General Method of Moments Bias and Specification Tests for Quantile Regression

by

Ziad H. Nejmeldeen

B.A. Economics, B.A. Statistics, B.A. Applied Mathematics
U.C. Berkeley, 1998

SUBMITTED TO THE DEPARTMENT OF ECONOMICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF

DOCTOR OF PHILOSOPHY IN ECONOMICS
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2003

Signature of Author: ______________________
Department of Economics
August 8, 2003

Certified By: ______________________
Primary Thesis Supervisor: Whitney Newey
Title: Professor of Economics

Accepted By: ______________________

Peter Temin
Professor of Economics
Departmental Thesis Administrator

General Method of Moments Bