t^{(i)}(v_{t+i}, \hat{\theta}_T^{(i)}(1))z_t^{(i)} - m_T^{(i)}(\hat{\theta}_{T,1}^{(i)})] \\ &\times [g_t^{(i)}(v_t, \hat{\theta}_T^{(i)}(1))z_t^{(i)} - m_T^{(i)}(\hat{\theta}_{T,1}^{(i)})]' \quad \text{for } i < 0. \qquad (1.108)\end{aligned}$$

where ω represents weights called *kernels*, b_T represents bandwidth and $\hat{\theta}_{T,1}^{(i)}$ represents the first-step GMM estimator for models $i = 1, 2$.

Under this case, Hall and Inoue (2003) show that the correct convergence rate for the GMM estimator is given by $(T/b_T)^{1/2}$ and is no longer equal to $T^{1/2}$.

Next, we state Assumptions (6) of Hall and Inoue (2003).

Assumption 17 *(i) For all $x \in \Re$, $|\omega(x)| \leq 1$, $\omega(-x) = \omega(x)$, $\omega(0) = 1$, $\omega(x)$ is continuous at zero and for almost all $x \in \Re$, $\int_{\Re}\omega(x)^2dx < \infty$, $\int_{\Re}\omega(x)e^{-ix\lambda}dx \geq 0$ for all $\lambda \in \Re$.*

(ii) $\int_{-\infty}^{\infty}\omega(x)dx = c$ where $0 < c < \infty$.

(iii) $\int_{-\infty}^{\infty}x\omega(x)dx < \infty$.

(iv) $b_T = o(T^{1/2})$ *and* $b_T \to \infty$.

Next, we state Assumption (7) of Hall and Inoue (2003).

Assumption 18 *Let* $\hat{\theta}^{(i)}_{1,T}$, $\theta^{(i)}_*$ *denote the first step GMM estimator and its probability limit, respectively, and* $\mu^{(i)}_* = \mu^{(i)}(\theta^{(i)}_*)$.

(i)

$$\left(\frac{T}{b_T}\right) vech(\bar{V}_T - V_T) \xrightarrow{d} \qquad N(0, \Omega_v) \tag{1.109}$$

$$lim_{T\to\infty} b^l (V_T - V) = C \tag{1.110}$$

where Ω_V *is a* $k(k+1)/2 \times k(k+1)/2$ *matrix,* $l > 0$ *is the characteristic exponent of the kernel* ω .

(ii)

$$\bar{V}_T = \bar{\Gamma}_0 + \sum_{i=1}^{T-1} \omega(i/b_T)(\bar{\Gamma}_i + \bar{\Gamma}^{'}_i) \tag{1.111}$$

$$V_T = \bar{\Gamma}_0 + \sum_{i=1}^{T-1} \omega(i/b_T)(\Gamma_i + \Gamma^{'}_i) \tag{1.112}$$

$$\bar{\Gamma}_i = T^{-1} \sum_{t=i+1}^{T} [g_t(v_t, \theta) z_t - \mu_*][g_t(v_t, \theta) z_t - \mu_*]^{'} \tag{1.113}$$

$$C = -\lim_{x\to 0}\left(\frac{1-\omega(x)}{|x|^l}\right) \sum_{j=-\infty}^{\infty} |j|^l \Gamma_j < \infty \tag{1.114}$$

.

(iii) $\hat{\theta}_{1,T} - \theta_* = O_p(T^{-1/2})$,

(iv) $\max_{1\le i\le T} \|T^{-1} \sum_{t=1}^{i} g_t(v_t, \theta_*) z_t - \mu_*\| = O_p(T^{-1/2})$.

Assumption 17-(iv) and Assumption 18-(iv) together imply the following:

$$(Tb_T)^{-1/2} \sum_{t=1}^{T} g^{(i)}_t(\theta^{(i)}_*) z^{(i)}_t - \mu^{(i)}_* = o_p(1) \tag{1.115}$$

Further, we make the following assumption:

Assumption 19

$$max_{1\leq i\leq T}\|T^{-1}\sum_{t=1}^{i}[vec(\hat{D}^{(1)}_{zz,t}-D^{(1)}_{zz})]\| = O_p(T^{-1/2}) \tag{1.116}$$

$$max_{1\leq i\leq T}\|g^{(2)}_t(\theta^{(2)}_*)^2 vech[z^{(2)}_t z^{(2)\prime}_t - E[z^{(2)}_t z^{(2)\prime}_t]]\| = O_p(T^{-1/2}) \tag{1.117}$$

Under Assumption (19), it follows that:

$$(b_T T)^{-1/2}\sum_{t=1}^{T} vec\big(\hat{D}^{(1)}_{zz,t}-D^{(1)}_{zz}\big) = o_p(1) \tag{1.118}$$

$$(b_T T)^{-1/2}\sum_{t=1}^{T} g^{(2)}_t(\theta^{(2)}_*)^2 vech[z^{(2)}_t z^{(2)\prime}_t - E[z^{(2)}_t z^{(2)\prime}_t]] = o_p(1) \tag{1.119}$$

Therefore, we can write the following:

$$(T/b_T)^{1/2}D_T = \frac{1}{b_T}\sum_{t=1}^{T}\mathscr{Z}_T \tag{1.120}$$

$$\mathscr{Z}_T = \begin{bmatrix} g^{(1)}_t(\theta^{(1)}_*)z^{(1)}_t - \mu^{(1)}_* \\ vec\big(\hat{D}^{(1)}_{zz,t}-D^{(1)}_{zz}\big) \\ g^{(1)}_t(\theta^{(1)}_*)^2 vech[z^{(1)}_t z^{(1)\prime}_t - E[z^{(1)}_t z^{(1)\prime}_t]] \\ vech\big([g^{(1)}_t(\theta^{(1)}_*)z^{(1)}_t - \mu^{(1)}_*][g^{(1)}_t(\theta^{(1)}_*)z^{(1)}_t - \mu^{(1)}_*]'\big) \\ g^{(2)}_t(\theta^{(2)}_*)z^{(2)}_t - \mu^{(2)}_* \\ vec\big(\hat{D}^{(2)}_{zz,t}-D^{(2)}_{zz}\big) \\ g^{(2)}_t(\theta^{(2)}_*)^2 vech[z^{(2)}_t z^{(2)\prime}_t - E[z^{(2)}_t z^{(2)\prime}_t]] \\ vech\big([g^{(2)}_t(\theta^{(2)}_*)z^{(2)}_t - \mu^{(2)}_*][g^{(2)}_t(\theta^{(2)}_*)z^{(2)}_t - \mu^{(2)}_*]'\big) \end{bmatrix} \tag{1.121}$$

The long-run variance of $T^{1/2}D_T$ is given by:

$$V_* = lim_{T\to\infty} var(T^{1/2}D_T) \tag{1.122}$$

$$= lim_{T\to\infty} E[T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathscr{Z}_t\mathscr{Z}_s] \tag{1.123}$$

$$= \bar{\Gamma}_0 + \sum_{i=1}^{\infty}[\bar{\Gamma}_i + \bar{\Gamma}_i^{'}] \tag{1.124}$$

Given the form of D_T in (1.120), the long-run variance of $T^{1/2}D_T$ in (1.124) can consistently be estimated with the help of (1.107). Further, we write B_T as follows:

$$B_T = \frac{(T/b_T)^{1/2}\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}}{\hat{\rho}_T/b_T} \tag{1.125}$$

Under the current case, L_T is still given by (1.97).
Next, we impose the following analog to Assumption (16):

Assumption 20 *(i)* $L_T(1/b_TT)^{1/2}\sum_{t=1}^{T}\mathscr{Z}_T \xrightarrow{d} N(0, L_*^{'}V_*L_*)$,

(ii) $Rank(G_*^{(i)}(\theta_*^{(i)})) = p_i$,

(iii) $\hat{V}_T \xrightarrow{p} V_*$,

(iv) $L_*^{'}V_*L_*$ *is p.s.d.*[22]

where $V_* = lim_{T\to\infty}T^{-1}var(\sum_{t=1}^{T}\mathscr{Z}_T(\theta_*^{(i)}))$ and $plim(\hat{V}_T) = V_*$.
Applying the same analysis in Equations (1.97)-(1.103) to (1.125), we write the following analog to Theorem (3).

Theorem 4 *Under Assumptions 1-14 and 15-20, we have*

$$B_T \xrightarrow{d} [L_*^{'}V_*L_*]^{-1/2}L_*^{'}V^{1/2}\eta_k \tag{1.126}$$

$$= N(0,1) \quad if \quad T^{1/2}/b_T^{(1/2+k)} \to \phi \in [0,\infty) \tag{1.127}$$

where k is such that Equation (18) of Hall and Inoue (2003) holds.

---

[22] Assumption 20-(iv) implies $[L_*^{'}V_*L_*]^{-1}$ exists.

1.5 Validity of tests

In this section, we prove the validity of the test statistics that we studied in Chapter 1. Namely, we prove the consistency of $TNR_T(\hat\theta_T^{(i)})$, $\mathscr{K}_T$ for $\mathscr{K}_*$ and B_T for B_*.

1.5.1 Consistency of $\boldsymbol{TNR_T(\hat\theta_T^{(i)})}$

In Chapter 1, we saw that $TNR_T^{(i)}(\hat\theta_T^{(i)})$ has an asymptotic distribution, which is a mixture of chi-square distributed variables where the weights used are eigenvalues of functions of parameters for the distribution of the data. Formally, it follows that:

$$TNR_T^{(i)}(\hat\theta_T^{(i)}) \xrightarrow{d} \sum_{j=1}^{k} \lambda_j n_{kj}^2 \tag{1.128}$$

where λ_j are eigenvalues, which are defined in Chapter 1, $n_{kj} \sim N(0,1)$, and it follows that $n_{kj}^2 \sim \chi_1^2$.

Next, we prove the validity of simulations that we will use for $TNR_T(\hat\theta_T^{(i)})$. For that purpose, we first propose consistent estimators for λ_j, which are eigenvalues of the following matrix:

$$F_*(\theta_*^{(i)}) = W(\theta_*^{(i)})^{-1}[I - P_*(\theta_*^{(i)})] \tag{1.129}$$

where $W(\theta_*^{(i)})^{-1}$ is consistently estimated by:

$$W_T^{(i)}(\hat\theta_T^{(i)}) = \left(T^{-1}\sum_{t=1}^{T} g_t^2(\hat\theta_T^{(i)})\right)\left(T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}\right) \tag{1.130}$$

and

$$P_*(\theta_*^{(i)}) = G_*^{(i)}(\theta_*^{(i)})[G_*^{(i)}(\theta_*^{(i)})' S(\theta_*^{(i)})^{-1} G_*(\bar\theta^{(i)})]^{-1} G_*^{(i)}(\theta_*^{(i)})' S(\theta_*^{(i)})^{-1} \tag{1.131}$$

which can be consistently estimated by their sample counterparts, $F_T^{(i)}(\hat\theta_T^{(i)}) = W_T^{(i)}(\hat\theta_T^{(i)})^{-1}[I - P_T^{(i)}(\hat\theta_T^{(i)})]$ where:

$$F_T^{(i)}(\hat\theta_T^{(i)}) = W_T^{(i)}(\hat\theta_T^{(i)})^{-1}[I - P_T^{(i)}(\hat\theta_T^{(i)})] \tag{1.132}$$

and

$$P_T^{(i)}(\hat{\theta}_T^{(i)}) = G_T^{(i)}(\hat{\theta}_T^{(i)})[G_T^{(i)}(\hat{\theta}_T^{(i)})' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} G_T^{(i)}(\bar{\theta}^{(i)})]^{-1} G_T^{(i)}(\hat{\theta}_T^{(i)})' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \tag{1.133}$$

$$= G_T^{(i)}(\hat{\theta}_T^{(i)})[G_T^{(i)}(\hat{\theta}_T^{(i)})' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} G_T^{(i)}(\hat{\theta}_T)]^{-1} G_T^{(i)}(\hat{\theta}_T^{(i)})' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} + o_p(1) \tag{1.134}$$

In what follows, we show that the λ_j, which are the eigenvalues $F_*(\theta_*^{(i)})$, can be estimated consistently by the eigenvalues of $F_T(\hat{\theta}_T^{(i)})$, which we denote by $\hat{\lambda}_{Tj}$. The $\hat{\lambda}_{Tj}$, as eigenvalues of $F_T(\hat{\theta}_T^{(i)})$, satisfy:

$$F_T(\hat{\theta}_T^{(i)}) Q_T(\hat{\theta}_T^{(i)}) = \Lambda_T(\hat{\theta}_T^{(i)}) Q_T(\hat{\theta}_T^{(i)}) \tag{1.135}$$

where $\Lambda_T(\hat{\theta}_T^{(i)})$ is a diagonal matrix of eigenvalues of $F_T(\hat{\theta}_T^{(i)})$, with diagonal entries $\hat{\lambda}_{Tj}$, and $Q_T(\hat{\theta}_T^{(i)})$ is the associated matrix of eigenvectors.

Next, we show that $\Lambda_T(\hat{\theta}_T^{(i)}) \xrightarrow{p} \Lambda$, or equivalently, $\lambda_{Tj}(\hat{\theta}_T^{(i)}) \xrightarrow{p} \lambda_j$.

Taking limits on both sides of (1.135) yields:

$$plim(F_T(\hat{\theta}_T^{(i)}) Q_T(\hat{\theta}_T^{(i)})) = plim(\Lambda_T(\hat{\theta}_T^{(i)}) Q_T(\hat{\theta}_T^{(i)})) \tag{1.136}$$

$$plim(F_T(\hat{\theta}_T^{(i)})) plim(Q_T(\hat{\theta}_T^{(i)})) = plim(\Lambda_T(\hat{\theta}_T^{(i)})) plim(Q_T(\hat{\theta}_T^{(i)})) \tag{1.137}$$

$$F_*(\theta_*^{(i)}) plim(Q_T(\hat{\theta}_T^{(i)})) = plim(\Lambda_T(\hat{\theta}_T^{(i)})) plim(Q_T(\hat{\theta}_T^{(i)})) \tag{1.138}$$

where Equation (1.137) follows from the continuous mapping theorem and Slutsky's Lemma, given $\hat{\theta}_T \xrightarrow{p} \theta_*$. Looking at Equation (1.138), we see that $plim(\Lambda_T(\hat{\theta}_T^{(i)}))$ and $plim(Q_T(\hat{\theta}_T^{(i)}))$ are the eigenvalues and the associated eigenvectors of $F_*(\theta_*^{(i)})$, which, of course, are none other than $\Lambda_*(\theta_*^{(i)})$ and $Q_T(\theta_*^{(i)})$, respectively. In this case, the eigenvalues for $F_*(\theta_*^{(i)})$ can be consistently estimated by the eigenvalues of $F_T(\hat{\theta}_T^{(i)})$.

1.5.2 Validity of $\mathscr{K}_T$

In this section, we show that $\mathscr{K}_T$, in Equation (1.54), is a consistent estimator of $\mathscr{K}_*(\theta_*^{(i)})$. This immediately follows from the analysis in Equations (1.29)-(1.48) and (1.57)-(1.62), where

$$F_*(\theta_*^{(i)}) = W^{(i)}(\theta_*^{(i)})^{-1}[I - P_*(\theta_*^{(i)})] \tag{1.139}$$

where $W(\theta_*^{(i)})^{-1}$ is consistently estimated by:

$$W_T^{(i)}(\hat{\theta}_T^{(i)}) = \left(T^{-1}\sum_{t=1}^{T} g_t^2(\hat{\theta}_T^{(i)})\right)\left(T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}\right) \tag{1.140}$$

and

$$P_*(\theta_*^{(i)}) = G_*^{(i)}(\theta_*^{(i)})[G_*^{(i)}(\theta_*^{(i)})' S(\theta_*^{(i)})^{-1} G_*(\bar{\theta}^{(i)})]^{-1} G_*^{(i)}(\theta_*^{(i)})' S(\theta_*^{(i)})^{-1} \tag{1.141}$$

which can be consistently estimated by their sample counterparts, $F_T(\hat{\theta}_T^{(i)}) = W_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}[I - P_T^{(i)}(\hat{\theta}_T^{(i)})]$, where:

$$F_T(\hat{\theta}_T^{(i)}) = W_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}[I - P_T^{(i)}(\hat{\theta}_T^{(i)})] \tag{1.142}$$

and

$$P_T^{(i)}(\hat{\theta}_T^{(i)}) = G_T^{(i)}(\hat{\theta}_T^{(i)})[G_T^{(i)}(\hat{\theta}_T^{(i)})' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} G_T^{(i)}(\bar{\theta}^{(i)})]^{-1} G_T^{(i)}(\hat{\theta}_T^{(i)})' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \tag{1.143}$$

$$= G_T^{(i)}(\hat{\theta}_T^{(i)})[G_T^{(i)}(\hat{\theta}_T^{(i)})' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} G_T^{(i)}(\hat{\theta}_T)]^{-1} G_T^{(i)}(\hat{\theta}_T^{(i)})' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} + o_p(1) \tag{1.144}$$

given that, under our assumptions, $\hat{\theta}_T \xrightarrow{p} \theta_*$.

1.6 Consistency (power) of tests

In this section, we prove that the tests we studied are consistent for the hypothesis covered in this chapter. The test statistics include $\mathscr{K}_T$ and B_T, where $\hat{\theta}_T^{(i)}$ represents the GMM estimator. We show that under the first alternative hypothesis (when the first model is better than the second one), the test statistics converge to $-\infty$. On the other hand, under the second alternative hypothesis (when the second model is better than the first model), we show that the test statistics converge to $+\infty$.

1.6.1 Consistency of the test using $\mathscr{K}_T$

In order to show that the test using the statistic $\mathscr{K}_T$ is consistent, we show that under the alternative hypothesis, H_1, $T^{-1}\mathscr{K}_T \xrightarrow{p} C_1 < 0$, where C is a finite constant. Likewise, we show that when H_2 is true, $T^{-1}\mathscr{K}_T \xrightarrow{p} C_2 > 0$.

Let $\mu_*^{(i)} = \mu_*^{(i)}(\theta_*^{(i)}) = E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}]$, $W_*^{(i)} = W_*^{(i)}(\theta_*^{(i)})$, $P_*^{(i)} = P_*^{(i)}(\theta_*^{(i)})$ and $F_*^{(i)} = F_*^{(i)}(\theta_*^{(i)})$. From Equations (54), (56) and (1.58), it follows that:

$$
\begin{aligned}
T^{-1}\mathscr{K}_T \xrightarrow{p} & \; \mu_*^{(1)'}F_*^{(1)}\mu_*^{(1)} \\
& - \mu_*^{(2)'}F_*^{(2)}\mu_*^{(2)} && (1.145)\\
& = \mu_*^{(1)'}W_*^{(1)-1}\mu_*^{(1)} \\
& - \mu_*^{(2)'}W_*^{(2)-1}\mu_*^{(2)} \\
& + \mu_*^{(1)'}W_*^{(1)-1}P_*^{(1)}\mu_*^{(1)} \\
& - \mu_*^{(2)'}W_*^{(2)-1}P_*^{(2)}\mu_*^{(2)} && (1.146)\\
& = -\mu_*^{(2)'}W_*^{(2)-1}\mu_*^{(2)} \\
& - \mu_*^{(2)'}W_*^{(2)-1}P_*^{(2)}\mu_*^{(2)} && (1.147)\\
& < 0 && (1.148)
\end{aligned}
$$

In this case, by assumption, the first model is correctly specified and $\mu_*^{(1)'}W_*^{(1)-1}\mu_*^{(1)} = \mu_*^{(1)'}W_*^{(1)-1}P_*^{(1)}\mu_*^{(1)} = 0$. Further, both $W_*^{(2)-1}$ and $P_*^{(2)-1}$ are positive definite[23]. It follows that $\mu_*^{(2)'}W_*^{(2)-1}\mu_*^{(2)} > 0$ and $\mu_*^{(2)'}W_*^{(2)-1}P_*^{(2)}\mu_*^{(2)}0 >$ and the result follows by (1.147).

1.6.2 Consistency of the test using B_T

First, scaling B_T in equation (1.91) by $T^{-1/2}$, it follows that:

$$
plim(T^{-1/2}B_T) = \frac{plim(NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)}))}{plim(\hat{\rho}_T)} \qquad (1.149)
$$

For the numerator, under H_1, (1.67) no longer follows from (1.66). However, if we let $C = N^{(1)}R_*(\theta_*^{(1)}) - N^{(2)}R_*(\theta_*^{(2)}) < 0$, under H_1, we can write the following analog to

---

[23] Under Assumption 8-(iv), $rank(P_*^{(i)}) = rank(S_*^{(i)-1}) = k_i$, hence $P_*^{(i)}$ is full rank, therefore, positive definite.

Equation (77):

$$\begin{aligned}
NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)}) = {} & \{NR_T^{(1)}(\theta_*^{(1)}) - NR_*^{(1)}(\theta_*^{(1)})\} \\
& - \{NR_T^{(2)}(\theta_*^{(2)}) - NR_*^{(2)}(\theta_*^{(2)})\} \\
& + \left.\frac{\partial NR_T^{(1)}}{\partial \theta^{(1)\prime}}\right|_{\theta^{(1)}=\bar{\theta}^{(1)}} (\hat{\theta}_T^{(1)} - \theta_*^{(1)}) \\
& - \left.\frac{\partial NR_T^{(2)}}{\partial \theta^{(2)\prime}}\right|_{\theta^{(2)}=\bar{\theta}^{(2)}} (\hat{\theta}_T^{(2)} - \theta_*^{(2)}) \\
& + C
\end{aligned} \tag{1.150}$$

where C is a positive constant. As for the denominator, $\{NR_T^{(i)}(\theta_*^{(i)}) - NR_*^{(i)}(\theta_*^{(i)})\}$ converges in probability to constants, using the same analysis as in Equation (1.77). Under our assumptions, $\partial NR_T^{(i)}/\partial\theta^{(i)\prime}|_{\theta^{(i)}=\bar{\theta}^{(i)}} = O_p(1)$, $\hat{\theta}_T^{(i)} - \theta_*^{(i)} = O_p(1)$ and $plim(NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)}))$ is a nonzero constant, under H_1.

Besides, under our assumptions, $plim(\hat{\rho}_T) \xrightarrow{p} [L_*^{\prime} V_* L_*]^{1/2} = O(1)$, $L_*^{\prime} V_* L_*^{-1/2} = O(1)$ and therefore, $plim(T^{-1/2}B_T) = C \neq 0$. Hence, under H_1, $plim(B_T) = -\infty$ and under H_2, $plim(B_T) = +\infty$.

1.7 Conclusion

In this chapter, it is seen that there is no unified model selection criteria when the models under consideration are possibly misspecified. Unless there is some underlying theoretical selection, it is imperative that the researcher first decide if they are dealing with correctly specified or misspecified models. For that purpose, it is recommended that econometricians follow a two stage strategy. In the first stage, the $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$ test statistic is used to determine if the models under consideration are individually correctly specified. Once the model validation is finished and models under consideration are found to be correctly specified, then $\mathscr{K}_T$, whose asymptotic distribution is given by Theorem 2, can be used for model selection. On the other hand, if in the first stage, both of the models under consideration are found to be misspecified, then B_T, whose asymptotic distribution is given by Theorem 3, can be used for model selection in the second stage.

The Rivers and Vuong statistic does not have an asymptotic $N(0,1)$ distribution for the case when the models under consideration are correctly specified for both the GMM estimator and CUE. Therefore, for the case of correctly specified models, there is no need to use the B_T, hence $T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}$[24].

---

[24] Given that B_T does not have a standard normal asymptotic distribution in the case of correctly specified models, it is redundant to compute the numerator of B_T.

Chapter 2

CUE Estimation of RE Models

In this chapter, we study the distribution of the Noise Ratio Statistic when the parameters of the RE model under consideration are estimated using the CUE (continuous updating estimator) of Hansen, Heaton, and Yaron (1996). We will keep the same notation, assumptions and definitions that we have outlined in Chapter 1 and make changes only when necessary.

2.1 The continuous updating estimator

The methodology behind the CUE is developed in Hansen, Heaton, and Yaron (1996). As in the previous chapter, we use the regression of $g_t^{(i)}(\hat{\theta}_T^{(i)})$ on to $z_t^{'}$, where $g_t^{(i)}(\hat{\theta}_T^{(i)})$ denotes an estimated Euler equation forecast error,[1] derived from the underlying economic model, and $z_t^{'}$ denotes the t^{th} row of the $T \times k$ instrumental variables matrix Z. Further, $"'"$ denotes matrix or vector transpose. The CUE estimator is defined as follows:[2],

$$\hat{\theta}_{CUE,T}^{(i)} = argmin_{\theta^{(i)} \in \Theta^{(i)}} Q_{CUE,T}^{(i)}(\theta^{(i)}) \tag{2.1}$$

$$Q_{CUE,T}^{(i)}(\theta^{(i)}) = m_T^{(i)}(\theta^{(i)}) S_T^{(i)}(\theta^{(i)})^{-1} m_T^{(i)}(\theta^{(i)}) \tag{2.2}$$

---

[1] The dependence of the Euler equation forecast error on $\hat{\theta}_T^{(i)}$ will play an important part in deriving the asymptotic properties of the NR_T, hence we will write $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ to denote our noise ratio statistic.

[2] See Equation (4) of Hansen, Heaton, and Yaron (1996).

where

$$m_T^{(i)}(\hat{\theta}_T^{(i)}) = T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)}, v_t) z_t \tag{2.3}$$

and for the RE models estimated by IV, we have the following weighting matrix:[3]

$$S_T^{(i)}(\hat{\theta}_T^{(i)}) = \left(T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}\right)\left(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})^2\right) \tag{2.4}$$

Given the optimization problem in (2.2), we have the following f.o.c.:

$$\begin{aligned}
&2m_T^{(i)}(\hat{\theta}_T^{(i)})' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} G_T^{(i)}(\hat{\theta}_T^{(i)}) \\
&- \left(m_T^{(i)}(\hat{\theta}_T^{(i)})' \otimes m_T^{(i)}(\hat{\theta}_T^{(i)})'\right)\left(S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\right)\left(\left.\frac{\partial vec(S_T)}{\partial \theta^{(i)\prime}}\right|_{\theta=\hat{\theta}_T^{(i)}}\right) && (2.5)\\
&= 2m_T^{(i)}(\hat{\theta}_T^{(i)})' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} G_T^{(i)}(\hat{\theta}_T^{(i)}) \\
&- vec\left(m_T^{(i)}(\hat{\theta}_T^{(i)}) m_T^{(i)}(\hat{\theta}_T^{(i)})'\right)'\left(S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\right)\left(\left.\frac{\partial vec(S_T)}{\partial \theta^{(i)\prime}}\right|_{\theta=\hat{\theta}_T}\right) = 0 && (2.6)
\end{aligned}$$

where the $k_i \times p_i$ derivative matrix $g_t^{(i)}(\theta^{(i)})$ is given by:

$$\begin{aligned}
G_t^{(i)}(\theta^{(i)}) &= \frac{\partial m_T^{(i)}}{\partial \theta^{(i)\prime}} && (2.7)\\
&= T^{-1}\sum_{t=1}^{T} z_t^{(i)}\left(\frac{\partial g_t^{(i)}(\theta^{(i)})}{\partial \theta^{(i)\prime}}\right) && (2.8)\\
& && (2.9)
\end{aligned}$$

Using the fact that, for two conformable matrices, A and B, $vec(A\,B) = (B' \otimes I_q)vec(A)$, and that $vec[m_T^{(i)}(\hat{\theta}_T^{(i)})] = m_T^{(i)}(\hat{\theta}_T^{(i)})$[4]

$$vec[m_T^{(i)}(\hat{\theta}_T^{(i)}) m_T^{(i)\prime}(\hat{\theta}_T^{(i)})] = [m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i}] m_T^{(i)}(\hat{\theta}_T^{(i)}) \tag{2.10}$$

Using (2.10), (2.6) can be written in the following form:

$$\mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)}) m_T^{(i)}(\hat{\theta}_T^{(i)}) = 0 \tag{2.11}$$

---

[3] See Equation (5) of Hansen, Heaton, and Yaron (1996).

[4] Given that $m_T^{(i)}(\hat{\theta}_T^{(i)})$ is a $q_i \times 1$ vector.

where

$$\mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)}) = 2G_T^{(i)}(\hat{\theta}_T^{(i)})'S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} - \left\{\frac{\partial vec[S_T^{(i)}(\hat{\theta}_T^{(i)})]}{\partial\theta'}\right\}'[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}](m_T^{(i)}(\hat{\theta}_T^{(i)})\otimes I_{k_i}) \tag{2.12}$$

In the rest of this section, using the f.o.c. in Equation (2.6) as a starting point, the limit distribution of $T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)})$ for CUE is derived. This will help in the derivation of the asymptotic results in the subsequent sections.

Letting $H_T(\hat{\theta}_T^{(i)}) = \mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)})m_T^{(i)}(\hat{\theta}_T^{(i)})$[5], a mean value expansion $H_T(\hat{\theta}_T^{(i)})$ around $\theta_*^{(i)}$ yields:

$$\begin{aligned} H_T(\hat{\theta}_T^{(i)}) &= H_T(\theta_*^{(i)}) + \left.\frac{\partial H_T}{\partial\theta^{(i)\prime}}\right|_{\theta^{(i)}=\bar{\theta}_T^{(i)}}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \qquad (2.13)\\ &= H_T(\theta_*^{(i)}) + \left\{\left(m_T^{(i)}(\theta_*^{(i)})'\otimes I_p\right)\left(\left.\frac{\partial vec(\mathscr{N}_T^{(i)}(\theta))}{\partial\theta^{(i)\prime}}\right|_{\theta^{(i)}=\bar{\theta}_T^{(i)}}\right) + \mathscr{N}_T^{(i)}(\theta_*^{(i)})'\left(\left.\frac{\partial m_T^{(i)}}{\partial\theta^{(i)\prime}}\right|_{\theta^{(i)}=\bar{\theta}_T^{(i)}}\right)\right\}\\ &\qquad \times(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \qquad (2.14) \end{aligned}$$

where $\bar{\theta}_T = \lambda_{T,i}\theta_* + (1-\lambda_{T,i})\hat{\theta}_T$, λ_T is the $(q\times 1)$ vector with i^{th} element $\lambda_{T,i}$ for some $0 < \lambda_{T,i} < 1$.

Next, $\partial vec(\mathscr{N}(\theta)_T)/\partial\theta^{(i)\prime}|_{\theta^{(i)}=\bar{\theta}_T^{(i)}}$ can be computed as follows:

$$\begin{aligned} \frac{\partial vec(\mathscr{N}_T^{(i)}(\theta^{(i)}))}{\partial\theta^{(i)\prime}} &= [S_T^{(i)}(\theta)^{-1}\otimes I_{k_i}]\left[\frac{\partial vec(g_t^{(i)}(\theta^{(i)})')}{\theta^{(i)\prime}}\right]\\ &\quad - \left[I\otimes g_t^{(i)}(\theta^{(i)})'\right]\left[\frac{\partial vec(S_T^{(i)}(\theta)^{-1})}{\theta^{(i)\prime}}\right]\\ &\quad + \frac{\partial vec(\mathscr{D}_T(\theta))}{\partial\theta^{(i)\prime}} \qquad (2.15) \end{aligned}$$

where

$$\mathscr{D}_T(\theta) = \left\{\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial\theta'}\right\}'[S_T^{(i)}(\theta)^{-1}\otimes S_T^{(i)}(\theta)^{-1}](m_T^{(i)}(\theta^{(i)})\otimes I_{k_i}) \tag{2.16}$$

---

[5] Given $\mathscr{N}_T$ is $p_i\times k_i$ and m_T is $k_i\times 1$, it follows that H_T is $p\times 1$.

Next, we can write the following:[6]

$$
\begin{aligned}
vec(\mathscr{D}_T(\theta)) &= \left(I_{k_i} \otimes \left\{\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial \theta'}\right\}' [S_T^{(i)}(\theta)^{-1} \otimes S_T^{(i)}(\theta)^{-1}]\right) vec(m_T^{(i)}(\theta^{(i)}) \otimes I_{k_i}) && (2.17)\\
&= \left([m_T^{(i)}(\theta^{(i)}) \otimes I_{k_i}] \otimes \left\{\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial \theta'}\right\}'\right) vec(S_T^{(i)}(\theta)^{-1} \otimes S_T^{(i)}(\theta)^{-1}) && (2.18)\\
&= \left((m_T^{(i)}(\theta^{(i)}) \otimes I_{k_i})[S_T^{(i)}(\theta)^{-1} \otimes S_T^{(i)}(\theta)^{-1}]\right) F_{k^2} vec\left(\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial \theta'}\right) && (2.19)
\end{aligned}
$$

where F_{k^2} is a $k^2 \times k^2$ commutation matrix:[7] such that

$$
vec([\partial vec[S_T^{(i)}(\theta)]/\partial \theta^{(i)'}]') = F_{k^2} vec(\partial vec[S_T^{(i)}(\theta)]/\partial \theta') \tag{2.20}
$$

Further, from Lütkepohl (1993, p. 467), it follows that:[8],

$$
\begin{aligned}
vec(m_T^{(i)}(\theta^{(i)}) \otimes I_{k_i}) &= (C_{k^2} \otimes I_{k_i})(m_T^{(i)}(\theta^{(i)}) \otimes vec(I_{k_i})) && (2.21)\\
vec(S_T^{(i)}(\theta)^{-1} \otimes S_T^{(i)}(\theta)^{-1}) &= (I_{k_i} \otimes K_{k^2} \otimes I_{k_i})[vec(S_T^{(i)}(\theta)^{-1}) \otimes vec(S_T^{(i)}(\theta)^{-1})] && (2.22)
\end{aligned}
$$

Equations (2.17)-(2.22) yield the following:

$$
\begin{aligned}
\frac{\partial vec(\mathscr{D}_T)}{\partial \theta^{(i)'}} &= \left\{I_{k_i} \otimes \left[\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial \theta'}\right]' [S_T^{(i)}(\theta)^{-1} \otimes S_T^{(i)}(\theta)^{-1}]\right\} \left[(C_{k^2} \otimes I_{k_i})[g_t^{(i)}(\theta^{(i)}) \otimes I_{k_i}]\right]\\
&+ \left\{\left([m_T^{(i)}(\theta^{(i)}) \otimes I_{k_i}] \otimes \left\{\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial \theta'}\right\}'\right)(I_{k_i} \otimes K_{k^2} \otimes I_{k_i})\right.\\
&\times \left.\left[\left(\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial \theta'}\right) \otimes vec(S_T^{(i)}(\theta)^{-1}) + vec(S_T^{(i)}(\theta)^{-1})\left(\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial \theta^{(i)'}}\right)\right]\right\}\\
&+ \left((m_T^{(i)}(\theta^{(i)}) \otimes I_{k_i})[S_T^{(i)}(\theta)^{-1} \otimes S_T^{(i)}(\theta)^{-1}]\right) F_{k^2} \frac{\partial [vec(\partial vec[S_T^{(i)}(\theta)]/\partial \theta^{(i)'})]}{\partial \theta^{(i)'}}
\end{aligned}
\tag{2.23}
$$

Letting $\mathscr{H}_T(\theta)$ denote $plim(\partial H_T(\theta)/\partial \theta^{(i)'})$, we assume that the following holds:

Assumption 21 *$\mathscr{H}_T(\theta)^{-1}$ exists.*

---

[6] See Corollary 4.1 of Dhrymes (2000)[p.119].

[7] See Lütkepohl (1993)[p.466].

[8] Let G and F be matrices that are $m \times n$ and $p \times q$, respectively. Then, it follows that: $vec(G \otimes F) = (I_n \otimes K_{mn} \otimes I_p)(vec(G)) \otimes (vec(F))$.

From Equations (2.14), (2.15) and (2.23), it follows that:

$$
\begin{aligned}
\frac{\partial H_T(\theta)}{\partial \theta^{(i)\prime}} &= -\Bigg(m_T^{(i)}(\theta_*^{(i)})' \otimes I_p\Bigg)\Bigg(\frac{\partial vec(\mathscr{N}_T)}{\partial \theta^{(i)\prime}}\bigg|_{\theta^{(i)}=\theta_*^{(i)}}\Bigg) \\
&\quad + \mathscr{N}_T^{(i)}(\theta_*^{(i)})' G_T^{(i)}(\theta_*^{(i)})
\end{aligned}
\tag{2.24}
$$

$$
\begin{aligned}
&= \Bigg(m_T^{(i)}(\theta_*^{(i)})' \otimes I_p\Bigg)\Bigg| S_T^{(i)}(\theta_T)^{-1} \otimes I_{k_i}\big]\Bigg[\frac{\partial vec(G_T^{(i)}(\theta_*^{(i)})')}{\theta^{(i)\prime}}\Bigg] \\
&- \big[I \otimes G_T^{(i)}(\theta_*^{(i)})'\big]\Bigg[\frac{\partial vec(S_T^{(i)}(\theta_*^{(i)})^{-1})}{\theta^{(i)\prime}}\Bigg] \\
&+ \Bigg[I_{k_i} \otimes \Bigg\{\frac{\partial vec[S_T^{(i)}(\theta_*^{(i)})]}{\partial \theta'}\Bigg\}' [S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_T^{(i)}(\theta_*^{(i)})^{-1}]\Bigg]\Bigg[(C_{k^2} \otimes I_{k_i})[G_T \otimes I_{k_i}]\Bigg] \\
&+ \Bigg[\Bigg([m_T^{(i)}(\theta_*^{(i)}) \otimes I_{k_i}] \otimes \Bigg\{\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial \theta'}\Bigg\}'\Bigg)(I_{k_i} \otimes K_{k^2} \otimes I_{k_i}) \\
&\times \Bigg\langle \Bigg(\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial \theta'}\Bigg) \otimes vec(S_T^{(i)}(\theta_*^{(i)})^{-1}) + vec(S_T^{(i)}(\theta_*^{(i)})^{-1})\Bigg(\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial \theta^{(i)\prime}}\Bigg)\Bigg\rangle\Bigg] \\
&+ \Bigg(\big(m_T^{(i)}(\theta_*^{(i)}) \otimes I_{k_i}\big)[S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_T^{(i)}(\theta_*^{(i)})^{-1}]\Bigg) F_{k^2} \frac{\partial[vec(\partial vec[S_T^{(i)}(\theta)]/\partial \theta^{(i)\prime})]}{\partial \theta^{(i)\prime}}\Bigg| \\
&+ \mathscr{N}_T^{(i)}(\theta_*^{(i)})' G_T^{(i)}(\theta_*^{(i)})
\end{aligned}
\tag{2.25}
$$

Equations (2.14), (2.25) and Assumption 21 yield:

$$
T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) = \Bigg[\frac{\partial H_T(\theta)}{\partial \theta^{(i)\prime}}\bigg|_{\theta^{(i)}=\bar{\theta}_T^{(i)}}\Bigg]^{-1} T^{1/2}\big[H_T(\hat{\theta}_T^{(i)}) - H_T(\theta_*^{(i)})\big] \tag{2.26}
$$

$$
= -\Bigg[\frac{\partial H_T(\theta)}{\partial \theta^{(i)\prime}}\bigg|_{\theta^{(i)}=\bar{\theta}_T^{(i)}}\Bigg]^{-1} T^{1/2} H_T(\theta_*^{(i)}) \tag{2.27}
$$

$$
= -\Bigg[\frac{\partial H_T(\theta)}{\partial \theta^{(i)\prime}}\bigg|_{\theta^{(i)}=\theta_*^{(i)}}\Bigg]^{-1} T^{1/2} H_T(\theta_*^{(i)}) + o_p(1) \tag{2.28}
$$

$$
\begin{aligned}
&= -\Bigg\{\Bigg(m_T^{(i)}(\theta_*^{(i)})' \otimes I_p\Bigg)\Bigg(\frac{\partial vec(\mathscr{N}_T)}{\partial \theta^{(i)\prime}}\bigg|_{\theta^{(i)}=\theta_*^{(i)}}\Bigg) \\
&\quad + \mathscr{N}_T^{(i)}(\theta_*^{(i)})'\Bigg(\frac{\partial m_T^{(i)}(\theta^{(i)})}{\partial \theta^{(i)\prime}}\bigg|_{\theta^{(i)}=\theta_*^{(i)}}\Bigg)\Bigg\}^{-1} \mathscr{N}_T^{(i)}(\theta_*^{(i)}) T^{1/2} m_T^{(i)}(\theta_*^{(i)})
\end{aligned}
\tag{2.29}
$$

$$\begin{aligned}
&= -\Bigg\{\Big(m_T^{(i)}(\theta_*^{(i)})' \otimes I_p\Big)\Bigg(\frac{\partial vec(\mathscr{N}_T)}{\partial\theta^{(i)\prime}}\bigg|_{\theta^{(i)}=\theta_*^{(i)}}\Bigg) \\
&\quad + \mathscr{N}_T^{(i)}(\theta_*^{(i)})' G_T^{(i)}(\theta_*^{(i)})\Bigg\}^{-1} \mathscr{N}_T^{(i)}(\theta_*^{(i)}) T^{1/2} m_T^{(i)}(\theta_*^{(i)}) + o_p(1)
\end{aligned} \tag{2.30}$$

where Equation (2.27) follows from Equation (2.11).

Equations (2.15), (2.23) and (2.30) together imply the following:

$$\begin{aligned}
T^{1/2}(\hat{\theta}_T^{(i)} - \theta^{(i)}) = {}&\Bigg\{\Big(m_T^{(i)}(\theta_*^{(i)})' \otimes I_p\Big)\Bigg| S_T^{(i)}(\theta_T)^{-1} \otimes I_{k_i}\Big]\Bigg[\frac{\partial vec(G_T^{(i)}(\theta_*^{(i)})')}{\theta^{(i)\prime}}\Bigg] \\
&- \big[I \otimes G_T^{(i)}(\theta_*^{(i)})'\big]\Bigg[\frac{\partial vec(S_T^{(i)}(\theta_*^{(i)})^{-1})}{\theta^{(i)\prime}}\Bigg] \\
&+ \Bigg[I_{k_i} \otimes \Bigg\{\frac{\partial vec[S_T^{(i)}(\theta_*^{(i)})]}{\partial\theta'}\Bigg\}' [S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_T^{(i)}(\theta_*^{(i)})^{-1}]\Bigg]\Bigg[(C_{k^2} \otimes I_{k_i})[G_T \otimes I_{k_i}]\Bigg] \\
&+ \Bigg[\Bigg([m_T^{(i)}(\theta_*^{(i)}) \otimes I_{k_i}] \otimes \Bigg\{\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial\theta'}\Bigg\}'\Bigg)(I_{k_i} \otimes K_{k^2} \otimes I_{k_i}) \\
&\times \Bigg\langle\Bigg(\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial\theta'}\Bigg) \otimes vec(S_T^{(i)}(\theta_*^{(i)})^{-1}) + vec(S_T^{(i)}(\theta_*^{(i)})^{-1})\Bigg(\frac{\partial vec[S_T^{(i)}(\theta)]}{\partial\theta^{(i)\prime}}\Bigg)\Bigg\rangle\Bigg] \\
&+ \Bigg(\big(m_T^{(i)}(\theta_*^{(i)}) \otimes I_{k_i}\big)[S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_T^{(i)}(\theta_*^{(i)})^{-1}]\Bigg) F_{k^2} \frac{\partial[vec(\partial vec[S_T^{(i)}(\theta)]/\partial\theta^{(i)\prime})]}{\partial\theta^{(i)\prime}}\Bigg| \\
&+ \mathscr{N}_T^{(i)}(\theta_*^{(i)})' G_T^{(i)}(\theta_*^{(i)})\Bigg\}^{-1} \mathscr{N}_T^{(i)}(\theta_*^{(i)}) T^{1/2} m_T^{(i)}(\theta_*^{(i)}) + o_p(1)
\end{aligned} \tag{2.31}$$

In the next sections, we will use Equation (2.31) to analyze the asymptotic behavior of:

(i) $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$,

(ii) $T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}$ and,

(iii) B_T.

The data generating processes covered include:

(i) $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{git}^2$, where $\nexists\, \sigma_{gi}^2 : \sigma_{git}^2 = \sigma_{gi}^2$ for all t,

(ii) $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is a dependent process.

In the case of correctly specified models, we study the distributions of $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$ and $T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}$ with data generating process (ii). On the other hand, when the models under consideration are misspecified in the non-local sense, we study the distribution of B_T with the data generating process given in (iii).

2.2 Correct specification

In this section, we study the distributions of $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$ and $B_T(\hat{\theta}_T^{(i)})$, where $\hat{\theta}_T^{(i)}$ represents CUE, for correctly specified models[9].

2.2.1 The limiting distribution of $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$: CUE

In this section, we analyze the probability limit of TNR_T when Assumptions 1-12 hold. We begin studying the asymptotic distribution of $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$ by a mean value expansion around $\theta_*^{(i)}$.

$$NR_T^{(i)}(\hat{\theta}_T^{(i)}) = NR_T^{(i)}(\theta_*^{(i)}) + \left.\frac{\partial NR_T^{(i)}}{\partial \theta^{(i)\prime}}\right|_{\theta^{(i)}=\bar{\theta}^{(i)}} (\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.32}$$

where $\bar{\theta} = \lambda_T \theta_* + (1-\lambda_T)\hat{\theta}_T$ for some $0 < \lambda_T < 1$ and $\hat{\theta}_T^{(i)}$ denotes the second-step GMM estimator for the i^{th} model.

Multiplying through by T yields:

$$TNR_T^{(i)}(\hat{\theta}_T^{(i)}) = TNR_T^{(i)}(\theta_*^{(i)}) + T^{1/2} \left.\frac{\partial NR_T^{(i)}}{\partial \theta^{(i)\prime}}\right|_{\theta^{(i)}=\bar{\theta}^{(i)}} T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.33}$$

$$\begin{aligned}
\frac{\partial NR_T^{(i)}}{\partial \theta^{(i)\prime}} &= \left(\frac{\partial NR_T^{(i)}}{\partial m_T^{(i)\prime}}\right)' \left(\frac{\partial m_T^{(i)}}{\partial \theta^{(i)\prime}}\right) \\
&\quad + \left(\frac{\partial NR_T^{(i)}}{\partial vec(W_T^{(i)-1})}\right) \left(\frac{\partial vec(W_T^{(i)-1})}{\partial vec(W_T^{(i)})'}\right) \left(\frac{\partial vec(W_T^{(i)})}{\partial \theta^{(i)\prime}}\right) && (2.34) \\
&= 2 m_T^{(i)}(\theta^{(i)})' W_T^{(i)}(\theta^{(i)})^{-1} g_t^{(i)}(\theta^{(i)}) \\
&\quad + \left(m_T^{(i)}(\theta^{(i)})' \otimes m_T^{(i)}(\theta^{(i)})'\right) \left(\frac{\partial vec(W_T^{(i)-1})}{\partial vec(W_T^{(i)})'}\right) \left(\frac{\partial vec(W_T^{(i)})}{\partial \theta^{(i)\prime}}\right) && (2.35)
\end{aligned}$$

---

[9] See Definition 2 in Chapter 1.

$$= 2m_T^{(i)}(\theta^{(i)})' W_T^{(i)}(\theta^{(i)})^{-1} g_t^{(i)}(\theta^{(i)})$$

$$- \left(m_T^{(i)}(\theta^{(i)})' \otimes m_T^{(i)}(\theta^{(i)})'\right) \left(W_T^{(i)}(\theta^{(i)})^{-1} \otimes W_T^{(i)}(\theta^{(i)})^{-1}\right) \left(\frac{\partial vec(W_T^{(i)})}{\partial \theta^{(i)\prime}}\right) \tag{2.36}$$

where

$$g_t^{(i)}(\theta^{(i)}) = \frac{\partial m_T^{(i)}}{\partial \theta^{(i)\prime}} \tag{2.37}$$

$$= T^{-1} \sum_{t=1}^{T} z_t^{(i)} \left(\frac{\partial g_t^{(i)}(\theta^{(i)})}{\partial \theta^{(i)\prime}}\right) \tag{2.38}$$

where $g_t^{(i)}(\theta^{(i)})$ is $k \times p$.

2.2.2 RE models estimated by IV: CUE

In the case of RE models estimated by IV, we have $W_T \equiv S_T$,[10] where:

$$S_T^{(i)}(\hat{\theta}_T^{(i)}) = W_T^{(i)}(\hat{\theta}_T^{(i)}) \tag{2.39}$$

$$= \left(T^{-1} \sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}\right) \left(T^{-1} \sum_{t=1}^{T} g_T^{(i)}(\hat{\theta}_T^{(i)})^2\right) \tag{2.40}$$

We note that $S_T^{(i)}(\hat{\theta}_T^{(i)})$ has the same form as (1.18). However, in the current case, $\hat{\theta}_T^{(i)}$ represents the CUE, not the GMM estimator.

It follows that, in this case, $NR_T^{(i)}(\theta^{(i)}) \equiv Q_{CUE,T}(\theta)$. However, in order to allow for conditional heteroschedasticity, given by Assumption 11, we use the following weighting matrix to compute the CUE:

$$S_T^{(i)}(\hat{\theta}_T^{(i)}) = T^{-1} \sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})^2 z_t^{(i)} z_t^{(i)\prime} \tag{2.41}$$

---

[10] See Equation (5) of Hansen, Heaton, and Yaron (1996).

Under the current case, $\partial NR_T^{(i)}/\partial\theta^{(i)\prime} \equiv \partial Q_{CUE,T}/\partial\theta^{(i)\prime}$ is given by:

$$T^{1/2}\frac{\partial NR_T^{(i)}}{\partial\theta^{(i)\prime}} = 2T^{1/2}m_T^{(i)}(\theta^{(i)})'S_T^{(i)}(\theta^{(i)})^{-1}g_t^{(i)}(\theta^{(i)}) - \left(T^{1/2}m_T^{(i)}(\theta^{(i)})' \otimes m_T^{(i)}(\theta^{(i)})'\right)\left(S_T^{(i)}(\theta)^{-1} \otimes S_T^{(i)}(\theta)^{-1}\right)\left(\frac{\partial vec(S_T^{(i)}(\theta^{(i)}))}{\partial\theta^{(i)\prime}}\right) \tag{2.42}$$

where $\partial NR_T^{(i)}/\partial\theta^{(i)\prime}$ is in the form given in Equation (1.34).

Given Equation (2.39), we can use the same analysis in Equations (1.34)-(1.36) of Chapter 1 to write the following:

$$\left.\frac{\partial S_T^{(i)}(\theta)}{\partial\theta^{(i)\prime}}\right|_{\theta^{(i)}=\theta_*^{(i)}} = O_p(1) \tag{2.43}$$

where $\theta_*^{(i)}$ is such that $plim(\hat{\theta}_T^{(i)}) = \theta_*^{(i)}$ and $\hat{\theta}_T^{(i)}$ represents the CUE.

The following assumption will be used for the next theorem:

Assumption 22 *We assume that* $T^{1/2}m_T^{(i)}(\theta_*^{(i)}) \xrightarrow{d} N(0, S^{(i)}(\theta_*^{(i)}))$.

Further, Assumption 22 implies $S^{-1/2}T^{1/2}m_T^{(i)}(\theta_*^{(i)}) \sim N(0, I_{k_i})$, where $\eta_k^{(i)} \sim N(0, I_{k_i})$ and it also follows that $T^{1/2}m_T^{(i)}(\theta_*^{(i)}) = O_p(1)$.

In the case of correctly specified models, $E[m_T^{(i)}(\theta_*^{(i)})] = 0$, and Equations (2.42) and (2.43) together imply:

$$\begin{aligned}
T^{1/2}\left.\frac{\partial NR_T^{(i)}}{\partial\theta^{(i)\prime}}\right|_{\theta^{(i)}=\theta_*^{(i)}} &= 2T^{1/2}m_T^{(i)}(\theta_*^{(i)})'S_T^{(i)}(\theta_*^{(i)})^{-1}G_T^{(i)}(\theta_*^{(i)}) \\
&\quad - \left(T^{1/2}m_T^{(i)}(\theta_*^{(i)})' \otimes m_T^{(i)}(\theta_*^{(i)})'\right) && (2.44) \\
&\quad \times \left(S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_T^{(i)}(\theta_*^{(i)})^{-1}\right)\left(\left.\frac{\partial vec(S_T^{(i)}(\theta^{(i)}))}{\partial\theta^{(i)\prime}}\right|_{\theta^{(i)}=\theta_*^{(i)}}\right) && (2.45) \\
&= 2T^{1/2}m_T^{(i)}(\theta_*^{(i)})'S_T^{(i)}(\theta_*^{(i)})^{-1}G_T^{(i)}(\theta_*^{(i)}) - O_p(1)o_p(1)O_p(1)O_p(1) && (2.46) \\
&= o_p(1) && (2.47)
\end{aligned}$$

Next, we study the probability limit of $T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)})$. First of all, in the case of correctly specified models, $E[m_T^{(i)}(\theta_*^{(i)})] = 0$, and we can use Equation (2.12) to write the following:

$$\mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)}) = 2G_T^{(i)}(\hat{\theta}_T^{(i)})' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} + o_p(1) \tag{2.48}$$

We also make the following assumption:

Assumption 23

$$\frac{\partial[vec(\partial vec[S_T^{(i)}(\theta)]/\partial\theta^{(i)\prime})]}{\partial\theta^{(i)\prime}} = O_p(1) \tag{2.49}$$

Under our assumptions, Equations (2.15) and (2.23) together imply:

$$\frac{\partial vec(\mathscr{N}_T^{(i)}(\theta))}{\partial\theta^{(i)\prime}} = O_p(1) \tag{2.50}$$

Further, in the case of correctly specified models, using Equation (2.50), Equation (2.48) can be used to re-write Equation (2.31) as follows:

$$\begin{aligned}
T^{1/2}(\hat{\theta}_T^{(i)} - \theta^{(i)}) &= \left\{\mathscr{N}_T^{(i)}(\theta_*^{(i)})' G_T^{(i)}(\theta_*^{(i)})\right\}^{-1} \mathscr{N}_T^{(i)}(\theta_*^{(i)}) T^{1/2} m_T^{(i)}(\theta_*^{(i)}) + o_p(1) && (2.51)\\
&= \left\{G_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} G_T^{(i)}(\theta_*^{(i)})\right\}^{-1} G_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} T^{1/2} m_T^{(i)}(\theta_*^{(i)})\\
&+ o_p(1) && (2.52)
\end{aligned}$$

Equations (2.42) and (2.52) can be used to re-write Equation (2.33) as follows:

$$\begin{aligned}
TNR_T^{(i)}(\hat{\theta}_T^{(i)}) &= TNR_T^{(i)}(\theta_*^{(i)}) + T^{1/2} \left.\frac{\partial NR_T^{(i)}}{\partial\theta^{(i)\prime}}\right|_{\theta^{(i)}=\bar{\theta}^{(i)}} T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) && (2.53)\\
&= T^{1/2} m_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta)^{-1} T^{1/2} m_T^{(i)}(\theta_*^{(i)}) + T^{1/2} m_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} G_T^{(i)}(\theta_*^{(i)})\\
& && (2.54)\\
&\times \left\{G_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} G_T^{(i)}(\theta_*^{(i)})\right\}^{-1} G_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} T^{1/2} m_T^{(i)}(\theta_*^{(i)})\\
&+ o_p(1) && (2.55)\\
&= T^{1/2} m_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta)^{-1/2} [I_{k_i} - \mathscr{P}_T^{(i)}] S_T^{(i)}(\theta)^{-1/2} T^{1/2} m_T^{(i)}(\theta_*^{(i)}) + o_p(1)\\
& && (2.56)
\end{aligned}$$

and

$$\mathscr{P}_T^{(i)} = \Big(I - S_T^{(i)}(\theta_*^{(i)})^{-1/2} G_T^{(i)}(\theta_*^{(i)}) \{G_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} G_T^{(i)}(\theta_*^{(i)})\}^{-1} G_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1/2}\Big) \tag{2.57}$$

Given Assumption 22, Equation (2.57) can be rewritten as follows:

$$TNR_T^{(i)}(\hat{\theta}_T^{(i)}) = T^{1/2} m_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1/2} [I - \mathscr{P}_T^{(i)}] S_T^{(i)}(\theta_*^{(i)})^{-1/2} T^{1/2} m_T^{(i)}(\theta_*^{(i)}) + o_p(1) \tag{2.58}$$

$$= \eta_k^{(i)\prime} [I - \mathscr{P}_*^{(i)}] \eta_k^{(i)} + o_p(1) \tag{2.59}$$

$$\xrightarrow{p} \eta_k^{(i)\prime} \bar{C}' \bar{\Lambda} \bar{C} \eta_k^{(i)} \tag{2.60}$$

where Equation (2.59) follows from (2.58), under Assumption 22. Further, $plim(\mathscr{P}_T^{(i)}) = \mathscr{P}_*^{(i)}$, $\bar{C}$ is an orthogonal matrix of eigenvectors of $[I - \mathscr{P}_*^{(i)}]$ and $\bar{\Lambda} = diag(\bar{\lambda}_1, \bar{\lambda}_2, ..., \bar{\lambda}_k)$, with corresponding eigenvalues $\bar{\lambda}_1, \bar{\lambda}_2, ..., \bar{\lambda}_k$. Finally, given Equation (2.60), we can state the following theorem:

Theorem 5 *Under Assumptions 1-12 and 22, it follows that:*

$$TNR_T^{(i)}(\hat{\theta}_T^{(i)}) \xrightarrow{d} \sum_{j=1}^{k} \bar{\lambda}_j n_{kj}^2 \tag{2.61}$$

2.2.3 Distribution of $T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}$ under correct specification: CUE

In this section, we study the asymptotic distribution of $T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}$, under the case when:

(i) $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{git}^2$,

(ii) The model under consideration is correctly specified, that is, $E[g_t^{(i)}(\theta_*^{(i)}, v_t) z_t^{(i)}] = 0$,

(iii) The parameters of the model are estimated using the CUE, i.e. $\hat{\theta}_T^{(i)}$ now represents the CUE rather than the GMM estimator that we studied in Chapter 1.

In the case of correctly specified models, we know from the previous section that when $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{git}^2$, $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$ is distributed as a mixture of χ^2 random variables. Therefore, we will use T as the correct rate of convergence to derive the asymptotic distribution of $T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}$. We begin our analysis by imposing the following assumption:

Assumption 24 *$T^{1/2}m_T^{(i)}(\theta_*^{(i)}) \xrightarrow{d} N(0, S(\theta_*^{(i)}))$ and*

(i) $S(\theta_*^{(i)})$ is a positive definite matrix of constants,

(ii) $rank(G_*^{(i)}(\theta_*^{(i)})) = p_i$,

(iii) $\hat{S}(\hat{\theta}^{(i)}) \xrightarrow{p} S^{(i)}(\theta_*^{(i)})$,

(iv) $\hat{\Pi}_T \xrightarrow{p} \Pi_*$.

where

$$m_T^{(i)}(\theta_*^{(i)}) = \begin{bmatrix} m_T^{(1)}(\theta_*^{(1)}) \\ m_T^{(2)}(\theta_*^{(2)}) \end{bmatrix} \tag{2.62}$$

and

$$S(\theta_*^{(i)}) = \begin{bmatrix} S^{(1)}(\theta_*^{(1)}) & S^{(1,2)}(\theta_*^{(i)}) \\ S^{(1,2)}(\theta_*^{(i)}) & S^{(2)}(\theta_*^{(2)}) \end{bmatrix} \tag{2.63}$$

First, we write Equation (2.56) for model i as follows:

$$TNR_T^{(i)}(\hat{\theta}_T^{(i)}) = T^{1/2}m_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1/2}[I_{k_i} - \mathscr{P}_T^{(i)}]S_T^{(i)}(\theta_*^{(i)})^{-1/2}T^{1/2}m_T^{(i)}(\theta_*^{(i)}) + o_p(1) \tag{2.64}$$

Subtracting (2.64) for $i = 1, 2$, yields the following:

$$\begin{aligned} T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\} &= T^{1/2}m_T(\theta_*^{(1)})S_T^{(1)}(\theta_*^{(1)})^{-1/2}[I_{k_i} - \mathscr{P}_T^{(1)}]S_T^{(1)}(\theta_*^{(1)})^{-1/2}T^{1/2}m_T(\theta_*^{(1)}) \\ &\quad - T^{1/2}m_T(\theta_*^{(2)})S_T^{(2)}(\theta_*^{(2)})^{-1/2}[I_{k_i} - \mathscr{P}_T^{(2)}]S_T^{(2)}(\theta_*^{(2)})^{-1/2}T^{1/2}m_T(\theta_*^{(2)}) \\ &\quad + o_p(1) && (2.65) \\ &= T^{1/2}m_T^{(i)}(\theta_*^{(i)})'\mathscr{U}_T(\theta_*^{(i)})T^{1/2}m_T^{(i)}(\theta_*^{(i)}) + o_p(1) && (2.66) \end{aligned}$$

where $\theta_* = [\theta_*^{(1)'}, \theta_*^{(2)'}]'$ and $m_T^{(i)}(\theta_*^{(i)}) = [m_T^{(1)}(\theta_*^{(1)})', m_T^{(2)}(\theta_*^{(2)})']'$ is $(k_1 + k_2) \times 1$ and the $(k_1 + k_2) \times (k_1 + k_2)$ matrix $\mathscr{U}_T(\theta_*^{(i)})$ is given as follows:

$$\mathscr{U}_T(\theta_*^{(i)}) = \begin{bmatrix} S_T^{(1)}(\theta_*^{(1)})^{-1/2}\mathscr{F}_T^{(1)}S_T^{(1)}(\theta_*^{(1)})^{-1/2} & 0 \\ 0 & -S_T^{(2)}(\theta_*^{(2)})^{-1/2}\mathscr{F}_T^{(2)}S_T^{(2)}(\theta_*^{(2)})^{-1/2} \end{bmatrix} \tag{2.67}$$

where $\mathscr{F}_T^{(i)} = [I_{k_i} - \mathscr{P}_T^{(i)}]$ and $\mathscr{P}_T^{(i)}(\theta_*^{(i)})$ is given by Equation (2.57).
Let $\eta_{k_1+k_2} \sim N(0, I_{k_1+k_2})$. Then, under our assumptions, it follows that:

$$T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\} \xrightarrow{d} \eta_{k_1+k_2}' S(\theta_*^{(i)})^{1/2}\mathscr{U}_*(\theta_*^{(i)})S(\theta_*^{(i)})^{1/2}\eta_{k_1+k_2} \tag{2.68}$$

where $plim(\mathscr{U}_T(\theta_*^{(i)})) = \mathscr{U}_*(\theta_*^{(i)})$.
Given Equation (2.68), the asymptotic distribution of $T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}$, under the current case, is given by the following theorem:

Theorem 6 *Under Assumptions 1-14 and 24 we have:*

$$T\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\} \xrightarrow{d} \eta_{k_1+k_2}' S(\theta_*^{(i)})^{1/2}\mathscr{U}_* S(\theta_*^{(i)})^{1/2}\eta_{k_1+k_2} \tag{2.69}$$

2.3 Distribution of B_T under misspecification: CUE

In this section, we derive the asymptotic distribution of the Rivers and Vuong's test statistic, B_T, given by Equation (1.20), when[11]

(i) The models under consideration have non-local misspecification[12],

(ii) $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is a dependent process,

(iii) The parameters of the RE models are estimated using the CUE, that is, $\hat{\theta}_T^{(i)}$ now represents the CUE rather than the GMM estimator that we studied in Chapter 1.[13]

For that purpose, we first define what we mean by a misspecified model[14].

---

[11] When the models under consideration have non-local misspecification, $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ can not be an m.d.s.

[12] See Definition 4.

[13] The minimand of IV-CUE is given by Equation (5) of Hansen, Heaton, and Yaron (1996).

[14] This form of misspecification, as defined in Hall and Inoue (2003), is called non-local misspecification.

Definition 4 *The model, $M_{(i)}$ is said to be misspecified if,*

$$\begin{aligned} E[g^{(i)}(v_t,\theta^{(i)})z_t] &= \mu^{(i)}(\theta^{(i)}) && (2.70)\\ &\neq 0 \qquad \forall\, \theta^{(i)} \in \Theta^{(i)} \textit{ for } i=1,2. && (2.71)\end{aligned}$$

We begin our analysis by multiplying Equation (2.32) through by $T^{1/2}$ as follows: [15],

$$T^{1/2}NR_T^{(i)}(\hat{\theta}_T^{(i)}) = T^{1/2}NR_T^{(i)}(\theta_*^{(i)}) + \left.\frac{\partial NR_T^{(i)}}{\partial\theta^{(i)\prime}}\right|_{\theta^{(i)}=\bar{\theta}^{(i)}} T^{1/2}(\hat{\theta}_T^{(i)}-\theta_*^{(i)}) \qquad (2.72)$$

Subtracting Equation (2.72) for models $i=1,2$, and recognizing that under H_0, we have $NR_*^{(1)}(\theta_*^{(1)}) = NR_*^{(2)}(\theta_*^{(2)})$. Hence, we can write the following:

$$\begin{aligned} T^{1/2}\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\} &= T^{1/2}\{NR_T^{(1)}(\theta_*^{(1)}) - NR_T^{(2)}(\theta_*^{(2)})\}\\ &+ \left.\frac{\partial NR_T^{(1)}}{\partial\theta^{(1)\prime}}\right|_{\theta^{(1)}=\bar{\theta}^{(1)}} T^{1/2}(\hat{\theta}_T^{(1)}-\theta_*^{(1)})\\ &- \left.\frac{\partial NR_T^{(2)}}{\partial\theta^{(2)\prime}}\right|_{\theta^{(2)}=\bar{\theta}^{(2)}} T^{1/2}(\hat{\theta}_T^{(2)}-\theta_*^{(2)}) && (2.73)\\ &= T^{1/2}\{NR_T^{(1)}(\theta_*^{(1)}) - NR_*^{(1)}(\theta_*^{(1)})\}\\ &- T^{1/2}\{NR_T^{(2)}(\theta_*^{(2)}) - NR_*^{(2)}(\theta_*^{(2)})\}\\ &+ \left.\frac{\partial NR_T^{(1)}}{\partial\theta^{(1)\prime}}\right|_{\theta^{(1)}=\bar{\theta}^{(1)}} T^{1/2}(\hat{\theta}_T^{(1)}-\theta_*^{(1)})\\ &- \left.\frac{\partial NR_T^{(2)}}{\partial\theta^{(2)\prime}}\right|_{\theta^{(2)}=\bar{\theta}^{(2)}} T^{1/2}(\hat{\theta}_T^{(2)}-\theta_*^{(2)}) && (2.74)\end{aligned}$$

First of all,

$$\begin{aligned} NR_T^{(i)}(\theta_*^{(i)}) - NR_*^{(i)}(\theta_*^{(i)}) &= m_T^{(i)}(\theta_*^{(i)})S_T^{(i)}(\theta_*^{(i)})^{-1}m_T^{(i)}(\theta_*^{(i)})\\ &- m_*^{(i)}(\theta_*^{(i)})S_*^{(i)}(\theta_*^{(i)})^{-1}m_*^{(i)}(\theta_*^{(i)}) && (2.75)\end{aligned}$$

---

[15] In the case of the correctly-specified models, we have $T^{1/2}NR_T^{(i)}(\theta_*^{(i)}) = 0$, whereas, in the case of models with non-local misspecification, we have $T^{1/2}NR_T^{(i)}(\theta_*^{(i)}) = T^{1/2}m_T^{(i)}(\theta_*^{(i)})W_T^{(i)}(\theta_*^{(i)})^{-1}m_T^{(i)}(\theta_*^{(i)}) = O_p(1)$. This allows us to use $T^{1/2}$ as the correct rate of convergence, hence we can compute the limit of the numerator of B_T directly.

From Equation (2.75), we see that $NR_T^{(1)}(\theta_*^{(1)}) - NR_*^{(1)}(\theta_*^{(1)})$ is in the form of $\hat{h}'\hat{S}\hat{h} - h'Sh$, which we can write as:[16]

$$\hat{h}'\hat{S}\hat{h} - h'Sh = \hat{h}'\hat{S}(\hat{h} - h) + \hat{h}'(\hat{S} - S)h + (\hat{h} - h)Sh \tag{2.76}$$

Applying the expansion in (2.76) to (2.75) results in the following:[17],

$$\begin{aligned} T^{1/2}\{NR_T^{(i)}(\theta_*^{(i)}) - NR_*^{(i)}(\theta_*^{(i)})\} = {} & 2\mu^{(i)}(\theta_*^{(i)})' S_*^{(i)}(\theta_*^{(i)})^{-1} T^{-1/2} \sum_{t=1}^{T} [z_t^{(i)} g_t^{(i)}(\theta_*^{(i)}) - \mu_*^{(i)}] \\ & + \mu_*^{(i)\prime} T^{-1/2}\{S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*^{(i)}(\theta_*^{(i)})^{-1}\}\mu_*^{(i)} \\ & + o_p(1) \end{aligned} \tag{2.77}$$

For CUE-IV, the choice of weighting matrix we use is:

$$S_T^{(i)}(\theta) = T^{-1} \sum_{t=1}^{T} g_t^{(i)}(\theta^{(i)})^2 z_t^{(i)} z_t^{(i)\prime} \tag{2.78}$$

Using $A^{-1} - B^{-1} = -B^{-1}[A - B]A^{-1}$, we can write the following analog to Equation (60) of Hall and Pelletier (2008):

$$\begin{aligned} \mu_*^{(i)\prime} T^{1/2}\{S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*^{(i)}(\theta_*^{(i)})^{-1}\}\mu_*^{(i)} = {} & vec\{\mu_*^{(i)\prime} \\ & \times T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*^{(i)}(\theta_*^{(i)})^{-1}]\mu_*^{(i)}\} && (2.79) \\ = {} & [\mu_*^{(i)\prime} \otimes \mu_*^{(i)\prime}] \\ & \times vec\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*^{(i)}(\theta_*^{(i)})^{-1}]\} && (2.80) \\ = {} & -[\mu_*^{(i)\prime} \otimes \mu_*^{(i)\prime}] vec\{S_*^{(i)}(\theta_*^{(i)})^{-1} \\ & \times T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*^{(i)}(\theta_*^{(i)})] S_T^{(i)}(\theta_*^{(i)})^{-1}\} && (2.81) \\ = {} & -[\mu_*^{(i)\prime} \otimes \mu_*^{(i)\prime}][S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_*^{(i)}(\theta_*^{(i)})^{-1}] && (2.82) \\ & \times vec\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*^{(i)}(\theta_*^{(i)})]\} && (2.83) \end{aligned}$$

---

[16] See Hall and Pelletier (2008)[p.35].

[17] See Equations (58)-(60) of Hall and Pelletier (2008).

$$\begin{aligned}
&= -\left[\mu_*^{(i)\prime} \otimes \mu_*^{(i)\prime}\right]\left[S_*^{(i)}(\theta_*^{(i)})^{-1} \otimes S_*^{(i)}(\theta_*^{(i)})^{-1}\right] B^{(i)} \\
&\times T^{-1/2} \sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)})^2 vech\left[z_t^{(i)} z_t^{(i)\prime} - E[z_t^{(i)} z_t^{(i)\prime}]\right] \\
&+ o_p(1)
\end{aligned} \tag{2.84}$$

where B is the permutation matrix such that $vec(A) = Bvech(A)$, for some conformable matrix A. Using Equation (2.84), we can re-write Equation (2.77) as follows:

$$\begin{aligned}
T^{1/2}\left\{NR_T^{(i)}(\theta_*^{(i)}) - NR_*^{(i)}(\theta_*^{(i)})\right\} &= 2\mu^{(i)}(\theta_*^{(i)})' S_*^{(i)}(\theta_*^{(i)})^{-1} T^{-1/2} \sum_{t=1}^{T} [z_t^{(i)} g_t^{(i)}(\theta_*^{(i)}) - \mu_*^{(i)}] \\
&+ \left[\mu_*^{(i)\prime} \otimes \mu_*^{(i)\prime}\right]\left[S_*^{(i)}(\theta_*^{(i)})^{-1} \otimes S_*^{(i)}(\theta_*^{(i)})^{-1}\right] \\
&\times B^{(i)} T^{-1} \sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)})^2 vech[z_t^{(i)} z_t^{(i)\prime} - E[z_t^{(i)} z_t^{(i)\prime}]] \\
&+ o_p(1)
\end{aligned} \tag{2.85}$$

2.3.1 Analysis of $\partial NR_T^{(i)} / \partial \theta^{(i)}$

Since we are studying the CUE-IV estimator, from Equation (2.40), we have $W_T^{(i)}(\hat{\theta}_T^{(i)}) = S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}$. Next, we let $E[g_t^{(i)}(\theta_*^{(i)}, v_t) z_t^{(i)}] = \mu_*^{(i)}$. In this case, the derivative matrix in (2.72) converges as follows:

$$\begin{aligned}
\left.\frac{\partial NR_T^{(i)}}{\partial \theta^{(i)\prime}}\right|_{\theta^{(i)} = \bar{\theta}^{(i)}} &= \left(\frac{\partial NR_T^{(i)}}{\partial m_T^{(i)\prime}}\right)' \left.\left(\frac{\partial m_T^{(i)}}{\partial \theta^{(i)\prime}}\right)\right|_{\theta^{(i)} = \bar{\theta}^{(i)}} \\
&+ \left.\left(\frac{\partial NR_T^{(i)}}{\partial vec(S_T^{-1})'}\right)\left(\frac{\partial vec(S_T^{-1})}{\partial vec(S_T)'}\right)\left(\frac{\partial vec(S_T)}{\partial \theta^{(i)\prime}}\right)\right|_{\theta^{(i)} = \bar{\theta}^{(i)}}
\end{aligned} \tag{2.86}$$

$$\begin{aligned}
&= 2m_T^{(i)}(\bar{\theta}^{(i)})' S_T^{(i)}(\bar{\theta}^{(i)})^{-1} G_T^{(i)}(\bar{\theta}^{(i)}) \\
&+ \left(m_T^{(i)}(\bar{\theta}^{(i)})' \otimes m_T^{(i)}(\bar{\theta}^{(i)})'\right)\left(\frac{\partial vec(S_T^{(i)-1})}{\partial vec(S_T^{(i)})'}\right)\left(\left.\frac{\partial vec(S_T^{(i)})}{\partial \theta^{(i)\prime}}\right|_{\theta^{(i)} = \bar{\theta}^{(i)}}\right)
\end{aligned} \tag{2.87}$$

$$\begin{aligned}
&= 2m_T^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} G_T^{(i)}(\theta_*^{(i)}) \\
&- \left(m_T^{(i)}(\theta_*^{(i)})' \otimes m_T^{(i)}(\theta_*^{(i)})'\right)\left(S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_T^{(i)}(\theta_*^{(i)})^{-1}\right) \Sigma_T^{(i)}(\theta_*^{(i)}) \\
&+ o_p(1)
\end{aligned} \tag{2.88}$$

$$
\begin{aligned}
&= 2\mu_*^{(i)'} S_*^{(i)}(\theta_*^{(i)})^{-1} G_*^{(i)}(\theta_*^{(i)}) \\
&+ (\mu_*^{(i)'} \otimes \mu_*^{(i)'})(S_*^{(i)}(\theta_*^{(i)})^{-1} \otimes S_*^{(i)}(\theta_*^{(i)})^{-1})\Sigma_*^{(i)} + o_p(1) \qquad (2.89) \\
&= \Pi_*^{(i)} + o_p(1) \qquad (2.90)
\end{aligned}
$$

where $\Sigma_T^{(i)}(\theta^{(i)}) = \partial vec(S_T^{(i)})/\partial \theta^{(i)'}$, $plim(\Sigma_T^{(i)}(\theta_*^{(i)})) = \Sigma^{(i)} NR_*^{(1)}(\theta_*^{(1)})$ and $\Pi^{(i)} = \partial NR_T^{(i)}/\partial \theta^{(i)'}$.

2.3.2 Distribution of CUE: Misspecified models

In what follows, using an analysis similar to Hall and Inoue (2003), we derive the asymptotic distribution of $T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)})$ for the CUE. First, a mean value expansion of $m_T^{(i)}(\hat{\theta}_T^{(i)})$ around $\theta_*^{(i)}$ yields:

$$
m_T^{(i)}(\hat{\theta}_T^{(i)}) = m_T^{(i)}(\theta_*^{(i)}) + G_T^{(i)}(\bar{\theta}_T^{(i)})(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \qquad (2.91)
$$

Pre-multiplying (2.91) by $\mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)})$ gives:

$$
\mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)}) m_T^{(i)}(\hat{\theta}_T^{(i)}) = \mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)}) m_T^{(i)}(\theta_*^{(i)}) + \mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)}) G_T^{(i)}(\bar{\theta}_T^{(i)})(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \qquad (2.92)
$$

The left hand side of (2.92) is zero by Equation (2.11). Assuming $[\mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)}) G_T^{(i)}(\bar{\theta}_T^{(i)})]^{-1}$ exists, we can write:

$$
T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) = H_{0,T}\{H_{1,T} + H_{2,T}\} \qquad (2.93)
$$

Also, we use the following:

$$
\Sigma_T(\theta) = \frac{\partial vec[S_T^{(i)}(\theta)]}{\partial \theta'} \qquad (2.94)
$$

Next, we give the forms of the functions in (2.93):

$$
\begin{aligned}
H_{0,T} &= -[\mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)}) G_T^{(i)}(\bar{\theta}_T^{(i)})]^{-1} \qquad (2.95) \\
H_{1,T} &= \mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)}) T^{-1/2} \sum_{t=1}^{T} [g_t^{(i)}(\theta_*^{(i)}, v_t) z_t^{(i)} - \mu_*^{(i)}] \qquad (2.96) \\
H_{2,T} &= \mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)}) T^{1/2} \mu_*^{(i)} \qquad (2.97) \\
&= H_{2,T}(1) + H_{2,T}(2) + H_{2,T}(3) + H_{2,T}(4) + H_{2,T}(5) + H_{2,T}(6) \qquad (2.98)
\end{aligned}
$$

$$H_{2,T}(1) = T^{1/2}[G_T^{(i)}(\hat{\theta}_T^{(i)}) - g_t^{(i)}(\theta_*^{(i)}, v_t)]' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\mu_*^{(i)} \tag{2.99}$$

$$H_{2,T}(2) = T^{1/2}[G_T^{(i)}(\theta_*^{(i)}) - G_*^{(i)}(\theta_*^{(i)})]' S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\mu_*^{(i)} \tag{2.100}$$

$$H_{2,T}(3) = G_*^{'} T^{1/2}[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} - S_T^{(i)}(\theta_*^{(i)})^{-1}]\mu_*^{(i)} \tag{2.101}$$

$$H_{2,T}(4) = G_*^{'} T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*(\theta_*^{(i)})^{-1}]\mu_*^{(i)} \tag{2.102}$$

$$H_{2,T}(5) = T^{1/2}[G_*^{'} S_*(\theta_*^{(i)})^{-1}\mu_*^{(i)}] \tag{2.103}$$

$$H_{2,T}(6) = -T^{1/2}\Sigma_T(\hat{\theta}_T^{(i)})'[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}](m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \tag{2.104}$$

$$= H_{2,T}(6,1) + H_{2,T}(6,2) + H_{2,T}(6,3) \tag{2.105}$$

$$H_{2,T}(6,1) = T^{1/2}[\Sigma_T(\hat{\theta}_T^{(i)}) - \Sigma_T(\theta_*^{(i)})]'[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}](m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \tag{2.106}$$

$$H_{2,T}(6,2) = T^{1/2}[\Sigma_T(\theta_*^{(i)}) - \Sigma_*(\theta_*^{(i)})]'[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}](m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \tag{2.107}$$

$$H_{2,T}(6,3) = T^{1/2}\Sigma_*(\theta_*^{(i)})'[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}](m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \tag{2.108}$$

$$= H_{2,T}(6,3,1) + H_{2,T}(6,3,2) + H_{2,T}(6,3,3) \tag{2.109}$$

$$H_{2,T}(6,3,1) = \Sigma_*(\theta_*^{(i)})'\{T^{1/2}[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} - S_T^{(i)}(\theta_*^{(i)})^{-1}] \otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\}(m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \tag{2.110}$$

$$H_{2,T}(6,3,2) = \Sigma_*(\theta_*^{(i)})'\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*(\theta_*^{(i)})^{-1}] \otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\}(m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \tag{2.111}$$

$$H_{2,T}(6,3,3) = T^{1/2}\Sigma_*(\theta_*^{(i)})'[S_*(\theta_*^{(i)})^{-1} \otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}](m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \tag{2.112}$$

$$= H_{2,T}(6,3,3,1) + H_{2,T}(6,3,3,2) + H_{2,T}(6,3,3,3) \tag{2.113}$$

$$\begin{aligned} H_{2,T}(6,3,3,1) = {} & \Sigma_*(\theta_*^{(i)})'\{S_*(\theta_*)^{-1} \otimes T^{1/2}[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} - S_T^{(i)}(\theta_*^{(i)})^{-1}]\} \\ & \times \ (m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \end{aligned} \tag{2.114}$$

$$H_{2,T}(6,3,3,2) = \Sigma_*(\theta_*^{(i)})'\{S_*(\theta_*)^{-1} \otimes T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*(\theta_*^{(i)})^{-1}]\}(m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \tag{2.115}$$

$$H_{2,T}(6,3,3,3) = T^{1/2}\Sigma_*(\theta_*^{(i)})'[S_*(\theta_*^{(i)})^{-1} \otimes S_*(\theta_*^{(i)})^{-1}](m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \tag{2.116}$$

$$H_{2,T}(6,3,3,3) = H_{2,T}(6,3,3,3,1) + H_{2,T}(6,3,3,3,2) + H_{2,T}(6,3,3,3,3) \tag{2.117}$$

$$H_{2,T}(6,3,3,3,1) = \Sigma_*(\theta_*^{(i)})'[S_*(\theta_*^{(i)})^{-1} \otimes S_*(\theta_*^{(i)})^{-1}](T^{1/2}[m_T^{(i)}(\hat{\theta}_T^{(i)}) - m_T^{(i)}(\theta_*^{(i)})] \otimes I_{k_i})\mu_*^{(i)} \tag{2.118}$$

$$H_{2,T}(6,3,3,3,2) = \Sigma_*(\theta_*^{(i)})'[S_*(\theta_*^{(i)})^{-1} \otimes S_*(\theta_*^{(i)})^{-1}](T^{1/2}[m_T^{(i)}(\theta_*^{(i)}) - \mu_*^{(i)}] \otimes I_{k_i})\mu_*^{(i)} \tag{2.119}$$

$$= H_{3,T}\left\{T^{-1/2}\left[\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)} - \mu_*^{(i)}\right] \otimes I_{k_i}\right\}\mu_*^{(i)} \tag{2.120}$$

$$H_{2,T}(6,3,3,3,3) = T^{1/2}\Sigma_*(\theta_*^{(i)})'[S_*(\theta_*^{(i)})^{-1} \otimes S_*(\theta_*^{(i)})^{-1}](\mu_*^{(i)} \otimes I_{k_i})\mu_*^{(i)} \tag{2.121}$$

where $H_{3,T} = \Sigma_*(\theta_*^{(i)})'[S_*(\theta_*^{(i)})^{-1} \otimes S_*(\theta_*^{(i)})^{-1}]$ and $plim(\Sigma_T(\theta_*^{(i)})) = \Sigma_*(\theta_*^{(i)})$. First, the population analog of the f.o.c. for CUE implies $H_{2,T}(5) - H_{2,T}(6,3,3,3,3) = 0$. Further, we can use the following expansions:[18]

$$H_{2,T}(1) = (\mu_*^{(i)'} S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \otimes I_p)vec\{T^{1/2}[G_T^{(i)}(\hat{\theta}_T^{(i)}) - G_T^{(i)}(\theta_*^{(i)})]\} \tag{2.122}$$

$$= (\mu_*^{(i)'} S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \otimes I_p)G_T^{(2,i)}(\tilde{\theta}_T^{(i)})T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.123}$$

$$= M_{1,T}T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.124}$$

$$H_{2,T}(3) = -G_*^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1}T^{1/2}[S_T^{(i)}(\hat{\theta}_T^{(i)}) - S_T^{(i)}(\theta_*^{(i)})]S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\mu_*^{(i)} \tag{2.125}$$

$$= -(\mu_*^{(i)'} S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes G_*^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1})vec\{T^{1/2}[S_T^{(i)}(\hat{\theta}_T^{(i)}) - S_T^{(i)}(\theta_*^{(i)})]\} \tag{2.126}$$

$$= (\mu_*^{(i)'} S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes G_*^{(i)}(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1})\Sigma_T(\tilde{\theta}_T^{(i)})T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.127}$$

$$= M_{2,T}T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.128}$$

$$H_{2,T}(6,1) = T^{1/2}[\Sigma_T(\hat{\theta}_T^{(i)}) - \Sigma_T(\theta_*^{(i)})]'[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}](m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \tag{2.129}$$

$$\begin{aligned} &= -\left\{ \left\langle \mu_*^{(i)'}[m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i}]'[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}] \right\rangle \otimes I_{k_i} \right\} \\ &\quad \times \ vec\{T^{1/2}[\Sigma_T(\hat{\theta}_T^{(i)}) - \Sigma_T(\theta_*^{(i)})]'\} \end{aligned} \tag{2.130}$$

$$\begin{aligned} &= -\left\{ \left\langle \mu_*^{(i)'}[m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i}]'[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}]) \right\rangle \otimes I_{k_i} \right\} \\ &\quad \times \ \Sigma_T^2(\check{\theta}_T)T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \end{aligned} \tag{2.131}$$

$$= -M_{3,T}T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.132}$$

where $\Sigma_T^2(\check{\theta}_T)$ is the $k^2 \times p$ matrix whose r^{th} row is the corresponding row of

$$(\partial/\partial)vec(\partial S_T^{(i)}(\check{\theta}_T^{(r)})z_t/\partial\theta^{(i)'})$$

---

[18] Equation (2.125) follows from the fact that, for two conformable matrices, A and B, it follows that: $A^{-1} - B^{-1} = B^{-1}(A - B)A^{-1}$.

with $\breve{\theta}_T^{(r)} = \epsilon^{(r)}\hat{\theta}_T + (1-\epsilon^{(r)})\theta_*$, where $0 \le \epsilon^{(r)} \le 1$.

Further, we have:

$$H_{2,T}(6,3,1) = \Sigma_*(\theta_*^{(i)})'\{T^{1/2}[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} - S_T^{(i)}(\theta_*^{(i)})^{-1}]\otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\}(m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \tag{2.133}$$

$$= -\Big\{ \Big\langle \mu_*^{(i)'}[m_T^{(i)}(\hat{\theta}_T^{(i)})' \otimes I_{k_i}][S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}] \Big\rangle \otimes \Sigma_*(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} \Big\}$$

$$\times \ vec\{T^{1/2}[S_T^{(i)}(\hat{\theta}_T^{(i)}) - S_T^{(i)}(\theta_*^{(i)})]\} \tag{2.134}$$

$$= -\Big\{ \Big\langle \mu_*^{(i)'}[m_T^{(i)}(\hat{\theta}_T^{(i)})' \otimes I_{k_i}]'[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}]) \Big\rangle \otimes \Sigma_*(\theta_*^{(i)})' S_T^{(i)}(\theta_*^{(i)})^{-1} \Big\}$$

$$\times \ \Sigma_T(\breve{\theta}_T)T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.135}$$

$$= -M_{4,T}T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.136}$$

$$H_{2,T}(6,3,3,1) = -\Big\{\mu_*^{(i)'}[m_T^{(i)}(\hat{\theta}_T^{(i)})' \otimes I_{k_i}]'[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\otimes\Sigma_*(\theta_*^{(i)})'][S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}]\Big\}$$

$$\times \ vec\{T^{1/2}[S_T^{(i)}(\hat{\theta}_T^{(i)}) - S_T^{(i)}(\theta_*^{(i)})]\} \tag{2.137}$$

$$= -\Big\{\mu_*^{(i)'}[m_T^{(i)}(\hat{\theta}_T^{(i)})' \otimes I_{k_i}]'[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\otimes\Sigma_*(\theta_*^{(i)})'][S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}]\Big\}$$

$$\times \ \Sigma_T(\breve{\theta}_T)T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.138}$$

$$= -M_{5,T}T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.139}$$

$$H_{2,T}(6,3,3,3,1) = \Big\{\mu_*^{(i)'}\Sigma_*(\theta_*^{(i)})[S_*(\theta_*^{(i)})^{-1} \otimes S_*(\theta_*^{(i)})^{-1}]\Big\}$$

$$\times \ vec\Big\{T^{1/2}[m_T^{(i)}(\hat{\theta}_T^{(i)}) - m_T^{(i)}(\theta_*^{(i)})] \otimes I_{k_i}\Big\} \tag{2.140}$$

$$= \Big\{\mu_*^{(i)'}\Sigma_*(\theta_*^{(i)})[S_*(\theta_*^{(i)})^{-1} \otimes S_*(\theta_*^{(i)})^{-1}]\Big\}$$

$$\times \ G_T^{(2,i)}(\tilde{\theta}_T^{(i)})T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.141}$$

$$= M_{6,T}T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{2.142}$$

where $G_T^{(2,i)}(\tilde{\theta}_T^{(i)})$ is the $kp \times p$ matrix whose r^{th} row is the corresponding row of

$$(\partial/\partial)vec(\partial g_t(\tilde{\theta}_T^{(r)})z_t/\partial\theta^{(i)'})$$

with $\tilde{\theta}_T^{(r)} = \phi^{(r)}\hat{\theta}_T + (1-\phi^{(r)})\theta_*$, where $0 \leq \phi^{(r)} \leq 1$.
$H_{2,T}(4)$, $H_{2,T}(6,3,2)$ and $H_{2,T}(6,3,3,2)$ all are functions of $T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*(\theta_*^{(i)})^{-1}]$. Therefore, we can join them into one factor.
First, we can write $H_{2,T}(4)$ as follows:

$$\begin{aligned}
H_{2,T}(4) &= [\mu_*^{(i)'} \otimes G_*']vec\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*(\theta_*^{(i)})^{-1}]\} && (2.143)\\
&= -[\mu_*^{(i)'} \otimes G_*']vec\{S_*(\theta_*^{(i)})^{-1}T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*(\theta_*^{(i)})]S_T^{(i)}(\theta_*^{(i)})^{-1}\} && (2.144)\\
&= -[\mu_*^{(i)'} \otimes G_*'][S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_*(\theta_*^{(i)})^{-1}]vec\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*(\theta_*^{(i)})]\} && (2.145)\\
&= -[\mu_*^{(i)'} \otimes G_*'][S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_*(\theta_*^{(i)})^{-1}]Bvech\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*(\theta_*^{(i)})]\} && (2.146)\\
&= -[\mu_*^{(i)'}S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes G_*'S_*(\theta_*^{(i)})^{-1}]Bvech\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*(\theta_*^{(i)})]\} && (2.147)
\end{aligned}$$

where (2.143) follows from the fact that $H_{2,T}(4)$ is $p \times 1$ using $vec(ABC) = (C' \otimes A)vec(B)$. Equation (2.144) follows from (2.143), using the fact that $A^{-1} - B^{-1} = -B^{-1}[A-B]A^{-1}$. Equation (2.145) follows from (2.144), using the fact that $vec(ABC) = (C' \otimes A)vec(B)$. Finally, (2.146) follows from $vec(A) = Bvech(A)$. Next, we can write $H_{2,T}(6,3,2)$ as follows:

$$\begin{aligned}
H_{2,T}(6,3,2) &= \Sigma_*(\theta_*^{(i)})'\{S_*(\theta_*)^{-1} \otimes T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*(\theta_*^{(i)})^{-1}]\}(m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} && (2.148)\\
&= \Sigma_*(\theta_*^{(i)})'\{S_*(\theta_*)^{-1} \otimes T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*(\theta_*^{(i)})^{-1}]\}vec(\mu_*^{(i)}m_T^{(i)}(\hat{\theta}_T^{(i)})') && (2.149)\\
&= \Sigma_*(\theta_*^{(i)})'[I \otimes S_T^{(i)}(\theta_*^{(i)})^{-1}\mu_*^{(i)}m_T^{(i)}(\hat{\theta}_T^{(i)})']vec\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*(\theta_*^{(i)})^{-1}]\} && (2.150)\\
&= -\Sigma_*(\theta_*^{(i)})'[I \otimes S_T^{(i)}(\theta_*^{(i)})^{-1}\mu_*^{(i)}m_T^{(i)}(\hat{\theta}_T^{(i)})']\\
&\quad \times vec\{S_*(\theta_*^{(i)})^{-1}T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*(\theta_*^{(i)})]S_T^{(i)}(\theta_*^{(i)})^{-1}\} && (2.151)
\end{aligned}$$

$$
\begin{aligned}
&= -\Sigma_*(\theta_*^{(i)})'[I \otimes S_T^{(i)}(\theta_*^{(i)})^{-1}\mu_*^{(i)} m_T^{(i)}(\hat{\theta}_T^{(i)})'][S_T^{(i)}(\theta_*^{(i)})^{-1}T^{1/2} \otimes S_*(\theta_*^{(i)})^{-1}] \\
&\quad \times vec\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*(\theta_*^{(i)})]\} \qquad (2.152) \\
&= -\Sigma_*(\theta_*^{(i)})'[S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_T^{(i)}(\theta_*^{(i)})^{-1}\mu_*^{(i)} m_T^{(i)}(\hat{\theta}_T^{(i)})' S_*(\theta_*^{(i)})^{-1}] \\
&\quad \times vec\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*(\theta_*^{(i)})]\} \qquad (2.153) \\
&= -\Sigma_*(\theta_*^{(i)})'[S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_T^{(i)}(\theta_*^{(i)})^{-1}\mu_*^{(i)} m_T^{(i)}(\hat{\theta}_T^{(i)})' S_*(\theta_*^{(i)})^{-1}] \\
&\quad \times Bvech\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*(\theta_*^{(i)})]\} \qquad (2.154)
\end{aligned}
$$

Also, we can write $H_{2,T}(6,3,3,2)$ as follows:

$$
\begin{aligned}
H_{2,T}(6,3,3,2) &= \Sigma_*(\theta_*^{(i)})'\{S_*(\theta_*)^{-1} \otimes T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*(\theta_*^{(i)})^{-1}]\}(m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)} \qquad (2.155) \\
&= \Sigma_*(\theta_*^{(i)})'\{S_*(\theta_*)^{-1} \otimes T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*(\theta_*^{(i)})^{-1}]\}vec(\mu_*^{(i)} m_T^{(i)}(\hat{\theta}_T^{(i)})') \qquad (2.156) \\
&= \Sigma_*(\theta_*^{(i)})'\{S_*(\theta_*^{(i)})^{-1}[\mu_*^{(i)} m_T^{(i)}(\hat{\theta}_T^{(i)})'] \otimes I_{k_i}\} \\
&\quad \times vec\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)})^{-1} - S_*(\theta_*^{(i)})^{-1}]\} \qquad (2.157) \\
&= -\Sigma_*(\theta_*^{(i)})'\{S_*(\theta_*^{(i)})^{-1}[\mu_*^{(i)} m_T^{(i)}(\hat{\theta}_T^{(i)})'] \otimes I_{k_i}\}[S_T^{(i)}(\theta_*^{(i)})^{-1}T^{1/2} \otimes S_*(\theta_*^{(i)})^{-1}] \\
&\quad \times vec\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*(\theta_*^{(i)})]\} \qquad (2.158) \\
&= -\Sigma_*(\theta_*^{(i)})'\{S_*(\theta_*^{(i)})^{-1}[\mu_*^{(i)} m_T^{(i)}(\hat{\theta}_T^{(i)})']S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_*(\theta_*^{(i)})^{-1}\} \\
&\quad \times Bvech\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*(\theta_*^{(i)})]\} \qquad (2.159)
\end{aligned}
$$

Let $\mathscr{M}_T = H_{2,T}(4) + H_{2,T}(6,3,2) + H_{2,T}(6,3,3,2)$. Using Equations (2.78), (2.147), (2.154) and (2.159), we can write the following:

$$
\begin{aligned}
\mathscr{M}_T &= -\Xi_T Bvech\{T^{1/2}[S_T^{(i)}(\theta_*^{(i)}) - S_*(\theta_*^{(i)})]\} \qquad (2.160) \\
&= -\Xi_T B T^{-1/2} \sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)})^2 vech\big[z_t^{(i)} z_t^{(i)\prime} - E[z_t^{(i)} z_t^{(i)\prime}]\big] \qquad (2.161)
\end{aligned}
$$

where B is a permutation matrix such that, for any matrix A, we have $vec(A) = Bvech(A)$ and:

$$\begin{aligned}\Xi_T &= [\mu_*^{(i)'} S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes G_*' S_*(\theta_*^{(i)})^{-1}] \\ &\quad + \Sigma_*(\theta_*^{(i)})' [S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_T^{(i)}(\theta_*^{(i)})^{-1} \mu_*^{(i)} m_T^{(i)}(\hat{\theta}_T^{(i)})' S_*(\theta_*^{(i)})^{-1}] \\ &\quad + \Sigma_*(\theta_*^{(i)})' \{S_*(\theta_*^{(i)})^{-1} [\mu_*^{(i)} m_T^{(i)}(\hat{\theta}_T^{(i)})'] S_T^{(i)}(\theta_*^{(i)})^{-1} \otimes S_*(\theta_*^{(i)})^{-1}\} \qquad (2.162)\end{aligned}$$

It also follows that:

$$\begin{aligned}H_{2,T}(2) &= \big([\mu_*^{(i)'} M_{zz}^{(i)-1}] \otimes I_{p_i}\big) T^{-1/2} \sum_{t=1}^{T} vec\big[D_{zz,t}^{(i)} - D_{zz}^{(i)}\big] \qquad (2.163) \\ &= H_{5,T} T^{-1/2} \sum_{t=1}^{T} vec\big[D_{zz,t}^{(i)} - D_{zz}^{(i)}\big] \qquad (2.164)\end{aligned}$$

where $D_{zz,t}^{(i)} = \partial g_t^{(i)}(\theta)/\partial\theta^{(i)'}]' z_t^{(i)'}$ and $D_{zz}^{(i)} = E[z_t^{(i)}(\partial g_t^{(i)}(\theta)/\partial\theta^{(i)'})]'$.
Given Equations (2.41) and (2.94), we have:

$$\begin{aligned}vec(\Sigma_T(\theta)) &= vec\Big\{\partial vec\big(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta^{(i)})^2 z_t z_t'\big)/\partial\theta^{(i)'}\Big\} \qquad (2.165) \\ &= vec\Big\{2T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta^{(i)}) vec[z_t z_t']\Big(\frac{g_t^{(i)}(\theta^{(i)})}{\theta^{(i)'}}\Big)\Big\} \qquad (2.166) \\ &= 2T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta^{(i)}) vec\Big\{vec[z_t z_t']\Big(\frac{g_t^{(i)}(\theta^{(i)})}{\theta^{(i)'}}\Big)\Big\} \qquad (2.167) \\ &= 2T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta^{(i)})\Big\{\Big[vec\big(z_t z_t'\big)\Big] \otimes \Big[vec\Big(\frac{g_t^{(i)}(\theta^{(i)})}{\theta^{(i)'}}\Big)\Big]\Big\} \qquad (2.168) \\ &= 2T^{-1}\sum_{t=1}^{T} \varsigma_t(\theta) \qquad (2.169)\end{aligned}$$

where

$$\varsigma_t(\theta) = g_t^{(i)}(\theta^{(i)})\Big\{\Big[vec\big(z_t z_t'\big)\Big] \otimes \Big[vec\Big(\frac{g_t^{(i)}(\theta^{(i)})}{\theta^{(i)'}}\Big)\Big]\Big\} \qquad (2.170)$$

Given Equations(2.78), (2.107) and (2.169), it follows that:

$$H_{2,T}(6,2) = vec\big(T^{1/2}[\Sigma_T(\theta_*^{(i)}) - \Sigma_*(\theta_*^{(i)})]'[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}](m_T^{(i)}(\hat{\theta}_T^{(i)}) \otimes I_{k_i})\mu_*^{(i)}\big) \tag{2.171}$$

$$= \Big\{ \big[\mu_*^{(i)\prime}\big(m_T^{(i)}(\hat{\theta}_T^{(i)})' \otimes I_{k_i}\big)\big]\big[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\big] \otimes I_p \Big\} \times vec\big\{T^{1/2}[\Sigma_T(\theta_*^{(i)}) - \Sigma_*(\theta_*^{(i)})]'\big\} \tag{2.172}$$

$$= 2\Big\{ \big[\mu_*^{(i)\prime}\big(m_T^{(i)}(\hat{\theta}_T^{(i)})' \otimes I_{k_i}\big)\big]\big[S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} \otimes S_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}\big] \otimes I_p \Big\} \times T^{-1/2}\sum_{t=1}^{T}\big[\varsigma_t(\theta_*^{(i)}) - E[\varsigma_t(\theta_*^{(i)})]\big]' \tag{2.173}$$

$$= H_{6,T}T^{-1/2}\sum_{t=1}^{T}\big[\varsigma_t(\theta_*^{(i)}) - E[\varsigma_t(\theta_*^{(i)})]\big]' \tag{2.174}$$

Finally, we can write the following:

$$T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) = \mho_T\bigg\{\bigg(\mathcal{N}_T^{(i)}(\hat{\theta}_T^{(i)})T^{-1/2}\sum_{t=1}^{T}[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)} - \mu_*^{(i)}]\bigg) + \bigg\langle H_{3,T}\bigg[\bigg(T^{-1/2}\sum_{t=1}^{T}[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)} - \mu_*^{(i)}]\bigg) \otimes I_{k_i}\bigg]\mu_*^{(i)}\bigg\rangle + H_T\bigg\} \tag{2.175}$$

$$= \mho_T\bigg\{\mathcal{N}_T^{(i)}(\hat{\theta}_T^{(i)})\bigg(T^{-1/2}\sum_{t=1}^{T}[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)} - \mu_*^{(i)}]\bigg) + H_{3,T}(I_{k_i} \otimes \mu_*^{(i)})\bigg(T^{-1/2}\sum_{t=1}^{T}[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)} - \mu_*^{(i)}]\bigg) + H_T\bigg\} \tag{2.176}$$

$$= \mho_T\bigg\{H_{4,T}\bigg(T^{-1/2}\sum_{t=1}^{T}[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)} - \mu_*^{(i)}]\bigg) + H_T\bigg\} \tag{2.177}$$

$$
\begin{aligned}
&= \mho_T \Bigg\{ H_{4,T} \Bigg[T^{-1/2} \sum_{t=1}^{T} [g_t^{(i)}(\theta_*^{(i)}, v_t) z_t^{(i)} - \mu_*^{(i)}] \Bigg] \\
&\quad + H_{5,T} \Bigg[T^{-1/2} \sum_{t=1}^{T} vec\big[D_{zz,t}^{(i)} - D_{zz}^{(i)}\big] \Bigg] \\
&\quad + H_{6,T} \Bigg[T^{-1/2} \sum_{t=1}^{T} \big[\varsigma_t(\theta_*^{(i)}) - E[\varsigma_t(\theta_*^{(i)})]\big]' \Bigg] \\
&\quad + H_{7,T} \Bigg[T^{-1/2} \sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)})^2 vech\big[z_t^{(i)} z_t^{(i)\prime} - E[z_t^{(i)} z_t^{(i)\prime}]\big] \Bigg] \Bigg\} \qquad (2.178)
\end{aligned}
$$

where

$$
\begin{aligned}
\mho_T &= [I - H_{0,T} M_T]^{-1} H_{0,T} && (2.179) \\
H_{4,T} &= \mathscr{N}_T^{(i)}(\hat{\theta}_T^{(i)}) + H_{3,T}(I_{k_i} \otimes \mu_*^{(i)}) && (2.180) \\
H_T &= H_{2,T}(2) + H_{2,T}(4) + H_{2,T}(6,2) + H_{2,T}(6,3,2) + H_{2,T}(6,3,3,2) && (2.181) \\
&= H_{2,T}(2) + H_{2,T}(6,2) + \mathscr{M}_T && (2.182) \\
H_{7,T} &= -\Xi_T B && (2.183)
\end{aligned}
$$

Let $\mathscr{B}_T = T^{1/2}\big\{NR_T^{(1)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\big\}$. Then, we have:

$$
\begin{aligned}
\mathscr{B}_T &= H_{8,T}^{(1)} T^{-1/2} \sum_{t=1}^{T} [g_t^{(1)}(\theta_*^{(1)}) z_t^{(1)} - \mu_*^{(1)}] \\
&\quad + H_{9,T}^{(1)} T^{-1/2} \sum_{t=1}^{T} g_t^{(1)}(\theta_*^{(1)})^2 vech[z_t^{(1)} z_t^{(1)\prime} - E[z_t^{(1)} z_t^{(1)\prime}]] \\
&\quad + H_{8,T}^{(2)} T^{-1/2} \sum_{t=1}^{T} [g_t^{(2)}(\theta_*^{(2)}) z_t^{(2)} - \mu_*^{(2)}] \\
&\quad + H_{9,T}^{(2)} T^{-1/2} \sum_{t=1}^{T} g_t^{(2)}(\theta_*^{(2)})^2 vech[z_t^{(2)} z_t^{(2)\prime} - E[z_t^{(2)} z_t^{(2)\prime}]] \\
&\quad + \Pi_*^{(1)} \mho_T^{(1)} \Bigg\{ H_{4,T}^{(1)} \Bigg[T^{-1/2} \sum_{t=1}^{T} [g_t^{(1)}(\theta_*^{(1)}) z_t^{(1)} - \mu_*^{(1)}] \Bigg] \qquad (2.184)
\end{aligned}
$$

$$+ H_{5,T}^{(1)}\Bigg[T^{-1/2}\sum_{t=1}^{T} vec\big(\hat{D}_{zz,t}^{(1)} - D_{zz}^{(1)}\big)\Bigg]$$

$$+ H_{6,T}^{(1)}\Bigg[T^{-1/2}\sum_{t=1}^{T} \big[\varsigma_t^{(1)}(\theta_*^{(i)}) - E[\varsigma_t^{(1)}(\theta_*^{(i)})]\big]'\Bigg] \tag{2.185}$$

$$+ H_{7,T}^{(1)}\Bigg[T^{-1/2}\sum_{t=1}^{T} g_t^{(1)}(\theta_*^{(1)})^2 vech\big[z_t^{(1)}z_t^{(1)'} - E[z_t^{(1)}z_t^{(1)'}]\big]\Bigg]\Bigg\}$$

$$+ \Pi_*^{(2)}\mho_T^{(2)}\Bigg\{H_{4,T}^{(2)}\Bigg[T^{-1/2}\sum_{t=1}^{T}[g_t^{(2)}(\theta_*^{(i)})z_t^{(2)} - \mu_*^{(2)}]\Bigg] \tag{2.186}$$

$$+ H_{5,T}^{(2)}\Bigg[T^{-1/2}\sum_{t=1}^{T} vec\big(\hat{D}_{zz,t}^{(2)} - D_{zz}^{(2)}\big)\Bigg]$$

$$+ H_{6,T}^{(2)}\Bigg[T^{-1/2}\sum_{t=1}^{T} \big[\varsigma_t^{(2)}(\theta_*^{(i)}) - E[\varsigma_t^{(2)}(\theta_*^{(i)})]\big]'\Bigg]$$

$$+ H_{7,T}^{(2)}\Bigg[T^{-1/2}\sum_{t=1}^{T} g_t^{(2)}(\theta_*^{(2)})^2 vech\big[z_t^{(2)}z_t^{(2)'} - E[z_t^{(2)}z_t^{(2)'}]\big]\Bigg]\Bigg\} \tag{2.187}$$

$$H_{8,T}^{(i)} = 2\mu^{(i)}(\theta_*^{(i)})' S_*^{(i)}(\theta_*^{(i)})^{-1} \tag{2.188}$$

$$H_{9,T}^{(i)} = \big[\mu_*^{(i)'} \otimes \mu_*^{(i)'}\big]\big[S^{(i)}(\theta_*^{(i)})^{-1} \otimes S^{(i)}(\theta_*^{(i)})^{-1}\big]B^{(i)} \tag{2.189}$$

Next, we rearrange Equation (2.184) as follows:

$$\begin{aligned}
\mathscr{B}_T = {} & \Bigg\{\Bigg[H_{8,T}^{(1)} + \Pi_*^{(1)}\mho_T^{(1)}H_{4,T}^{(1)}\Bigg]\Bigg[T^{-1/2}\sum_{t=1}^{T}[g_t^{(1)}(\theta_*^{(i)})z_t^{(1)} - \mu_*^{(1)}]\Bigg]\Bigg\} \\
& + \Bigg\{\Bigg[H_{9,T}^{(1)} + \Pi_*^{(1)}\mho_T^{(1)}H_{7,T}^{(1)}\Bigg]\Bigg[T^{-1/2}\sum_{t=1}^{T} g_t^{(1)}(\theta_*^{(1)})^2 vech[z_t^{(1)}z_t^{(1)'} - E[z_t^{(1)}z_t^{(1)'}]]\Bigg]\Bigg\} \\
& + \Bigg\{\Bigg[H_{8,T}^{(2)} + \Pi_*^{(2)}\mho_T^{(2)}H_{4,T}^{(2)}\Bigg][T^{-1/2}\sum_{t=1}^{T}[g_t^{(2)}(\theta_*^{(i)})z_t^{(2)} - \mu_*^{(2)}]\Bigg]\Bigg\} \\
& + \Bigg\{\Bigg[H_{9,T}^{(2)} + \Pi_*^{(2)}\mho_T^{(2)}H_{7,T}^{(2)}\Bigg]\Bigg[T^{-1/2}\sum_{t=1}^{T} g_t^{(2)}(\theta_*^{(2)})^2 vech[z_t^{(2)}z_t^{(2)'} - E[z_t^{(2)}z_t^{(2)'}]]\Bigg]\Bigg\} \\
& + \Pi_*^{(1)}\mho_T^{(1)}\Bigg\{H_{5,T}^{(1)}\Bigg[T^{-1/2}\sum_{t=1}^{T} vec\big[D_{zz,t}^{(1)} - D_{zz}^{(1)}\big]\Bigg] + H_{6,T}^{(1)}\Bigg[T^{-1/2}\sum_{t=1}^{T}\big[\varsigma_t^{(1)}(\theta_*^{(i)}) - E[\varsigma_t^{(1)}(\theta_*^{(i)})]\big]'\Bigg]\Bigg\}
\end{aligned} \tag{2.190}$$

$$+\,\Pi_*^{(2)}\mho_T^{(2)}\Bigg\{H_{5,T}^{(2)}\Big[T^{-1/2}\sum_{t=1}^{T} vec\big[D_{zz,t}^{(2)} - D_{zz}^{(2)}\big]\Big]$$

$$+\,H_{6,T}^{(2)}\Big[T^{-1/2}\sum_{t=1}^{T}\big[\varsigma_t^{(2)}(\theta_*^{(i)}) - E[\varsigma_t^{(2)}(\theta_*^{(i)})]\big]'\Big]\Bigg\} \tag{2.191}$$

$$= L_T' T^{1/2} D_T + o_p(1) \tag{2.192}$$

where

$$L_T = \begin{bmatrix} R_{1,T}^{(1)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & R_{2,T}^{(1)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & R_{3,T}^{(1)} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & R_{4,T}^{(1)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & R_{1,T}^{(2)} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & R_{2,T}^{(2)} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & R_{2,T}^{(3)} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & R_{2,T}^{(4)} \end{bmatrix} \tag{2.193}$$

where

$$R_{1,T}^{(i)} = H_{8,T}^{(i)} + \Pi_*^{(i)}[I_{k_i} - H_{0,T}^{(i)} M_T^{(i)}]^{-1} H_{0,T}^{(i)} H_{4,T}^{(i)} \tag{2.194}$$

$$R_{2,T}^{(i)} = H_{9,T}^{(i)} + \Pi_*^{(i)}[I_{k_i} - H_{0,T}^{(i)} M_T^{(i)}]^{-1} H_{0,T}^{(i)} H_{7,T}^{(i)} \tag{2.195}$$

$$R_{3,T}^{(i)} = \Pi_*^{(i)}[I_{k_i} - H_{0,T}^{(i)} M_T^{(i)}]^{-1} H_{0,T}^{(i)} H_{5,T}^{(i)} \tag{2.196}$$

$$R_{4,T}^{(i)} = \Pi_*^{(i)}[I_{k_i} - H_{0,T}^{(i)} M_T^{(i)}]^{-1} H_{0,T}^{(i)} H_{6,T}^{(i)} \tag{2.197}$$

Finally, let

$$T^{1/2} D_T = T^{-1/2}\sum_{t=1}^{T} \xi_t \tag{2.198}$$

where

$$\xi_t = \begin{bmatrix} g_t^{(1)}(\theta_*^{(i)})z_t^{(1)} - \mu_*^{(1)} \\ g_t^{(1)}(\theta_*^{(1)})^2 vech[z_t^{(1)}z_t^{(1)'} - E[z_t^{(1)}z_t^{(1)'}]] \\ vec\big[D_{zz,t}^{(1)} - D_{zz}^{(1)}\big] \\ \varsigma_t^{(1)}(\theta_*^{(i)}) - E[\varsigma_t^{(1)}(\theta_*^{(i)})] \\ g_t^{(2)}(\theta_*^{(i)})z_t^{(2)} - \mu_*^{(2)} \\ g_t^{(2)}(\theta_*^{(2)})^2 vech[z_t^{(2)}z_t^{(2)'} - E[z_t^{(2)}z_t^{(2)'}]] \\ vec\big[D_{zz,t}^{(2)} - D_{zz}^{(2)}\big] \\ \varsigma_t^{(2)}(\theta_*^{(i)}) - E[\varsigma_t^{(2)}(\theta_*^{(i)})] \end{bmatrix} \tag{2.199}$$

Next, the following assumption is imposed:

Assumption 25 *(i)* $L_T^{'}T^{-1/2}\sum_{t=1}^{T}\xi_t \xrightarrow{d} N(0, L_*^{'}V_*L_*)$,

(ii) $Rank(G_*^{(i)}(\theta_*^{(i)})) = p_i$,

(iii) $T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) = O_p(1)$,

(iv) $\hat{V}_T \xrightarrow{p} V_*$,

(iv) $L_*^{'}V_*L_*$ *is p.s.d.*

where $V_* = lim_{T\to\infty}T^{-1}var(\sum_{t=1}^{T}\xi_t(\theta_*^{(i)}))$.

Given Equations (2.192), (2.199) and Assumption 25, the numerator of B_T converges in distribution to the following:

$$L_*^{'}V^{1/2}\eta_k \tag{2.200}$$

where $\eta_k \sim N(0, I_s)$.

Next, we study the limiting distribution of the denominator of B_T. For that purpose, we define the following:

$$R_T(\theta_*^{(i)}) = L_T \tag{2.201}$$

and

$$V_T = var(T^{1/2}D_T) \tag{2.202}$$

$$= var(T^{-1/2}\sum_{t=1}^{T}\xi_t) \tag{2.203}$$

where ξ_t is given by (1.99).

Then, the limiting distribution of the denominator of B_T is given as:

$$\rho_T \xrightarrow{p}= [L_*{}'V_*L_*]^{1/2} \tag{2.204}$$

where $plim(L_T) = L_*$ and $V_* = var(T^{-1/2}\sum_{t=1}^{T}\xi_t)$.

Equations (2.200) and (2.204), under our assumptions, imply the following limiting distribution of B_T.

Theorem 7 *When Assumptions 1-14 and 25 hold, we have:*

$$B_T \xrightarrow{d} [L_*^{'}V_*L_*]^{-1/2}L_*^{'}V_*^{1/2}\eta_k \tag{2.205}$$

$$= N(0,1) \tag{2.206}$$

2.4 Difference of the NR_T from Pesaran and Smith's GR^2

Pesaran and Smith (1994) propose what they call a Generalized $R^2(GR^2)$ as a model selection statistic, which is given as:

$$GR^2 = 1 - \frac{\hat{e}^{'}\hat{e}}{S_{yy}} \tag{2.207}$$

where $\hat{e}$ is the residual vector from the second stage of the two-stage least squares procedure (TSLS) [19], and $S_{yy} = \sum_{i=1}^{T}(y_t - \bar{y}_T)^2$, where $\bar{y} = T^{-1}\sum_{t=1}^{T} y_t$, y_t is the dependent variable in the structural form equation system given by Equation (2.1) in Pesaran and Smith (1994). The TSLS procedure can be outlined as follows:

(i) In the first step of the TSLS regression X, the matrix of explanatory variables is regressed on the set of instruments as follows:[20]

$$X = Z\hat{\Phi} + \nu \tag{2.208}$$

$$= P_z X + \nu \tag{2.209}$$

where $P_z = Z(Z^{'}Z)Z^{'}$ is the projection vector that linearly projects a variable onto the column space of Z.

---

[19] See Pesaran and Smith (1994, p. 706).

[20] X consists of both exogenous and endogenous regressors, such as past values of the dependent variable.

(ii) In the second step of the TSLS regression, Y is regressed on P_zX from the first stage as follows:

$$Y = P_zX\hat{\delta}_T + \hat{e} \tag{2.210}$$

where the OLS estimator $\hat{\delta}_T$ is given as,

$$\begin{aligned} \hat{\delta}_T &= [(P_zX)^{'}P_zX]^{-1}(P_zX)^{'}Y && (2.211) \\ &= [X^{'}P_zX]^{-1}X^{'}P_zY && (2.212) \end{aligned}$$

Using Equation (2.212) we can compute $\hat{e}$ as follows:

$$\hat{e} = (I - P_zX[X^{'}P_zX]^{-1}X^{'}P_z)Y \tag{2.213}$$

In light of these, we can summarize the differences and similarities between the $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ and Pesaran's GR^2 as follows:

(i) From (2.207) and (2.210), we see that the GR^2 is the adjusted R^2 statistic of the regression given by (2.210), whereas the NR_T, given by (1.11), is the R^2 of the regression given by (1.9). From this, we can see that Y is the analog of $g_t^{(i)}(\hat{\theta}_T^{(i)})$. GR^2 tries to answer how much of the variation in Y is explainable by the variation in P_zX, and the noise ratio tries to explain how much of the variation in $g_t^{(i)}(\hat{\theta}_T^{(i)})$ is due to noise($\hat{g}(\hat{\theta}_T^{(i)})$).

(ii) On the other hand, from the same equations, we can see that $g_t^{(i)}(\hat{\theta}_T^{(i)})$ is neither the usual error from the OLS regression nor the residual from the second stage of the IV regression, but rather, a nonlinear RE error[21] that is a function of the GMM estimator, $\hat{\theta}_T^{(i)}$.

---

[21] An RE error satisfies a conditional moment derived from the optimizing behavior of economic agents.

Chapter 3

Monte-Carlo Simulations for Tests of Non-nested Model Selection

3.1 Simulations

In this chapter, we use Monte-Carlo simulations to study the empirical size and power of the test statistics whose asymptotic distributions were studied in Chapter 1, for comparing non-nested models. The statistic we use is $\mathscr{K}_T$, where:

$$\mathscr{K}_T = T\{NR_T^{(2)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\} \tag{3.1}$$

where $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ is the noise ratio statistic for model i, evaluated at the GMM estimator for model i, which we denote by $\hat{\theta}_T^{(i)}$, where $i = 1, 2$.
Our analysis covers two main cases:

(i) In the first case, we compute the empirical size of our test for comparing two correctly specified models using the $\mathscr{K}_T$ statistic.

(ii) In the second case, we compute the empirical power of our test for comparing two models using the $\mathscr{K}_T$ statistic. The first model under consideration is correctly specified and the second model under consideration is misspecified.

3.1.1 Empirical size and power for hypotheses tests

In all of our analysis, we consider tests constructed using the noise ratio statistic as a goodness of fit measure. Using Monte-Carlo simulations, we compute empirical size and power of the tests using the $\mathscr{K}_T$ statistic. In what follows, we outline the hypotheses covered.

Under the null hypothesis, we assume that the two models, M_1 and M_2, are asymptotically equivalent from a goodness of fit perspective, which we can formally state as follows:

$$NR_*^{(1)}(\theta_*^{(1)}) = NR_*^{(2)}(\theta_*^{(2)}) \tag{3.2}$$

where $NR_*^{(i)}(\theta_*^{(i)})$ denotes the probability limit of the noise ratio for i=1,2.

There are two alternative hypotheses of interest. Under the first alternative hypothesis, we have M_1 is better than M_2, that is,

$$NR_*^{(1)}(\theta_*^{(1)}) < NR_*^{(2)}(\theta_*^{(2)}) \tag{3.3}$$

Under the second alternative hypothesis, we have M_2 is better than M_1, that is,

$$NR_*^{(1)}(\theta_*^{(1)}) > NR_*^{(2)}(\theta_*^{(2)}) \tag{3.4}$$

Under the first case, we consider two correctly specified models with parameter values of our choosing, such that they are asymptotically equivalent from a goodness of fit perspective, to represent the null hypothesis. Under this scenario, the rejection frequency of our test statistic of these correctly specified models, which we compute under our simulation routine, should represent the empirical size of our test.

On the other hand, under Case 2, in order to compute the empirical power of our tests, we construct a scenario under which one of the two models under consideration is correctly specified and the other is misspecified in the non-local sense. We do this by picking up two different sets of instruments for the same model under consideration. The first set of instruments is valid and relevant, whereas, the second set of instruments is relevant but invalid, that is, $E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_{t2}] \neq 0$, where z_{t2} represents the second instrument set. Under the latter scenario, we know that the truth is represented by (3.3), hence the relative simulation rejection frequency represents the empirical power of our tests.

3.1.2 Data generating process and goodness of fit

In all of our analysis, we make assumptions about asymptotic goodness of fit. Here, we measure asymptotic goodness of fit by the limit of the noise ratio statistic evaluated at probability limit of the the GMM estimator, which we denote as $NR_*^{(i)}(\theta_*^{(i)})$, for models $i = 1, 2$. We construct our tests of hypotheses by making different assumptions about this goodness of fit statistic.

The data generating process and the models and instrument sets we consider are due to Hall and Pelletier (2008) and are given as follows:

$$y_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + u_{0,t} \tag{3.5}$$

$$x_{1,t} = \gamma_1 z_{1,t} + \gamma_2 z_{2,t} + \gamma_3 z_{3,t} + \gamma_4 z_{4,t} + \gamma_5 z_{5,t} + \gamma_6 z_{6,t} + u_{1,t} \tag{3.6}$$

$$x_{2,t} = \alpha_1 z_{2,t} + \alpha_2 z_{2,t} + \alpha_3 z_{3,t} + \alpha_4 z_{4,t} + \alpha_5 z_{5,t} + \alpha_6 z_{6,t} + u_{2,t} \tag{3.7}$$

Also, the two models we compare are given as follows:

$$y_t = \tilde{\beta}_1 x_{1,t} + \tilde{u}_{1,t} \tag{3.8}$$

$$y_t = \tilde{\beta}_2 x_{2,t} + \tilde{u}_{2,t} \tag{3.9}$$

Under this scenario, the noise ratio statistic for model i is given by:

$$NR_T^{(i)}(\hat{\beta}_i) = (T^{-1}\sum_{t=1}^{T} \hat{u}_{i,t}(\hat{\beta}_i) z_t^{(i)})' W_T^{(i)}(\hat{\beta}_i)^{-1} (T^{-1}\sum_{t=1}^{T} \hat{u}_{i,t}(\hat{\beta}_i) z_t^{(i)}) \tag{3.10}$$

where

$$W_T^{(i)}(\hat{\beta}_i) = \left[T^{-1}\sum_{t=1}^{T} \hat{u}_{i,t}(\hat{\beta}_i)^2\right]\left[T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}\right] \tag{3.11}$$

where $plim(\hat{\beta}_i) = \tilde{\beta}_i$.

Under this scenario, the limit of the noise ratio statistic for model i can be computed as follows:

$$NR_*^{(i)}(\tilde{\beta}_i) = \mu_*^{(i)}(\tilde{\beta}_i)' W_*^{(i)}(\tilde{\beta}_i)^{-1} \mu_*^{(i)}(\tilde{\beta}_i) \tag{3.12}$$

$$= \mu_*^{(i)}(\tilde{\beta}_i)' \left[\frac{E[z_t^{(i)} z_t^{(i)\prime}]}{E[\tilde{u}_{i,t}(\tilde{\beta}_i)^2]}\right]^{-1} \mu_*^{(i)}(\tilde{\beta}_i) \tag{3.13}$$

where $\mu_*^{(i)}(\tilde{\beta}_i) = E[\tilde{u}_{i,t}(\tilde{\beta}_i)z^{(i)}]$ and $z_t^{(i)}$ represents the instrument set for model i.
Let $z^{(i)} = [z_{1,t}, ..., z_{k_i,t}]$ denote the $k_i \times 1$ instrument set for model $i = 1, 2$. Next, we can compute $\mu_*^{(i)}$ for $i = 1, 2$ as follows:

$$E[\tilde{u}_{1,t}z_t^{(1)}] = E[(y_t - \tilde{\beta}_1 x_{1,t})z_t^{(1)}] \tag{3.14}$$

$$= E[\{\beta_1 x_{1,t} + \beta_2 x_{2,t} + u_{0,t} - \tilde{\beta}_1 x_{1,t}\}z_t^{(1)}] \tag{3.15}$$

$$= E[\{(\beta_1 - \tilde{\beta}_1)x_{1,t} + \beta_2 x_{2,t} + u_{0,t}\}z_t^{(1)}] \tag{3.16}$$

$$\begin{aligned} &= E\Bigg[\Bigg\{(\beta_1 - \tilde{\beta}_1)\Bigg[\sum_{j=1}^{6}\gamma_j z_{j,t} + u_{1,t}\Bigg] \\ &+ \beta_2\Bigg[\sum_{k=1}^{6}\alpha_k z_{j,t} + u_{2,t}\Bigg] + u_{0,t}\Bigg\}z_t^{(1)}\Bigg] \end{aligned} \tag{3.17}$$

$$= (\beta_1 - \tilde{\beta}_1)\Bigg[\sum_{j=1}^{6}\gamma_j E[z_{j,t}z_t^{(1)}]\Bigg] + \beta_2\Bigg[\sum_{k=1}^{6}\alpha_k E[z_{j,t}z_t^{(1)}]\Bigg] \tag{3.18}$$

It also follows that:

$$E[\tilde{u}_{2,t}z_t^{(2)}] = E[(y_t - \tilde{\beta}_2 x_{2,t})z_t^{(2)}] \tag{3.19}$$

$$= E[\{\beta_1 x_{1,t} + \beta_2 x_{2,t} + u_{0,t} - \tilde{\beta}_2 x_{2,t}\}z_t^{(2)}] \tag{3.20}$$

$$= E[\{(\beta_2 - \tilde{\beta}_2)x_{2,t} + \beta_1 x_{1,t} + u_{0,t}\}z_t^{(2)}] \tag{3.21}$$

$$\begin{aligned} &= E\Bigg[\Bigg\{(\beta_2 - \tilde{\beta}_2)\Bigg[\sum_{j=1}^{6}\alpha_j z_{j,t} + u_{2,t}\Bigg] \\ &+ \beta_1\Bigg[\sum_{k=1}^{6}\gamma_k z_{k,t} + u_{1,t}\Bigg] + u_{0,t}\Bigg\}z_t^{(2)}\Bigg] \end{aligned} \tag{3.22}$$

$$= (\beta_2 - \tilde{\beta}_2)\Bigg[\sum_{j=1}^{6}\alpha_j E[z_{j,t}z_t^{(2)}]\Bigg] + \beta_1\Bigg[\sum_{k=1}^{6}\gamma_k E[z_{k,t}z_t^{(2)}]\Bigg] \tag{3.23}$$

Further, we have:

$$
\begin{aligned}
E[\tilde{u}_{1,t}^2] &= E[(y_t - \tilde{\beta}_1 x_{1,t})^2] && (3.24)\\
&= E[(\beta_1 x_{1,t} + \beta_2 x_{2,t} + u_{0,t} - \tilde{\beta}_1 x_{1,t})^2]\\
&= E[(\beta_1 x_{1,t} + \beta_2 x_{2,t} + u_{0,t} - \tilde{\beta}_1 x_{1,t})^2] && (3.25)\\
&= E[\{(\beta_1 - \tilde{\beta}_1)(\gamma_1 z_{1,t} + \gamma_2 z_{2,t} + \gamma_3 z_{3,t} + \gamma_4 z_{4,t} + \gamma_5 z_{5,t} + \gamma_6 z_{6,t} + u_{1,t})\\
&+ \beta_2[\alpha_1 z_{1,t} + \alpha_2 z_{2,t} + \alpha_3 z_{3,t} + \alpha_4 z_{4,t} + \alpha_5 z_{5,t} + \alpha_6 z_{6,t} + u_{2,t}] + u_{0,t}\}^2] && (3.26)\\
&= (\beta_1 - \tilde{\beta}_1)^2 \Bigg(\sum_{j=1}^{6} \gamma_j^2 E[z_{t,j}^2] \Bigg) + \beta_2^2 \Bigg(\sum_{j=1}^{6} \alpha_j^2 E[z_{t,j}^2] \Bigg)\\
&+ 2\beta_2(\beta_1 - \tilde{\beta}_1) \Bigg(\sum_{j=1}^{6} \sum_{k=1}^{6} \gamma_j \alpha_k E[z_{t,j} z_{t,k}] \Bigg) && (3.27)
\end{aligned}
$$

Likewise, we can compute $E[\tilde{u}_{2,t}^2]$ as follows:

$$
\begin{aligned}
E[\tilde{u}_{2,t}^2] &= E[(y_t - \tilde{\beta}_2 x_{2,t})^2] && (3.28)\\
&= E[(\beta_1 x_{1,t} + \beta_2 x_{2,t} + u_{0,t} - \tilde{\beta}_2 x_{2,t})^2]\\
&= E[(\beta_1 x_{1,t} + \beta_2 x_{2,t} + u_{0,t} - \tilde{\beta}_2 x_{2,t})^2] && (3.29)\\
&= E[\{\beta_1[\gamma_1 z_{1,t} + \gamma_2 z_{2,t} + \gamma_3 z_{3,t} + \gamma_4 z_{4,t} + \gamma_5 z_{5,t} + \gamma_6 z_{6,t} + u_{1,t}]\\
&+ (\beta_2 - \tilde{\beta}_2)[\alpha_1 z_{1,t} + \alpha_2 z_{2,t} + \alpha_3 z_{3,t} + \alpha_4 z_{4,t} + \alpha_5 z_{5,t} + \alpha_6 z_{6,t} + u_{2,t}] + u_{0,t}\}^2]\\
& && (3.30)\\
&= \beta_1^2 \Bigg(\sum_{j=1}^{6} \gamma_j^2 E[z_{t,j}^2] \Bigg) + (\beta_2 - \tilde{\beta}_2)^2 \Bigg(\sum_{j=1}^{6} \alpha_j^2 E[z_{t,j}^2] \Bigg)\\
&+ 2\beta_1(\beta_2 - \tilde{\beta}_2) \Bigg(\sum_{j=1}^{6} \sum_{k=1}^{6} \gamma_j \alpha_k E[z_{t,j} z_{t,k}] \Bigg) && (3.31)
\end{aligned}
$$

where, under our assumptions, $E[z_{t,j} z_{t,k}] = 0 \ \forall j \neq k$.

3.1.3 Correctly specified models: Case 1

Under the first case, we consider two correctly specified models and we compute the empirical size of our test for comparing two models using the $\mathscr{K}_T$ statistic. We pick up our parameters of interest, such that the truth is represented by the null hypothesis in (3.2). The variables $z_{1,t}$, $z_{2,t}$ and $z_{3,t}$ will be used for M_1, Equation (3.8), and variables $z_{4,t}$, $z_{5,t}$ and $z_{6,t}$ will be used as instruments for M_2, Equation (3.9).

Also, following Hall and Pelletier (2008), we draw the error terms and instruments independently from an $N(0,1)$ distribution, and assume the following parameter values , $\beta_1 = \beta_2 = 0.5$:

$$\gamma_1 = \gamma_2 = \gamma_3 = 0.5 \tag{3.32}$$

$$\gamma_4 = \gamma_5 = \gamma_6 = 0 \tag{3.33}$$

$$\alpha_1 = \alpha_2 = \alpha_3 = 0 \tag{3.34}$$

$$\alpha_4 = \alpha_5 = \alpha_6 = 0.5 \tag{3.35}$$

Given these parameter values, using Equation (3.17), we can compute the population moments for each model as follows:

$$\begin{aligned}
E[\tilde{u}_{1,t} z_t^{(1)}] &= E[\{(0.5 - \tilde{\beta}_1)(0.5z_{1,t} + 0.5z_{2,t} + 0.5z_{3,t} + u_{1,t}) \\
&\quad - 0.5(0.5z_{4,t} + 0.5z_{5,t} + 0.5z_{6,t} + u_{2,t})\}[z_{1,t}, z_{2,t}, z_{3,t}]] && (3.36) \\
&= E\begin{bmatrix} 0.5(0.5 - \tilde{\beta}_1)[z_{1,t} + z_{2,t} + z_{3,t} + u_{1,t}]z_{1,t} \\ 0.5(0.5 - \tilde{\beta}_1)[z_{1,t} + z_{2,t} + z_{3,t} + u_{1,t}]z_{2,t} \\ 0.5(0.5 - \tilde{\beta}_1)[z_{1,t} + z_{2,t} + z_{3,t} + u_{1,t}]z_{3,t} \end{bmatrix} && (3.37) \\
&\quad + E\begin{bmatrix} 0.25[z_{4,t} + z_{5,t} + z_{6,t} + u_{2,t}]z_{1,t} \\ 0.25[z_{4,t} + z_{5,t} + z_{6,t} + u_{2,t}]z_{2,t} \\ 0.25[z_{4,t} + z_{5,t} + z_{6,t} + u_{2,t}]z_{3,t} \end{bmatrix} && (3.38) \\
&= \begin{bmatrix} 0.5(0.5 - \tilde{\beta}_1)E[z_{1,t}^2] \\ 0.5(0.5 - \tilde{\beta}_1)E[z_{2,t}^2] \\ 0.5(0.5 - \tilde{\beta}_1)E[z_{3,t}^2] \end{bmatrix} && (3.39) \\
&= \begin{bmatrix} 0.5(0.5 - \tilde{\beta}_1) \\ 0.5(0.5 - \tilde{\beta}_1) \\ 0.5(0.5 - \tilde{\beta}_1) \end{bmatrix} && (3.40)
\end{aligned}$$

Likewise, under our assumptions, we can compute the population moment for M_2 as follows:

$$E[\tilde{u}_{2,t}z_t^{(2)}] = \begin{bmatrix} 0.5(0.5-\tilde{\beta}_2) \\ 0.5(0.5-\tilde{\beta}_2) \\ 0.5(0.5-\tilde{\beta}_2) \end{bmatrix} \tag{3.41}$$

Also, in this case we have:

$$E[z_t^{(i)}z_t^{(i)'}] = I_3 \tag{3.42}$$

where I_3 represents the identity matrix with dimension 3 for models $i = 1, 2$. Further, we have:

$$E[\tilde{u}_{1,t}^2] = 0.75(0.5-\tilde{\beta}_1)^2 + 1.5(0.5-\tilde{\beta}_1) + 0.1875 \tag{3.43}$$

$$E[\tilde{u}_{2,t}^2] = 0.75(0.5-\tilde{\beta}_2)^2 + 1.5(0.5-\tilde{\beta}_2) + 0.1875 \tag{3.44}$$

3.1.4 Simulation results: Case 1

In what follows, we provide Monte-Carlo simulation results for the empirical power of our test. The following table summarizes the results of a Monte-Carlo simulation with 4000 replications, and sample sizes $T = 100, 500, 1000$ with $S = 100, 200, 400$ draws from the limiting distribution of $\mathscr{K}_T$ to compute the critical values. Here, the empirical size is given as the two-sided rejection frequency computed in our simulations. As we can see from Table 3.1, the empirical size of the test is very close to the percentile value, 5 %, used for the rejection region constructed from draws from the empirical distribution of the limit of the $\mathscr{K}_T$ test statistic.

3.1.5 Correctly specified and misspecified models: Case 2

In case 2, we compute the empirical power of our test for comparing two models using the $\mathscr{K}_T$ statistic. In order to compute the empirical power of our tests, we construct a scenario under which one of the two models under consideration is correctly specified and the other is misspecified in the non-local sense[1]. We do this by picking up two different sets of instruments for the same model under consideration. The first set of instruments is

---

[1] See Hall and Inoue (2003)[p.366] for a definition of non-local misspecification.

Table 3.1: Empirical Size: Case 1

T	S	$\mathscr{K}_T$	J_T
100	100	0.04775	0.09050
	200	0.03900	0.08550
	400	0.04325	0.09025
500	100	0.05450	0.08375
	200	0.05850	0.09175
	400	0.05275	0.08875
1000	100	0.05575	0.09300
	200	0.05900	0.09850
	400	0.05575	0.09100

[a] Case 1: M_1 and 2 are correctly specified.

[b] These results were generated by 4000 replications.

valid and relevant, whereas, the second set of instruments is relevant but invalid, that is, $E[\tilde{u}_t(\tilde{\beta}_2)z_t^{(2)}] \neq 0$, where $z_t^{(2)}$ represents the second instrument set. Under this scenario, we know that the truth is represented by:

$$NR_*^{(1)}(\theta_*^{(1)}) < NR_*^{(2)}(\theta_*^{(2)}) \tag{3.45}$$

hence, the relative simulation rejection frequency of (3.45) represents the empirical power of our test[2]. In this case, we can assume that M_1 is asymptotically better than M_2, that is, we have:

$$lim_{T\to\infty}T\{NR_*^{(1)}(\theta_*^{(1)}) - NR_*^{(2)}(\theta_*^{(1)})\} = -\infty \tag{3.46}$$

First, we let $\beta_1 = \beta_2 = 0.5$. Also, we use the following set of parameters:

$$\gamma_1 = \gamma_2 = \gamma_3 = 0.5 \tag{3.47}$$

$$\gamma_4 = \gamma_5 = \gamma_6 = 0 \tag{3.48}$$

$$\alpha_1 = \alpha_2 = \alpha_3 = 0 \tag{3.49}$$

$$\alpha_4 = \alpha_5 = \alpha_6 = 2 \tag{3.50}$$

---

[2] The power of a test statistic is defined as the probability of rejecting H_o, given that the alternative hypothesis is true.

The variables $z_{1,t}$, $z_{2,t}$ and $z_{3,t}$ will be used as instruments for M_1, Equation (3.8), and variables $z_{3,t}$, $z_{4,t}$, $z_{5,t}$ will be used as instruments for model 2, Equation (3.9). Further, we increase the variances of instruments $z_{4,t}$, $z_{5,t}$ to 4. Given this data generating process and model choices, we can compute the population moment for Models 1 and 2 as follows:

$$E[\tilde{u}_{2,t}z^{(2)}] = \begin{bmatrix} 0.5(0.5 - \tilde{\beta}_1) \\ 0.5(0.5 - \tilde{\beta}_1) \\ 0.5(0.5 - \tilde{\beta}_1) \end{bmatrix} \tag{3.51}$$

$$E[\tilde{u}_{2,t}z^{(2)}] = \begin{bmatrix} .25 \\ 8(0.5 - \tilde{\beta}_2) \\ 8(0.5 - \tilde{\beta}_2) \end{bmatrix} \tag{3.52}$$

From which, it follows that:

$$\mu_*^{(2)} = \sqrt{128(0.5 - \tilde{\beta}_2)^2 + 0.0625} \tag{3.53}$$

Further, it follows that:

$$E[\tilde{u}_{2,t}^2] = 0.1875 + 36(0.5 - \tilde{\beta}_2)^2 \tag{3.54}$$

Since M_1 is correctly specified, everything remains the same as in Case 1. Since the truth is given by the alternative hypothesis in (3.45), the rejection frequency represents the power of the test. Table 3.2 summarizes the results of Monte-Carlo simulations for 4000 replications, sample sizes $T = 100, 500, 1000$ and $S = 100, 200, 400$.[3] In both Cases 2a and 2b, M_1 is correctly specified, whereas, model 2 is misspecified. However, the degree of misspecification of M_2 is larger in Case 2a. It would be desirable for our test to distinguish a mildly misspecified model, M_2, from a correctly specified model, M_1, in our case. For that purpose,

---

[3] S represents the number of draws from the asymptotic distribution of the $\mathscr{K}_T$ test statistic for the data generating processes outlined in this section.

we modify our data generating process given above as follows:

$$\gamma_1 = \gamma_2 = \gamma_3 = 0.5 \tag{3.55}$$

$$\gamma_4 = \gamma_5 = \gamma_6 = 0 \tag{3.56}$$

$$\alpha_1 = \alpha_2 = \alpha_3 = 0 \tag{3.57}$$

$$\alpha_4 = \alpha_5 = \alpha_6 = 0.5 \tag{3.58}$$

The variables $z_{1,t}$, $z_{2,t}$ and $z_{3,t}$ will be used as instruments for M_1 and variables $z_{3,t}$, $z_{5,t}$, $z_{6,t}$ will be used as instruments for M_2, Equation (3.9). This time, however, all instruments are assumed to be independently drawn from $N(0,1)$. The results of a Monte-Carlo simulation with 4000 replications and 300 draws from the limiting distribution of $\mathscr{K}_T$, with a correctly specified M_1 and a mildly misspecified M_2, for this new data generating process, are given in Table (3.2). As we can see from these results, for a given S, as we increase the sample size from 100 to 1000, the power of the test goes upward toward 1, which is the desirable result.

Table 3.2: Empirical Power: Case 2

		2a		2b	
T	S	$\mathscr{K}_T$	J_T	$\mathscr{K}_T$	J_T
100	100	0.51375	0.29775	0.23925	0.28250
	200	0.50225	0.29550	0.22350	0.28200
	400	0.49200	0.30150	0.21900	0.28100
500	100	0.98725	0.98275	0.94775	0.98225
	200	0.98325	0.98375	0.95375	0.98350
	400	0.98650	0.98075	0.94825	0.98000
1000	100	1	1	0.99875	1
	200	0.99950	1	0.99900	1
	400	1	1	0.99900	1

[a] Case 2: M_1 is correctly specified and model 2 is misspecified.

[b] These results were generated by 4000 replications.

Chapter 4

Production Smoothing and Buffer Stock Model of Inventories

4.1 Introduction

It is established that inventories play a major role in the business cycle and one of the most popular inventory models is the Production Smoothing, Buffer Stock Model of Inventories (PSBSMI). In this model, cost-minimizing firms with convex cost functions, when faced with variable demand will smooth out production. Additionally, in the case of stochastic demand, the firm will use inventories as a buffer stock. However, the Pure Production Smoothing Model of Inventories (PPSMI) has been found to be contrary to data. One potential source of this contradiction is econometric estimation techniques that have been used. Durlauf and Maccini (1995) have used GMM to estimate the inventory models which they compare with the J_T test statistic. When the models under consideration are misspecified, the OIR test statistic can not be used for model comparison. Durlauf and Maccini (1995) also use the Noise Ratio (NR) statistic of Durlauf and Hall (1989), whose asymptotic distribution, when the parameters of the model are estimated by GMM, was studied in Chapter 1 of this dissertation. Also, in Chapter 1, the NR statistic was evaluated in the Rivers and Vuong (2002) framework of model selection. It was seen that there is no unified model selection methodology using the Rivers and Vuong (2002) framework where the goodness of fit measure is the NR statistic. In this chapter, different inventory models are compared using certain statistics that are developed in Chapter 1 of this dissertation.

The particular models compared are the production smoothing model of inventories and its variants, covered in Durlauf and Maccini (1995). The statistics used for comparing these models include $T\{NR_T^{(2)}(\hat{\theta}_T^{(1)}) - NR_T^{(2)}(\hat{\theta}_T^{(2)})\}$ and B_T of Rivers and Vuong (2002). All statistics are evaluated at the GMM estimator $\hat{\theta}^{(i)}$ for the corresponding model $i = 1, 2$.

Inventories play a major role in the business cycle. Most of the drop in US GNP in the average post WWII recession is accounted by he drop in inventories. Macroeconomic evidence suggests that inventories are a destabilizing factor. In one of the first studies about the macroeconomics of inventory investment, Metzler (1941) showed that inventories, when added to simple Keynesian models, can generate cycles through the inventory-accelerator mechanism. Abramovitz (1950) and Lovell (1961) also showed that inventories can destabilize the economy. On the other hand, some microeconomic models of inventories disagree. In particular, the PSBSMI suggest that firms use inventories to smooth production. The model is based on an intuitively simple idea. When faced with variable demand, firms with convex cost functions use inventories to smooth production. If, in addition, the demand is stochastic, the firm would use inventories as a buffer stock. That is, all the demand shock would come solely out of inventories without any change in production. In this paper, we will attempt to reconcile microeconomic theory with facts using different model selection criteria studied in Chapter 1 of this dissertation.

4.1.1 Production smoothing buffer stock model of inventories

The production smoothing buffer stock model of inventories has three implications:

- $Var(Y) > Var(X)$
- $Cov(X, \Delta N) < 0$
- Interest rates and inventory investment are negatively correlated.

where Y is output, X is sales and ΔN is inventory investment.
In most studies it has been found that macroeconomic evidence contradicts all the implications of the PPSMI. First of all, macroeconomic evidence suggests that $Var(Y) > Var(X)$, that is, GDP is more variable than final sales. Further, Blanchard (1983) showed that this

inequality between GDP and final sales holds at the industry level, that is, production is more variable than sales in most industries. However, evidence suggests that sales and inventory investment are positively, not negatively, correlated. Finally, there is little evidence that interest rates and inventory investment are negatively correlated.

All these facts have motivated researchers to come up with scenarios that would reconcile microeconomic theory with macroeconomic data.

First, it has been suggested that the Department of Commerce data might be inaccurate. Also, it has been suggested that it is more appropriate to work with physical unit data. However, Fair (1989), using physical unit data, shows that only a few industries have production less variable than sales and only for short periods of time. Other studies, such as Blanchard (1983), have shown that production is more variable than sales in the automobile industry, even in the case of physical unit data.

Second, Ramey (1991) has suggested that the marginal cost functions might be positively, not negatively, sloped. However, this contradicts the law of supply, one of the main pillars of microeconomic theory. Other authors, in an attempt to save the PSBSMI, have added cost shocks. Cost shocks help in making output more variable than sales and also yield a positive variance between inventory investment and sales. When costs are high, firms lower production and use inventories to satisfy demand and when costs are unexpectedly low, firms raise output and build up inventories to take advantage of low costs. However, even adding cost shocks has not reconciled theory with facts in the case of the PPBSMI. Eichenbaum (1989) showed that the only case in which cost shocks work is when they are unobserved technology shocks. However, even when we allow for such cost shocks, it is not possible to explain facts using the PSBSMI.

4.2 The pure production smoothing model

The simplest version of the PSBSMI has linear demand together with quadratic production and inventory holding costs. Firms that face convex production cost functions minimize the expected present value of total costs. Formally, this optimization is given as[1]:

$$Min\, E_t \sum_{s=0}^{\infty} \beta^s TC_{t+s} \tag{4.1}$$

---

[1] The notation and equations used in this Chapter are taken from Durlauf and Maccini (1995)

where $E_t[.]$ is expectation conditional on information, as of period t, $\beta = 1/(1+r)$ is the discount factor with r as the real interest rate and TC_{t+s} is total production costs as of period $t+s$, which is given as:

$$TC_{t+s} = C(Y_{t+s}) + B(N_{t+s}, X_{t+s}) \tag{4.2}$$

Production costs and inventory holding costs, given in (4.2), are defined as follows:

$$C(Y_{t+s}) = (c_1 + \Gamma_{t+s})Y_{t+s} + \frac{1}{2c}Y_{t+s}^2 \tag{4.3}$$

where Γ_{t+s} represents a cost shock, Y_{t+s} is output at the end of period $t+s$ and c, c_1 are constants. The value of the parameter c is critical to production smoothing. Notice the marginal cost of production equals $(c_1 + \Gamma_{t+s}) + \frac{1}{c}Y_{t+s}$. Therefore, a lower value of c implies a steeper marginal cost function, hence a higher motive to smooth out production.

The inventory holding costs are given as

$$B(N_{t+s}, X_{t+s}) = b_1 N_{t+s} + \frac{b}{2}(N_{t+s} - \alpha X_{t+s})^2 \tag{4.4}$$

where N_{t+s} is the stock of finished goods inventories at the end of period $t+s$, X_{t+s} is real sales and b, b_1 are constants. Further, we assume that $b > 0$, $\alpha > 0$. The parameter b is also critical for production smoothing. A large value of b makes it progressively more costly to increase inventories, hence will give the firm a disincentive to smooth production by changing inventories.

Inventories are accumulated according to the following identity:

$$N_{t+s} - N_{t+s-1} = Y_{t+s} - X_{t+s} \tag{4.5}$$

In what follows, we use lower case letters to denote the conditional expectation of a variable given all available information at period t, i.e. $y_{t+s} = E[Y_{t+s}|\mathscr{F}_t]$.

This optimization problem can be solved as follows:

First, using (4.5), we can re-write (4.2) as:

$$\begin{aligned} TC_{t+s} &= C(N_{t+s} - N_{t+s-1} + X_{t+s}) + B(N_{t+s}, X_{t+s}) && (4.6)\\ &= (c_1 + \Gamma_{t+s})(N_{t+s} - N_{t+s-1} + X_{t+s}) + \frac{1}{2c}(N_{t+s} - N_{t+s-1} + X_{t+s})^2 \\ &+ b_1 N_{t+s} + \frac{b}{2}(N_{t+s} - \alpha X_{t+s})^2 && (4.7) \end{aligned}$$

Now, we can substitute (4.7) in (4.1) and differentiate it with respect to N_{t+s} for all $s = 0, 1, 2, ...$ to get the following Euler equation,

$$\beta n_{t+s+1} - [1 + \beta + bc] n_{t+s} + n_{t+s-1} = (1 - \alpha bc) x_{t+s} - \beta x_{t+s+1} + c\gamma_{t+s} - \beta c \gamma_{t+s+1} + \bar{c} \quad (4.8)$$

for $s = 0, 1, 2, ...$ where $\bar{c}$ is a constant.

The Euler equation basically states that the firm should equate the marginal gain from producing one more unit today instead of tomorrow to the cost of holding that additional unit in inventory.

We can re-write (4.8) in lag operator notation as follows:

$$\begin{aligned} \left[L^2 - \left(\frac{1 + \beta + bc}{\beta}\right) L + \frac{1}{\beta}\right] n_{t+s+1} &= \left(\frac{1 - \alpha bc}{\beta}\right) x_{t+s} - x_{t+s+1} \\ &+ \frac{c}{\beta}\gamma_{t+s} - c\gamma_{t+s+1} + \bar{c} && (4.9) \end{aligned}$$

where L denotes the lag operator, which is defined as $L^n X_t = X_{t-n}$ for $n = ..., -2, -1, 0, 1, 2, ...$ and $\{X_t\}$ is a sequence of real numbers, $t = ..., -2, -1, 0, 1, 2, ...$

Using Equation 2.3.3 of Hamilton (1994), any second degree difference equation can be written as $(1 + \phi_1 L + \phi_2 L^2) y_t = w_t$. We can factor the lag polynomial as follows:

$(1 + \phi_1 L + \phi_2 L^2) = (1 - \lambda_1 L)(1 - \lambda_2 L) = (1 - [\lambda_1 + \lambda_2] L + \lambda_1 \lambda_2 L^2)$, from which we have: $\lambda_1 + \lambda_2 = \phi_1$ and $\lambda_1 \lambda_2 = -\phi_2$.

From (4.9) we get, $\phi_1 = (1 + \beta + bc)/\beta$ and $\phi_2 = 1/\beta$. Therefore, we have $\lambda_1 + \lambda_2 =$

$(1+\beta+bc)/\beta$ and $\lambda_1\lambda_2 = -1/\beta$. Thus, we can re-write (4.9) as follows:

$$\begin{aligned}[(1-\lambda L)(1-\beta^{-1}\lambda^{-1}L)]n_{t+s+1} = {} & \left(\frac{1-\alpha bc}{\beta}\right)x_{t+s} - x_{t+s+1} \\ & + \frac{c}{\beta}\gamma_{t+s} - c\gamma_{t+s+1} + \bar{c}\end{aligned} \tag{4.10}$$

Solving the characteristic equation, we get its stable root, λ, as,

$$\lambda = 1 + \frac{1}{2}\left(r + \frac{bc}{\beta}\right) - \left[\left(r + \frac{bc}{\beta}\right)^2 + \frac{4bc}{\beta}\right]^{1/2} \tag{4.11}$$

Given the assumptions about the parameters of the model, $0 < \lambda < 1$.

The expectational difference equation (4.10) can be solved to yield the following solution for the optimal level of inventories:

$$\begin{aligned}n_{t+s} = {} & \lambda_{n+s-1} + [1-(1-\alpha bc)\lambda]\sum_{i=0}^{\infty}(\beta\lambda)^i x_{t+s+i} \\ & - x_{t+s} + c[1-\lambda]\sum_{i=0}^{\infty}(\beta\lambda)^i\gamma_{t+s+i} - c\gamma_{t+s} + \bar{d}\end{aligned} \tag{4.12}$$

where $\bar{d}$ is a constant.

Next, we derive a relationship for actual inventory accumulation. Setting $s = 0$ and given $n_{t_1} = N_{t-1}$, we can re-write Equation (4.12) as follows:

$$\begin{aligned}n_t - N_{t-1} = {} & (1+\lambda)N_{t-1} + [1-(1-\alpha bc)\lambda]\sum_{i=0}^{\infty}(\beta\lambda)^i x_{t+i} \\ & - x_t + c[1-\lambda]\sum_{i=0}^{\infty}(\beta\lambda)^i\gamma_{t+i} - c\gamma_t + \bar{d}\end{aligned} \tag{4.13}$$

Taking expectations of both sides of Equation (4.5) we have:

$$E[N_t - N_{t-1}] = n_t - Nt - 1 \tag{4.14}$$

$$= E[Y_t - X_t] \tag{4.15}$$

$$= Y_t - x_t \tag{4.16}$$

where, under our assumptions $E[Y_t] = Y_t$.

Again, using the inventory identity at $s = 0$, we get

$$\begin{aligned} N_t - N_{t-1} &= Y_t - X_t && (4.17)\\ &= Y_t - x_t + x_t - X_t && (4.18)\\ &= N_t - n_{t-1} + x_t - X_t && (4.19)\end{aligned}$$

where the last line follows from (4.14).

Finally, substituting (4.13) in (4.19) and rearranging terms, the actual inventory accumulation is given by:

$$\begin{aligned} N_t &= \lambda N_{t-1} - X_t + [1 - \alpha bc)\lambda]x_t \\ &+ [1 - (1 - \alpha bc)\lambda] \sum_{i=1}^{\infty} (\beta\lambda)^i x_{t+i} \\ &- c\lambda\gamma_t + c[1 - \lambda] \sum_{i=0}^{\infty} (\beta\lambda)^i \gamma_{t+i} + \bar{e} \qquad (4.20)\end{aligned}$$

where $\bar{e}$ denotes a constant.

To derive g_{t+1}, we define the following data-based variable:

$$\Delta_t = N_t - \lambda N_{t-1} + (1 - \alpha bc)\lambda X_t + c\lambda\Gamma_t \qquad (4.21)$$

Assuming that agents observe X_t and Γ_t at the time inventories are chosen,[2] we can use (4.20) to re-write (4.21) as follows:

$$\Delta_t = [1 - (1 - \alpha bc)\lambda] \sum_{i=1}^{\infty} (\beta\lambda)^i x_{t+i} + c[1 - \lambda] \sum_{i=1}^{\infty} (\beta\lambda)^i \gamma_{t+i} + \bar{d} \qquad (4.22)$$

We note that this variable is purely based on conditional expectations. We will denote this variable as follows:

$$\Delta_t^e = [1 - (1 - \alpha bc)\lambda] \sum_{i=1}^{\infty} (\beta\lambda)^i x_{t+i} + c[1 - \lambda] \sum_{i=1}^{\infty} (\beta\lambda)^i \gamma_{t+i} \qquad (4.23)$$

---

[2] This implies $x_t = X_t$ and $\gamma_t = \Gamma_t$, where lower case letters denote conditional expectations of the corresponding uppercase variables, as of time t.

Of course, no model is a perfect reflection of reality and there is a cost to simplifying reality to a set of mathematical equations. This cost is called model noise and is given by $\Delta_t - \Delta_t^e$. In this context the variable $\bar{d}$ gives the size of misspecification of a model[3]. It measures how far our data Δ_t is from what our model says it is equal to, namely Δ_t^e. Hence, we say that our model is correctly specified when $\Delta_t - \Delta_t^e$ equals to zero.

Since (4.23) is a purely conditional expectations based variable, we can define a perfect foresight counterpart as follows:[4]

$$\Delta_t^* = [1-(1-\alpha bc)\lambda]\sum_{i=1}^{\infty}(\beta\lambda)^i X_{t+i} + c[1-\lambda]\sum_{i=1}^{\infty}(\beta\lambda)^i \Gamma_{t+i} + \bar{d} \tag{4.24}$$

where we used Δ_t^* to denote the perfect foresight analog of Δ_t. Further under RE[5] we have the following:

$$\Delta_t^* = [1-(1-\alpha bc)\lambda]\sum_{i=1}^{\infty}(\beta\lambda)^i x_{t+i} + c[1-\lambda]\sum_{i=1}^{\infty}(\beta\lambda)^i \gamma_{t+i} + \eta_t \tag{4.25}$$

where η_t is a RE error term.

Durlauf and Maccini (1995) use $(1-\beta\lambda L^{-1})$[6] as a filter to make $\Delta_t - \Delta_t^*$ white noise.

$$\begin{aligned} g_{t+1} &= (1-\beta\lambda L^{-1})(\Delta_t - \Delta_t^*) && (4.26)\\ &= \Delta_t - \beta\lambda\Delta_{t+1} - (1-\beta\lambda L^{-1})\Delta_t^* && (4.27)\\ &= -\beta\lambda N_{t+1} + (1+\beta\lambda^2)N_t - \lambda N_{t-1} - \beta\lambda X_{t+1} \\ &\quad + (1-\alpha bc)\lambda X_t - c\beta\lambda\Gamma_{t+1} + c\lambda\Gamma_t + k && (4.28) \end{aligned}$$

---

[3] Misspecification is also called noise.

[4] In this context, perfect foresight means that firms face no uncertainty when making inventory decisions, about all future values of sales and cost shocks. This will allow us to replace conditional expectations with actual values.

[5] RE implies that the difference between the conditional expectation of a variable and its perfect foresight counterpart is an error term whose conditional expectation is zero, which in our context, is given as $\Delta_t^* - \Delta_t^e = \eta_t$, where $E_t[\eta_t] = 0$.

[6] L denotes the backshift (lag) operator.

where $k = -(1-\beta\lambda)\bar{d}$[7].

In the following, we show how to derive (4.28) from (4.27):

Using (4.23), we can write the following:

$$\Delta_t^* = [1-(1-\alpha bc)\lambda]\sum_{i=1}^{\infty}(\beta\lambda)^i X_{t+i} + c[1-\lambda]\sum_{i=1}^{\infty}(\beta\lambda)^i \Gamma_{t+i} + \bar{d} \tag{4.32}$$

$$= [1-(1-\alpha bc)\lambda]\left(\frac{\beta\lambda L^{-1}}{1-\beta\lambda L^{-1}}\right)X_t + c[1-\lambda]\left(\frac{\beta\lambda L^{-1}}{1-\beta\lambda L^{-1}}\right)\Gamma_t \tag{4.33}$$

where (4.33) results from (4.32) by the following fact:[8]

$$\frac{\beta\lambda L^{-1}}{1-\beta\lambda L^{-1}} = \sum_{i=1}^{\infty}(\beta\lambda)^i L^{-i} \tag{4.34}$$

given $\beta < 1$ and $0 < \lambda < 1$ implying $\beta\lambda < 1$.

Next, we multiply (4.33) through by $1-\beta\lambda L^{-1}$ to get the desired result.

$$(1-\beta\lambda L^{-1})\Delta_t^* = (1-\beta\lambda L^{-1})[[1-(1-\alpha bc)\lambda]\left(\frac{\beta\lambda L^{-1}}{1-\beta\lambda L^{-1}}\right)X_t$$

$$+ c[1-\lambda]\left(\frac{\beta\lambda L^{-1}}{1-\beta\lambda L^{-1}}\right)\Gamma_t] \tag{4.35}$$

$$= [1-(1-\lambda\alpha bc)]\beta\lambda X_{t+1} - c(1-\lambda)\beta\lambda\Gamma_{t+1} + k \tag{4.36}$$

Also, using (4.21) for periods t and $t+1$, we can write the following:

$$\Delta_t - \beta\lambda\Delta_{t+1} = N_t - \lambda N_{t-1} + (1-\alpha bc)\lambda)X_t + c\lambda\Gamma_t$$

$$- \beta\lambda N_{t+1} + \beta\lambda^2 N_t - (1-\alpha bc)\lambda^2\beta X_{t+1} - c\beta\lambda^2\Gamma_{t+1} \tag{4.37}$$

$$= -\beta\lambda N_{t+1} + (1+\beta\lambda^2)N_t - \lambda N_{t-1}(1-\alpha bc)\lambda)X_t$$

$$+ c\lambda\Gamma_t - (1-\alpha bc)\lambda^2\beta X_{t+1} - c\beta\lambda^2\Gamma_{t+1} \tag{4.38}$$

---

[7] Notice from (4.22) and (4.24) that

$$(1-\beta\lambda L^{-1})(\Delta_t - \Delta_t^*) = (1-\beta\lambda L^{-1})[\bar{d}_t - \eta_t] \tag{4.29}$$

$$= (1-\beta\lambda L^{-1})\bar{d}_t - (1-\beta\lambda L^{-1})\eta_t \tag{4.30}$$

$$= S_t + \nu_t \tag{4.31}$$

hence, we have $S_t = (1-\beta\lambda L^{-1})\bar{d}_t$ and $\nu_t = -(1-\beta\lambda L^{-1})\eta_t$.

[8] For some real number a, for which we have $|a| < 1$, $\sum_{i=0}^{\infty} a^i = 1/(1-a)$.

Finally, substituting (4.36) and (4.38) into (4.27), we get the what Durlauf and Maccini (1995) call a Euler equation forecast error as:

$$\begin{aligned} g_{t+1} = & -\beta\lambda N_{t+1} + (1+\beta\lambda^2)N_t - \lambda N_{t-1} - \beta\lambda X_{t+1} \\ & + (1-\alpha bc)\lambda X_t - c\beta\lambda\Gamma_{t+1} + c\lambda\Gamma_t + k \end{aligned} \tag{4.39}$$

4.3 Data

The data used in this paper are the seasonally adjusted chained goods sales and inventories data published by the Bureau of Economic Analysis. We use monthly data that covers the period, 1967:01-2008-09.

4.4 Econometric methodology

4.4.1 Noise ratio statistic

The following is from Durlauf and Hall (1989):
"Under rather general conditions, observed covariances place a useful lower bound on the variance of misspecification, or noise, in models based on expectations. For a correctly specified model, the lower bound will be zero. We construct an optimal bound on model noise that captures the complete set of testable restrictions on an expectations based model."
The noise ratio is a tool for measuring the extent to which a model explains the movements in data. Technically, it is given as the following ratio:

$$NR = \frac{var(S_{t|t})}{var(g_{t+1})} \tag{4.40}$$

where $var(.)$ is the variance of the associated variable, $S_{t|t}$ is an estimate of the noise variable, which we will denote by S_t, and g_{t+1} is what we will call a Euler equation forecast error.
Durlauf and Maccini (1995) use the signal extraction methodology developed by Durlauf and Hall (1989), where using GMM, an estimate of unobservable model noise, $g_{t+1}(\hat{\theta}_T^{(i)})$, and the associated noise ratio statistic, are computed from data as follows:

- First, we derive what is called a Euler equation forecast error, which will be denoted as g_{t+1}.
- Second, using GMM, estimates of the parameters in g_{t+1}, which we denote by $\hat{\theta}_T^{(i)}$, are computed. Then, using these parameter estimates and data, an estimate for g_{t+1} and an estimate for its associated variance is computed.
- Third, g_{t+1} is projected onto the information set of the econometrician using different sets of instruments in order to get an estimate of noise. This estimate of the noise variable will be denoted by $S_{t|t}$. Further, we compute an estimate for the variance of $S_{t|t}$.
- Finally, the noise ratio is computed as: $NR = var(S_{t|t})/var(g_{t+1})$.

4.4.2 GMM estimation of parameters, g_{t+1} and its associated variance

Now, under the null hypothesis of correct specification, we have:

$$Ho : g_{t+1} = \nu_{t+1} \tag{4.41}$$

where ν_{t+1} is an RE error, for which we have $E_t[\nu_{t+1}] = 0$. We can use g_{t+1} in the following moment condition to estimate the parameters of (4.39), namely $\mu = 1 - \alpha bc$, λ and c, using GMM.

$$E[g_{t+1} Z_t] = 0 \tag{4.42}$$

where z_t is a set of instruments from the information set of agents.

Once we have estimates of the parameters, we can compute an estimate for g_{t+1} using (4.39). Further, we can compute an estimate of the the variance of g_{t+1}.

4.4.3 Estimating noise and its associated variance

When the model is not correctly specified, there will be a noise component to g_{t+1}. Under such an alternative hypothesis, we would have:

$$Ho : g_{t+1} = \nu_{t+1} + S_t \tag{4.43}$$

where S_t is model noise (or misspecification).

We can get an estimate of S_t, which we will denote by $S_{t|t}$, by linearly projecting g_{t+1} onto the econometrician's information set as follows:

$$\begin{aligned} S_{t|t} &= proj(z_t)g_{t+1} && (4.44) \\ &= proj(z_t)(\nu_{t+1} + S_t) && (4.45) \\ &= proj(Z_t)\nu_{t+1} + proj(z_t)S_t && (4.46) \\ &= proj(z_t)(S_t) && (4.47) \end{aligned}$$

where $proj(z_t)$ is an operator which linearly projects onto the econometrician's information set. Equation (4.47) follows from (4.45), because v_{t+1} is orthogonal to any variable in the information set of the econometrician by construction, hence, $proj(Z_t)\nu_t = 0.^9$.

We can use the following to compute the estimates of S_t for periods $t = 1, 2, ..., T$:

$$S_{t|t} = z_t^{'}(Z^{'}Z)^{-1}Z^{'}g_{t+1} \tag{4.48}$$

where Z is the $T \times k$ matrix of instruments with t^{th} row $z_t^{'}$ and k is the number of instruments. Once we have $S_{t|t}$, we can compute an estimate for its variance.

Finally, the noise ratio can be computed as follows:

$$NR = \frac{var(S_{t|t})}{var(g_{t+1})} \tag{4.49}$$

This statistic tells us how much of the variability in g_{t+1} is due to variability in the estimate of noise. Given that noise is a measure of how much the model explains data, this statistic also measures how well our model explains movements in data.

4.5 Extensions of the PPSMI

4.5.1 Current sales and cost shocks are unobservable

This case is more realistic, in that many firms do not know the current sales and cost shocks at the time they make production decisions. Hence, there is a role for inventories to buffer

---

[9] This projection can be computed by a regression of g_{t+1} onto Δ_t.

production from current sales and cost shocks.

If we solve the model under the assumption that current sales and cost shocks are unobservable, we get the following Euler equation forecast error:

$$\begin{aligned} g_{t+1} = &-\beta\lambda N_{t+1} + (1+\beta\lambda^2)N_t \\ &-\lambda N_{t-1} - \beta\lambda X_{t+1} + (1-\alpha bc)\lambda X_t \\ &-c\beta\lambda\Gamma_{t+1} + c\lambda\Gamma_t + k \end{aligned} \tag{4.50}$$

where m is an unimportant constant. As we can see, this again has exactly the same form as when both current sales and costs shocks were observable. However, this case is different. Under this case, the information set can not include current sales and cost shocks. Moreover, given the inventory accumulation equation, current inventories are also unknown and can not be included in the information set. For this reason, we have to modify our projection, hence the noise ratio as follows:

$$S_{t|t-1} = proj(Z_{t-1})g_{t+1} \tag{4.51}$$

and our noise ratio becomes

$$NR = \frac{var(S_{t|t-1})}{var(g_{t+1})} \tag{4.52}$$

4.6 Econometric results

The data used in this paper are the seasonally adjusted chained goods sales and inventories data published by the Bureau of Economic Analysis. The data are at monthly frequency and cover the period, 1967:01-2008-09. Following Durlauf and Maccini (1995), we detrend our series using a quadratic time trend. We use two different instrument sets in our estimations. The first instrument set includes current values of inventories and sales for each industry as well as three or less lags of these variables. The second instrument set excludes current values of these variables, which captures the idea that inventories buffer sales surprises.

Next, we report estimates for the parameters of different formulations of the PPSMI for 5 different industries, using the GMM estimator. We also report the estimates of the noise ratio and OIR statistics. Finally, we compare different formulations of the PPSMI, using the $\mathscr{K}_T$ statistic, whose theory, in the case of the GMM estimator, was studied in Chapter

1 of this dissertation.

4.6.1 Pure production smoothing model of inventories

In Table 4.1, we report results for the PPSMI for different industries. Specifically, we set $\beta = 0.995$, $\alpha = 0$, or equivalently, $\mu = 1$, and estimate λ using the GMM. Also, we report the values of the noise ratio, OIR and $\mathscr{K}_T$ test statistics for each industry.

As Table 4.1 illustrates, the J_T tets rejects PPSMI with instrument set 1. On the other hand, the PPSMI performs much better with the second instrument set. The J test statistic rejects mildly only the chemicals industry with instrument set 2, at the 5 percent significance level. These findings are also corroborated by the $\mathscr{K}_T$ test statistic. As we can see from the associated p values, the statistic chooses PPSMI with instrument set 2 over instrument set 1. However, the values of λ for the model with instrument set 2 are quite high. In this model, $1-\lambda$ is a measure of the adjustment speed of inventories toward desired levels. Hence, the values in Table 1 imply unusually low adjustment speeds.

Table 4.1: Pure production smoothing model of inventories.

Industry	Instrument	λ	NR_T	J_T	$\mathscr{K}_T$ (p-value)
Petroleum	Z1	0.682	0.486	111.592	241.27
	Z2	0.764	0.019	9.215	(0)
Chemicals	Z1	0.778	0.246	55.128	104.52
	Z2	0.835	0.036	15.062	(0)
Food	Z1	0.786	0.240	42.217	99.45
	Z2	0.857	0.040	9.960	(0)
Rubber	Z1	0.803	0.221	66.610	101.64
	Z2	0.930	0.016	8.692	(0)

[a] PPSMI occurs when $\alpha = 0$ and there are no cost shocks.

[b] Instrument set 1: $[N_t, N_{t-1}, N_{t-2}, N_{t-3}, X_t, X_{t-1}, X_{t-2}, X_{t-3}, constant]$
Instrument set 2: $[N_{t-1}, N_{t-2}, N_{t-3}, X_{t-1}, X_{t-2}, X_{t-3}, constant]$.

Next, in Table 4.2, we report the results for the stock out avoidance model of inventories. Stock out avoidance occurs when $\alpha \neq 0$. In this case, we estimate λ and $\mu = 1 - \alpha bc$ using GMM estimation. We use the same two sets of instruments that we used for the PPSMI. Again, we see that the J_T statistics all reject the formulations of the stock out avoidance model with instrument set one. However, all formulations of the model pass the J_T test with instrument set 2. Likewise, the values of the $\mathscr{K}_T$ statistic unanimously reject the model with instrument set 1 in favor of the model with instrument set 2. Again, the values of λ imply unusually low adjustment speeds for inventories to their desired levels.

4.7 Conclusion

All the results in this paper more or less confirm the findings of Durlauf and Maccini (1995). However, this paper has a more sound theoretical foundation, as it uses econometric results that we derived in Chapter 1, for comparing possibly misspecified models and hence is an improvement in that regard. The J statistic can only be used to see whether a model is misspecified or not. Further, contrary to its usage by Durlauf and Maccini (1995), the noise ratio by itself is not a metric to measure how far a model is from the truth. When the models under consideration are correctly specified, the noise ratio converges to zero. However, when the models under consideration are misspecified in the non-local sense, the noise ratio for a model converges to a constant. On the other hand, when correctly scaled, the noise ratio or model selection statistics constructed from it can be used as valid model selection tools for comparing both correctly or non-locally misspecified models in a similar fashion to the Rivers and Vuong (2002).

Table 4.2: Stock out avoidance model of inventories.

Industry	Instrument	λ	μ	NR_T	J_T	$\mathscr{K}_T$ (p-value)
Petroleum	Z1	0.695	0.968	0.478	110.455	232.05
	Z2	0.832	0.937	0.012	4.550	(0)
Chemicals	Z1	0.776	0.969	0.236	54.835	103.72
	Z2	0.828	0.965	0.028	11.666	(0)
Food	Z1	0.800	0.884	0.203	31.106	90.45
	Z2	0.847	0.905	0.021	3.629	(0)
Rubber	Z1	0.773	0.940	0.214	65.998	103.03
	Z2	0.898	0.952	0.006	2.657	(0)

[a] $\alpha \neq 0$, implies stockout avoidance.

[b] Instrument set 1: $[N_t, N_{t-1}, N_{t-2}, N_{t-3}, X_t, X_{t-1}, X_{t-2}, X_{t-3}, constant]$.
Instrument set 2: $[N_{t-1}, N_{t-2}, N_{t-3}, X_{t-1}, X_{t-2}, X_{t-3}, constant]$.

BIBLIOGRAPHY

ABRAMOVITZ, M. (1950): *Inventories and business cycles : with special reference to manufacturers inventories.* National Bureau of Economic Research, New York.

BLANCHARD, O. J. (1983): "The Production and Inventory Behavior of the American Automobile Industry," *The Journal of Political Economy*, 91(3), 365–400.

DHRYMES, P. J. (2000): *Mathemeatics for Econometrics.* Springer, 3 edn.

DURLAUF, S. N., AND R. E. HALL (1989): "Bounds on the Variances of Specification Errors in Models with Ex- pectations," NBER Working Papers 2936, National Bureau of Economic Research, Inc.

DURLAUF, S. N., AND L. J. MACCINI (1995): "Measuring noise in inventory models," *Journal of Monetary Economics*, 36(1), 65–89.

EICHENBAUM, M. (1989): "Some Empirical Evidence on the Production Level and Production Cost Smoothing Models of Inventory Investment," *American Economic Review*, 79(4), 853–64.

FAIR, R. C. (1989): "The production-smoothing model is alive and well," *Journal of Monetary Economics*, 24(3), 353–370.

HALL, A. R. (2005): *Generalized Method of Moments (Advanced Texts in Econometrics).* Oxford University Press, USA.

HALL, A. R., AND A. INOUE (2003): "The large sample behaviour of the generalized method of moments estimator in misspecified models," *Journal of Econometrics*, 114(2), 361–394.

Hall, A. R., and D. Pelletier (2008): "Non-Nested Testing in Models Estimated via Generalized Method of Moments," Working Paper Series 011, North Carolina State University, Department of Economics.

Hamilton, J. D. (1994): *Time Series Analysis.* Princeton University Press.

Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14(3), 262–280.

Lovell, M. (1961): "Manufacturers' Inventories, Sales Expectations, and the Acceleration Principle," *Econometrica*, 29(3), 293–314.

Lütkepohl, H. (1993): *Introduction to Multiple Time Series Analysis.* Springer, 2 edn.

Metzler, L. A. (1941): "The Nature and Stability of Inventory Cycles," *The Review of Economics and Statistics*, 23(3), 113–129.

Pesaran, M. H., and R. J. Smith (1994): "A Generalized R^2 Criterion for Regression Models Estimated by the Instrumental Variables Method," *Econometrica*, 62(3), 705–10.

Ramey, V. A. (1991): "Nonconvex Costs and the Behavior of Inventories," *The Journal of Political Economy*, 99(2), 306–334.

Rivers, D., and Q. Vuong (2002): "Model selection tests for nonlinear dynamic models," *Econometrics Journal*, 5(1), 1–39.

White, H. (2001): *Asymptotic Theory for Econometricians : Revised Edition (Economic Theory, Econometrics and Mathematical Economics).* Academic Press.

Wooldridge, J. M. (1994): "Estimation and inference for dependent processes," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4 of *Handbook of Econometrics*, chap. 45, pp. 2639–2738. Elsevier.

APPENDICES

In this appendix, we analyze the asymptotic behavior of the $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ under the case when $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{gi}^2$.

Formally, we make the following assumptions:

Assumption 26 $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ *is an m.d.s. (We have* $E[g_t^{(i)}(\theta_*)|\Omega_t] = 0$*).*

Assumption 27 $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t^{(i)}) = \sigma_{gi}^2 < \infty$.

Next, we analyze the asymptotic behavior of $T^{-1}J_T^{(i)}(\hat{\theta}_T^{(i)})$ and $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ under the case when $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{gi}^2$.

Once we assume $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ..., T\}$ is m.d.s, $\sigma_{gi}^2 < \infty$ and constant, and assume a certain form for $\hat{S}_T^{(i)}$, it follows that $T^{-1}J_T^{(i)} \equiv NR_T^{(i)}$. Hence, once we derive $plim(T^{-1}J_T^{(i)})$, we also have $plim(NR_T)$.

First, we analyze the probability limit of $T^{-1}J_T^{(i)}$[10].

Theorem 8 *Under Assumptions 1-9 we have:*

$$plim(T^{-1}J_T^{(i)}) = 0 \tag{A.1}$$

Proof:

$$\begin{aligned} plim(T^{-1}J_T^{(i)}) &= (\sigma_{gi}^2)^{-1}\psi_*^{(i)\prime}\zeta^{(i)-1}\psi_*^{(i)} && \text{(A.2)} \\ &= \psi_*^{(i)\prime}(\sigma_{gi}^2\zeta^{(i)} - 1)\psi_*^{(i)} && \text{(A.3)} \\ &= 0 && \text{(A.4)} \end{aligned}$$

where

$$\psi_*^{(i)} = E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}] \tag{A.5}$$

where $\zeta^{(i)} = E[z_t^{(i)}z_t^{(i)\prime}]$ and $E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}] = 0$ when the model under consideration is correctly specified.

From Theorem 8, we can see that the limiting distribution of $T^{-1}J_T^{(i)}$ is degenerate in the current case[11]. Next, we show why Equation (A.4) holds. From, Equation(A.6) and

---

[10] $J_T^{(i)} = TQ_T$, where Q_T is the GMM minimand.

[11] Convergence in probability implies convergence in distribution.

Slutsky's Theorem, we can write this probability limit as follows:

$$plim(T^{-1}J_T^{(i)}) = plim\big(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})z_t\big)'\hat{S}_T^{-1}\big(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})z_t\big) \tag{A.6}$$

$$= \Big(plim\big(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})z_t\big)'\Big)plim\big(\hat{S}_T^{-1}\big)\Big(plim\big(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})z_t\big)\Big) \tag{A.7}$$

Next, we analyze the limiting behavior of the components of (A.7). Hence, we begin with $m_T(v_t, \hat{\theta}_T) = T^{-1}\sum_{t=1}^{T} g_t(v_t, \hat{\theta}_T)z_t$. To simplify the notation, we suppress the dependence of m_T on v_t. It is important to recognize the dependence of $m_T^{(i)}(\hat{\theta}_T^{(i)})$ on the GMM estimator $\hat{\theta}_T^{(i)}$. Given that $g_t^{(i)}(\hat{\theta}_T^{(i)})$ depends on $\hat{\theta}_T^{(i)}$, a random vector, we can not apply WLLNs and CLTs directly. This will prevent us from directly applying WLLN's to $m_T^{(i)}(\hat{\theta}_T^{(i)})$. To circumvent this, we will follow a two stage strategy to derive the limiting behavior of the sample moment $m_T(.)$:

(i) in the first stage, we show that $m_T^{(i)}(\hat{\theta}_T^{(i)}) - m_T^{(i)}(\theta_*^{(i)}) = o_p(1)$, which we state in Lemma 2,

(ii) in the second stage, we show that the Ergodic Theorem implies $m_T^{(i)}(\theta_*^{(i)}) \xrightarrow{p} E[m_T^{(i)}(\theta_*^{(i)})]$, which is stated in Lemma 3.

We begin by the following lemma:

Lemma 2 *If Assumptions 1-10 and 27 hold, it follows that:*

$$T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})z_t \xrightarrow{p} T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)} \tag{A.8}$$

A mean value expansion of $m_T^{(i)}(\hat{\theta}_T^{(i)})$ around $\theta_*^{(i)}$ gives:

$$m_T^{(i)}(\hat{\theta}_T^{(i)}) = m_T^{(i)}(\theta_*^{(i)}) + G_T^{(i)}(\bar{\theta}_T^{(i)})(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{A.9}$$

$$= m_T^{(i)}(\theta_*^{(i)}) + o_p(1) \tag{A.10}$$

In what follows, we prove that (A.10) is indeed correct. Using the Cauchy-Schwartz inequality, we can write the following:

$$\|G_T^{(i)}(\bar{\theta}_T^{(i)})(\hat{\theta}_T^{(i)} - \theta_*^{(i)})\| \leq \|G_T^{(i)}(\bar{\theta}_T^{(i)})\|\|(\hat{\theta}_T^{(i)} - \theta_*^{(i)})\| \tag{A.11}$$

From Lemma 1, we have $(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) = o_p(1)$. Therefore, we only need to show $G_T^{(i)}(\bar{\theta}_T^{(i)}) = O_p(1)$. Given $G_*^{(i)}(\theta_*^{(i)}) = O_p(1)$, it suffices to show that $G_T^{(i)}(\bar{\theta}_T^{(i)}) - G_*^{(i)}(\theta_*^{(i)}) = o_p(1)$. Adding and subtracting terms yields the following:

$$\begin{aligned} G_T^{(i)}(\bar{\theta}_T^{(i)}) - G_*^{(i)}(\theta_*^{(i)}) &= G_T^{(i)}(\bar{\theta}_T^{(i)}) - G_T^{(i)}(\hat{\theta}_T^{(i)}) \\ &+ G_T^{(i)}(\hat{\theta}_T^{(i)}) - G_*(\hat{\theta}_T^{(i)}) \\ &+ G_*(\hat{\theta}_T^{(i)}) - G_*^{(i)}(\theta_*^{(i)}) \end{aligned} \tag{A.12}$$

By the continuity of $g_t^{(i)}(\theta^{(i)})$ in Assumption 8 and the facts that $\hat{\theta}_T \xrightarrow{p} \theta_*$ and $\bar{\theta}_T = \lambda\hat{\theta}_T + (1-\lambda)\theta_*$, we have:

$$G_T^{(i)}(\bar{\theta}_T^{(i)}) - G_T^{(i)}(\hat{\theta}_T^{(i)}) = o_p(1) \tag{A.13}$$

From (A.13), it follows that $\left\|G_T^{(i)}(\bar{\theta}_T^{(i)}) - G_T^{(i)}(\hat{\theta}_T^{(i)})\right\| = o_p(1)$. Hence, using (A.13) and the triangle inequality, we have:

$$\begin{aligned} \left\|G_T^{(i)}(\bar{\theta}_T^{(i)}) - G_*^{(i)}(\theta_*^{(i)})\right\| &\leq \left\|G_T^{(i)}(\hat{\theta}_T^{(i)}) - G_*(\hat{\theta}_T^{(i)})\right\| \\ &+ \left\|G_*(\hat{\theta}_T^{(i)}) - G_*^{(i)}(\theta_*^{(i)})\right\| + o_p(1) \end{aligned} \tag{A.14}$$

Since $\hat{\theta}_T \xrightarrow{p} \theta_*$, that is $lim_{T\to\infty}P(\hat{\theta}_T \in N_\epsilon)$, we can restrict our attention to N_ϵ[12]. By definition of the supremum, we have:

$$\left\|G_T^{(i)}(\hat{\theta}_T^{(i)}) - G_*(\hat{\theta}_T^{(i)})\right\| \leq sup_{\theta\in N_\epsilon}\left\|g_t^{(i)}(\theta^{(i)}) - G_*(\theta)\right\| \tag{A.15}$$

By Assumption 8 part (iii), $sup_{\theta\in N_\epsilon}\left\|g_t^{(i)}(\theta^{(i)}) - G_*(\theta)\right\| = o_p(1)$. Hence, it follows from (A.15) that:

$$\left\|G_T^{(i)}(\hat{\theta}_T^{(i)}) - G_*(\hat{\theta}_T^{(i)})\right\| = o_p(1) \tag{A.16}$$

---

[12] N_ϵ denotes an ϵ neighborhood of $\theta_*^{(i)}$ for any $\epsilon > 0$.

Next, we study the second term on the right-hand side of (A.14). By a continuity argument, we can show that:

$$\left\|G_*(\hat{\theta}_T^{(i)}) - G_*^{(i)}(\theta_*^{(i)})\right\| = o_p(1) \tag{A.17}$$

From (A.14), (A.16) and (A.17), it follows that $G_T^{(i)}(\bar{\theta}_T^{(i)}) = O_p(1)$. Given the fact that $(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) = o_p(1)$, it follows that (A.10) is indeed correct.

Next, we state our second lemma:

Lemma 3 *Under Assumptions 1, 2 and 3,*

$$plim(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}) = E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}] \tag{A.18}$$

Proof: Given Assumptions 1 and 2 part (i), $\{g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}, t = 1, 2, ...\}$ is a strictly stationary and ergodic sequence. Assumption 3 implies $\|E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}]\| < E[\|g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}\|] < E[sup_{\theta^{(i)}\in\Theta^{(i)}}\|g_t^{(i)}(v_t, \theta^{(i)})z_t\|] < \infty$ and hence satisfies the Ergodic theorem.

$$plim(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}) = T^{-1}\sum_{t=1}^{T} E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}] \tag{A.19}$$

$$= E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}] \tag{A.20}$$

where (A.19) follows from the stationarity[13] of this sequence. The proof is complete.

Combining Lemmas 2-3, we have the following theorem:

Theorem 9 *Under Assumptions 1-9, we have:*

$$plim\big(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})z_t\big) = 0 \tag{A.21}$$

Proof: First, from Lemmas 2 and 3:

$$plim\big(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})z_t\big) = E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}] \tag{A.22}$$

---

[13] All the moments of a stationarity time series sequence are independent of time. This allows us to move $E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)}]$ out of the infinite sum.

Further, under the m.d.s assumption:

$$\begin{align} E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)})] &= E[E[g_t^{(i)}(\theta_*^{(i)}, v_t)z_t^{(i)})|\Omega_t]] \tag{A.23} \\ &= E[z_t E[g_t^{(i)}(\theta_*^{(i)}, v_t))|\Omega_t]] \tag{A.24} \\ &= 0 \tag{A.25} \end{align}$$

where (A.23) follows from the LIE, (A.24) follows from the fact that $z_t \in \Omega_t$ and (A.25) follows from the m.d.s assumption. This completes the proof.

Next, we analyze the form of the long-run covariance matrix in the special case when $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{gi}^2$.

$$\begin{align} S &= lim_{T\to\infty} var(T^{-1/2}\sum_{t=1}^{T} z_t g_t^{(i)}(\theta_*^{(i)}, v_t)) \tag{A.26} \\ &= lim_{T\to\infty} V_T \tag{A.27} \end{align}$$

where we have used V_T to denote $var(T^{-1/2}Z'g(\theta_*^{(i)}))$.

In order to understand the behavior of $lim_{T\to\infty}V_T(\theta_*^{(i)})$ better, the matrix in (A.27) can be written as follows:[14]

$$\begin{align} V_T &= T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} E[g_t^{(i)}(\theta_*^{(i)}, v_t) g_s^{(i)}(\theta_*^{(i)}) z_t^{(i)} z_s^{(i)\prime}] \tag{A.28} \\ &= T^{-1}\sum_{t=1}^{T} E[g_t^2(\theta_*^{(i)}) z_t^{(i)} z_t^{(i)\prime}] \notag \\ &+ T^{-1}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T} E[g_t^{(i)}(\theta_*^{(i)}, v_t) g_{t-s}(\theta_*^{(i)}) z_t^{(i)} z_{t-s}^{(i)\prime} + g_{t-s}(\theta_*^{(i)}) g_t^{(i)}(\theta_*^{(i)}, v_t) z_{t-s} z_t^{\prime}] \tag{A.29} \\ &= T^{-1}\sum_{t=1}^{T} var(z_t g_t^{(i)}(\theta_*^{(i)}, v_t)) \notag \\ &+ T^{-1}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T} \Big[cov(z_t g_t^{(i)}(\theta_*^{(i)}, v_t), g_{t-s}(\theta_*^{(i)}) z_{t-s}) + cov(z_{t-s} g_{t-s}(\theta_*^{(i)}), g_t^{(i)}(\theta_*^{(i)}, v_t) z_t^{(i)})\Big] \tag{A.30} \end{align}$$

---

[14] See White (2001, p. 138).

Analyzing (A.30), we see that the limiting behavior of V_T depends on the limiting behavior of an average of the variances of $z_t g_t^{(i)}(\theta_*^{(i)}, v_t)$, the average of covariances of $z_t g_t^{(i)}(\theta_*^{(i)}, v_t)$ and $z_{t-s} g_{t-s}(\theta_*^{(i)})$ for $t = 1, ..., T$ and $s = 1, ..., T$.

Lemma 4 *When* $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ *is an m.d.s., the autocovariances over* $\theta_*^{(i)}$ *are all zero, that is:*

$$E[z_s^{(i)} g_s^{(i)}(\theta_*^{(i)}) g_t^{(i)}(\theta_*^{(i)}, v_t) z_t^{(i)'}] = 0 \tag{A.31}$$

Proof:

$$\begin{aligned} E[z_s^{(i)} g_s^{(i)}(\theta_*^{(i)}) g_t^{(i)}(\theta_*^{(i)}, v_t) z_t^{(i)'}] &= E[E[z_s^{(i)} g_s^{(i)}(\theta_*^{(i)}) g_t^{(i)}(\theta_*^{(i)}, v_t) z_t^{(i)'} | \Omega_t]] && \text{(A.32)} \\ &= E[z_s^{(i)} g_s^{(i)}(\theta_*^{(i)}) E[g_t^{(i)}(\theta_*^{(i)}, v_t) | \Omega_t] z_t^{'}] \qquad \forall s < t && \text{(A.33)} \\ &= 0 && \text{(A.34)} \end{aligned}$$

where (A.32) follows from LIE, (A.33) follows from $z_s^{(i)} g_s^{(i)}(\theta_*^{(i)}) \in \Omega_t \ \forall s < t$, $z_t \in \Omega_t \ \forall t$ and (A.34) follows from the fact that $E[g_t^{(i)}(\theta_*^{(i)}, v_t) | \Omega_t] = 0$.
Under this case, we have:

$$\begin{aligned} S &= lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2 z_t^{(i)} z_t^{(i)'}] && \text{(A.35)} \\ &= E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2 z_t^{(i)} z_t^{(i)'}] && \text{(A.36)} \end{aligned}$$

where Equation (A.36) follows from Assumptions 1 and 2 [15].

Lemma 5 *If* $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ *is an m.d.s. and* $var(g_t^{(i)}(\theta_*^{(i)}, v_t) | \Omega_t) = \sigma_{gi}^2$, *we have:*

$$S = \sigma_{gi}^2 \zeta^{(i)} \tag{A.37}$$

where S *is given by (A.26).*

---

[15] The strict stationarity of $\{g_t^{(i)}(\theta_*^{(i)}, v_t)\}$ and $\{z_t\}$ together implies that $\{g_t^{(i)}(\theta_*^{(i)}, v_t)^2 z_t^{(i)} z_t^{(i)'}\}$ is strictly stationary and $lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2 z_t^{(i)} z_t^{(i)'}] = E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2 z_t^{(i)} z_t^{(i)'}]$.

Proof:

$$\begin{aligned}
S &= E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2 z_t^{(i)} z_t^{(i)\prime}] && \text{(A.38)}\\
&= E[E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2 z_t^{(i)} z_t^{(i)\prime} | \Omega_t]] && \text{(A.39)}\\
&= E[E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2 | \Omega_t] z_t^{(i)} z_t^{(i)\prime}] && \text{(A.40)}\\
&= E[\sigma_{gi}^2 z_t^{(i)} z_t^{(i)\prime}] && \text{(A.41)}\\
&= \sigma_{gi}^2 E[z_t^{(i)} z_t^{(i)\prime}] && \text{(A.42)}\\
&= \sigma_{gi}^2 \, \zeta^{(i)} && \text{(A.43)}
\end{aligned}$$

where (A.39) follows from LIE and (A.40) follows from the fact that $z_t \in \Omega_t$ $\forall t$.

Given the form of S in (A.43), intuition suggests that we could consistently estimate it using the following matrix:

$$\hat{S}_T(\hat{\theta}_T^{(i)}) \equiv \left(T^{-1} \sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})^2 \right) \left(T^{-1} \sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime} \right) \tag{A.44}$$

Next, we define the following:

$$\zeta^{(i)} = plim(\zeta_T^{(i)}) \tag{A.45}$$

where $\zeta_T^{(i)} = T^{-1} \sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}$.

The following lemma proves that this intuition is indeed correct.

Lemma 6 *$plim(\hat{S}_T(\hat{\theta}_T^{(i)}))$*

When $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{gi}^2$, a finite constant, we have:

$$\hat{S}_T(\hat{\theta}_T^{(i)}) \xrightarrow{p} S \tag{A.46}$$

Proof:

$$plim(\hat{S}_T(\hat{\theta}_T^{(i)})) = plim\Big(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})^2\Big)plim\Big(T^{-1}\sum_{t=1}^{T} z_t^{(i)}z_t^{(i)\prime}\Big) \tag{A.47}$$

$$= lim_{T\to\infty}\Big(T^{-1}\sum_{t=1}^{T} E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2]\Big)lim_{T\to\infty}\Big(T^{-1}\sum_{t=1}^{T} E[z_t^{(i)}z_t^{(i)\prime}]\Big) \tag{A.48}$$

$$= E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2]E[z_t^{(i)}z_t^{(i)\prime}] \tag{A.49}$$

$$= \sigma_{gi}^2\,\zeta^{(i)} \tag{A.50}$$

where (A.48) holds as follows:

Using a mean value expansion of $T^{-1}\sum_{t=1}^{T} g_T^{(i)}(\hat{\theta}_T^{(i)})^2$ around $\theta_*^{(i)}$, we have the following:

$$T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})^2 = T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)^2 + 2\gamma_T(\bar{\theta}_T^{(i)})(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{A.51}$$

$$= T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)^2 + o_p(1) \tag{A.52}$$

where $\gamma_T(\theta) = T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta^{(i)})\dfrac{\partial g_t^{(i)}(\theta^{(i)})}{\partial\theta^{(i)\prime}}$, (A.51) follows under our assumptions under the facts that $\hat{\theta}_T \xrightarrow{p} \theta_*$ and $\bar{\theta}_T = \lambda\hat{\theta}_T + (1-\lambda)\theta_*$ for $\lambda \in [0,1]$.

Further, using the Cauchy-Schwartz inequality, we have:

$$\gamma_T(\theta) = T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta^{(i)})\frac{\partial g_t^{(i)}(\theta^{(i)})}{\partial\theta^{(i)\prime}} \tag{A.53}$$

$$\leq \Big(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta^{(i)})^2\Big)^{1/2}\Big(T^{-1}\sum_{t=1}^{T}\Big[\frac{\partial g_t^{(i)}(\theta^{(i)})}{\partial\theta^{(i)\prime}}\Big]^2\Big)^{1/2} \tag{A.54}$$

$$= O_p(1)O_p(1) \tag{A.55}$$

$$= O_p(1) \tag{A.56}$$

From (A.52), we have:

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})^2 &\xrightarrow{p} lim_{T\to\infty}T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)^2 && \text{(A.57)}\\
&\xrightarrow{p} lim_{T\to\infty}T^{-1}\sum_{t=1}^{T} E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2] && \text{(A.58)}\\
&= lim_{T\to\infty}T^{-1}\sum_{t=1}^{T} E[E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2|\Omega_t]] && \text{(A.59)}\\
&= lim_{T\to\infty}T^{-1}\sum_{t=1}^{T} E[\sigma_{gi}^2] && \text{(A.60)}\\
&= \sigma_{gi}^2 && \text{(A.61)}\\
& && \text{(A.62)}
\end{aligned}$$

where (A.59) follows from the LIE. Given Assumptions 1 and 2, from the Ergodic theorem we have:

$$T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime} \xrightarrow{p} E[z_t^{(i)} z_t^{(i)\prime}] \tag{A.63}$$

. Also, (A.49) follows from Assumptions 1 and 2 [16]. Given Equation (1.16), $plim(NR_T)$ is given as:

$$plim(NR_T) = plim\frac{\left(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})z_t\right)'\left(T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})z_t\right)}{T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})^2} \tag{A.64}$$

$$= \frac{plim\left(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})z_t\right)' plim\left(T^{-1}\sum_{t=1}^{T} z_t^{(i)} z_t^{(i)\prime}\right)^{-1} plim\left(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})z_t\right)}{plim\left(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})^2\right)} \tag{A.65}$$

---

[16] Any measurable function of a strictly stationary sequence is also strictly stationary. Further, expectations of strictly stationary functions are independent of time. Therefore, we can move the expectations in Equation (A.47) outside of the sum.

From Equations (A.7), (A.44) and (A.65), we can see that under the current case, $T^{-1}J_T \equiv NR_T$, hence both statistics have the same probability limit given in Theorem 8[17]. Hence, under the current case the asymptotic behavior of $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ is identical to that of the GMM minimand $T^{-1}J_T^{(i)}$ and hence the Rivers and Vuong statistic B_T has the same asymptotic behavior when the goodness of fit measure is the noise ratio or the GMM minimand. That is, the asymptotic results concerning B_T are the same when the goodness of fit criteria is either $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ or $T^{-1}J_T^{(i)}(\hat{\theta}_T^{(i)})$, as those derived by Hall and Pelletier (2008) when Assumptions 1-26 and 27 hold. Therefore, for model selection based on B_T, it makes no difference asymptotically to use either statistic as goodness of fit criteria.

Next, we analyze the probability limit of NR_T when Assumptions a3571, 1 and12 hold.

Theorem 10 *If Assumptions 1-12 hold, then* $plim(NR_T) = 0$.

Proof:

$$plim(NR_T) = \phi_2^{-1}\psi_*^{(i)'}\zeta^{(i)-1}\psi_*^{(i)} \tag{A.66}$$

$$= \psi_*^{(i)'}(\phi_2\zeta^{(i)} - 1)\psi_*^{(i)} \tag{A.67}$$

$$= 0 \tag{A.68}$$

---

[17] The following is footnote 5 of Durlauf and Maccini (1995):

> When the Euler equation implicitly defines a non-i.i.d, sequence, the noise ratio will not be proportional to the J-statistic because of a need in the latter case to correct the covariance matrix of the equation's innovations in computing whether the model's coefficients obey the overidentifying restrictions. Such a correction is unnecessary in computing the variance of the model noise. However, the two are still closely related asymptotically as the noise ratio will converge to zero if the J-statistic converges to a bounded value. We thank Spencer Krane for these observations.

The probability limit of the denominator of the $NR_T^{(i)}(\hat{\theta}_T^{(i)})$ can be obtained as follows: Using a mean value expansion of $T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})^2$ around $\theta_*^{(i)}$, we have the following:

$$T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})^2 = T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)^2 + 2\gamma_T(\bar{\theta}_T^{(i)})(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{A.69}$$

$$= T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)^2 + o_p(1) \tag{A.70}$$

where $\gamma_T(\theta) = T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta^{(i)})\dfrac{\partial g_t^{(i)}(\theta^{(i)})}{\partial\theta^{(i)\prime}}$. Given the facts that $\hat{\theta}_T \xrightarrow{p} \theta_*$ and $\bar{\theta}_T = \lambda\hat{\theta}_T + (1-\lambda)\theta_*$ for $\lambda \in [0,1]$, Equation (A.69) follows under our assumptions.

Further, using the Cauchy-Schwartz inequality, we have:

$$\gamma_T(\theta) = T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta^{(i)})\frac{\partial g_t^{(i)}(\theta^{(i)})}{\partial\theta^{(i)\prime}} \tag{A.71}$$

$$\leq \left(T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta^{(i)})^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\left[\frac{\partial g_t^{(i)}(\theta^{(i)})}{\partial\theta^{(i)\prime}}\right]^2\right)^{1/2} \tag{A.72}$$

$$= O_p(1)O_p(1) \tag{A.73}$$

$$= O_p(1) \tag{A.74}$$

From (A.70), we have:

$$T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\hat{\theta}_T^{(i)})^2 \xrightarrow{p} lim_{T\to\infty}T^{-1}\sum_{t=1}^{T} g_t^{(i)}(\theta_*^{(i)}, v_t)^2 \tag{A.75}$$

$$\xrightarrow{p} lim_{T\to\infty}T^{-1}\sum_{t=1}^{T} E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2] \tag{A.76}$$

$$= lim_{T\to\infty}T^{-1}\sum_{t=1}^{T} E[E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2|\Omega_t]] \tag{A.77}$$

$$= lim_{T\to\infty}T^{-1}\sum_{t=1}^{T} E[\sigma_{git}^2] \tag{A.78}$$

where (A.77) follows from the LIE.

Next, we analyze the $plim(T^{-1}J_T^{(i)})$ under Assumptions 1-12.

Theorem 11 *If Assumptions 1-26 hold, then* $plim(T^{-1}J_T^{(i)}) = 0$.

Proof:

Under the m.d.s. assumption, all covariances are zero, hence S has the form given by Equation (A.35)[18]. Further, we can study the form of S given Assumption 11 as follows:

$$S = lim_{T\to\infty}\left(T^{-1}\sum_{t=1}^{T} E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2 z_t^{(i)} z_t^{(i)\prime}]\right) \tag{A.79}$$

$$= lim_{T\to\infty}\left(T^{-1}\sum_{t=1}^{T} E[E[g_t^{(i)}(\theta_*^{(i)}, v_t)^2|\Omega_t] z_t^{(i)} z_t^{(i)\prime}]\right) \tag{A.80}$$

$$= lim_{T\to\infty}\left(T^{-1}\sum_{t=1}^{T} E[\sigma_{git}^2 z_t^{(i)} z_t^{(i)\prime}]\right) \tag{A.81}$$

where $\Xi = E[\sigma_{git}^2 z_t^{(i)} z_t^{(i)\prime}]$ and $\|\Xi^{-1}\| < \infty$, under Assumptions 1 and 12. Hence, we have:

$$plim(T^{-1}J_T^{(i)}) = \psi_*^{(i)\prime}\Xi^{-1}\psi_*^{(i)} \tag{A.82}$$

$$= 0 \tag{A.83}$$

Finally, for the case of correctly specified models, we derive the limiting distribution of $NR_T^{(i)}(\hat{\theta}_T^{(i)})$. Before we proceed we make the following assumption about the sample moment.

Assumption 28

$$S_*^{-1/2}T^{1/2}m_T^{(i)}(\theta_*^{(i)}) \xrightarrow{d} N(0, I_{k_i}) \tag{A.84}$$

where $S_* = lim_{T\to\infty}var(T^{1/2}m_T^{(i)}(\theta_*^{(i)}))$.

Theorem 12 *Under Assumptions 1-10, 27 and 28, we have:*

$$J_T^{(i)}(\hat{\theta}_T^{(i)}) \sim \chi_{k-p}^2 \tag{A.85}$$

Proof:

If $\{g_t^{(i)}(\theta_*^{(i)}, v_t), t = 1, 2, ...\}$ is an m.d.s. and $var(g_t^{(i)}(\theta_*^{(i)}, v_t)|\Omega_t) = \sigma_{gi}^2$, then we can use

---

[18] However, S no longer has the form given in Equation (A.36), because of the dependence of σ_{git}^2 on t.

an analysis similar to that on page 69 to 73 of Hall (2005) to study the distribution of $J_T^{(i)}(\hat{\theta}_T^{(i)})$ and $TNR_T^{(i)}(\hat{\theta}_T^{(i)})$. We know from the Appendix that under this case, we have $J_T^{(i)}(\hat{\theta}_T^{(i)}) \equiv NR_T^{(i)}(\hat{\theta}_T^{(i)})$.

Using the relationship between the sample moment and the $J_T^{(i)}(\hat{\theta}_T^{(i)})$, we begin our analysis by a mean value expansion of $m_T^{(i)}(\hat{\theta}_T^{(i)})$ around $\theta_*^{(i)}$, in Equation (A.9). First, we multiply Equation (A.9) through by $T^{1/2}\hat{S}_T^{-1/2}$.[19]

$$\hat{S}_T^{-1/2}T^{1/2}m_T^{(i)}(\hat{\theta}_T^{(i)}) = \hat{S}_T^{-1/2}T^{1/2}m_T^{(i)}(\theta_*^{(i)}) + \hat{S}_T^{-1/2}G_T T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) \tag{A.86}$$

Since $\hat{\theta}_T^{(i)}$ represents the second step GMM estimator[20], we can substitute the following[21] into (A.86):

$$T^{1/2}(\hat{\theta}_T^{(i)} - \theta_*^{(i)}) = -[G_T^{(i)}(\hat{\theta}_T^{(i)})'\hat{S}_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}G_T^{(i)}(\bar{\theta}^{(i)})]^{-1}G_T^{(i)}(\hat{\theta}_T^{(i)})\hat{S}_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}T^{1/2}m_T^{(i)}(\theta_*^{(i)}) \tag{A.87}$$

which yields:[22]

$$\hat{S}_T^{-1/2}T^{1/2}m_T^{(i)}(\hat{\theta}_T^{(i)}) = K_T(\hat{\theta}_T^{(i)})\hat{S}_T^{-1/2}m_T^{(i)}(\theta_*^{(i)}) \tag{A.88}$$

where

$$K_T(\hat{\theta}_T^{(i)}) = I_{k_i} - \hat{S}_T^{-1/2}G_T^{(i)}(\bar{\theta}^{(i)})[G_T^{(i)}(\hat{\theta}_T^{(i)})'\hat{S}_T^{(i)}(\hat{\theta}_T^{(i)})^{-1}G_T^{(i)}(\bar{\theta}^{(i)})]^{-1}G_T^{(i)}(\hat{\theta}_T^{(i)})'\hat{S}_T^{-1/2\prime} \tag{A.89}$$

Equations (A.88) and (A.89) together imply:

$$\hat{S}_T^{-1/2}T^{1/2}m_T^{(i)}(\hat{\theta}_T^{(i)}) = [I_{k_i} - P_o(\theta_*^{(i)})]S_*^{-1/2}T^{1/2}m_T^{(i)}(\theta_*^{(i)}) + o_p(1) \tag{A.90}$$

where $S_* = plim(\hat{S}_T)$ and $P_o(\theta_*^{(i)}) = plim(K_T(\hat{\theta}_T^{(i)}))$.

Now, $J_T^{(i)} = \{\hat{S}_T^{1/2}T^{1/2}m_T^{(i)}(\hat{\theta}_T^{(i)})\}'\{\hat{S}_T^{1/2}m_T^{(i)}(\hat{\theta}_T^{(i)})\}$. Hence, using Equation (A.90), we can

---

[19] Since $\hat{S}_T^{-1}$ is p.s.d., such a square root matrix exists by a Cholesky decomposition.

[20] In the first step, a suboptimal weighting matrix is used to estimate $\hat{\theta}_{T,1}$, which is called the first step estimator. On the second step, this first step estimator is used to construct a weighting matrix, $\hat{S}_T^{-1}(\hat{\theta}_{T,1})$, in order to estimate the second step GMM estimator.

[21] See Equation 3.26 of Hall (2005)[p.70].

[22] See Equations 3.35 and 3.36 of Hall (2005)[p.73].

derive the following: [23]

$$J_T^{(i)}(\hat{\theta}_T^{(i)}) = T^{1/2} m_T^{(i)}(\theta_*^{(i)})' S_*^{-1/2\,\prime} [I_{k_i} - P_o(\theta_*^{(i)})] S_*^{-1/2} m_T^{(i)}(\theta_*^{(i)}) + o_p(1) \tag{A.91}$$

Given Assumption 28, Equation (A.91) implies:[24]

$$J_T^{(i)}(\hat{\theta}_T^{(i)}) \xrightarrow{d} n_{k_i}[I_{k_i} - P_o(\theta_*^{(i)})] n_{k_i} \tag{A.92}$$

where $n_{k_i} \sim N(0, I_{k_i})$ and from (A.92) we claim the following:[25]

$$n_{k_i}[I_{k_i} - P_o(\theta_*^{(i)})] n_{k_i} \sim \chi_d^2 \tag{A.93}$$

where $d = rank([I_{k_i} - P_o(\theta_*^{(i)})])$. Given $[I_{k_i} - P_o(\theta_*^{(i)})]$ is idempotent, $rank[I_{k_i} - P_o(\theta_*^{(i)})] = trace[I_{k_i} - P_o(\theta_*^{(i)})]$.

Finally, we have:

$$\begin{aligned}
trace[I_{k_i} - P_o(\theta_*^{(i)})] &= trace(I_{k_i}) - trace(P_o(\theta_*^{(i)})) && \text{(A.94)}\\
&= k - trace(P_o(\theta_*^{(i)})) && \text{(A.95)}\\
&= k - trace[plim(\hat{S}_T^{-1/2} G_T^{(i)}(\bar{\theta}^{(i)})[G_T^{(i)}(\hat{\theta}_T^{(i)})' \hat{S}_T^{(i)}(\hat{\theta}_T^{(i)})^{-1} G_T^{(i)}(\bar{\theta}^{(i)})]^{-1} \\
&\quad \times G_T^{(i)}(\hat{\theta}_T^{(i)}) \hat{S}_T^{-1/2\,\prime})] && \text{(A.96)}\\
&= k - trace[S_*^{-1/2} G_*^{(i)}(\theta_*^{(i)})[G_*^{(i)}(\theta_*^{(i)})' S_*(\theta_*^{(i)})^{-1} G_*^{(i)}(\theta_*^{(i)})]^{-1} \\
&\quad \times G_*^{(i)}(\theta_*^{(i)}) S_*^{-1/2\,\prime}] && \text{(A.97)}\\
&= k - trace([G_*^{(i)}(\theta_*^{(i)})' S_*(\theta_*^{(i)})^{-1} G_*^{(i)}(\theta_*^{(i)})]^{-1} [G_*^{(i)}(\theta_*^{(i)}) S_*(\theta_*^{(i)})^{-1} G_*^{(i)}(\theta_*^{(i)})]) \\
& && \text{(A.98)}\\
&= k - trace(I_p) && \text{(A.99)}\\
&= k - p && \text{(A.100)}
\end{aligned}$$

where (A.98) follows from the fact that, for conformable matrices A, B and C, $trace(ABC) = tr(BCA)$.

When Assumptions 26 and 27 hold and $\hat{S}_T^{(i)}$ is given by Equation (A.44), we have $J_T^{(i)} \equiv$

---

[23] We also used the fact that $T^{1/2} m_T^{(i)}(\theta_*^{(i)})'$ and $[I_{k_i} - P_o(\theta_*^{(i)})] S_*^{-1/2}$ are orthogonal.

[24] See Gallant 1987[p.122].

[25] See Gallant 1987[p.122].

TNR_T. Hence, from Theorem 12 we conclude that $TNR_T^{(i)}(\hat{\theta}_T^{(i)}) \sim \chi^2_{k-p}$.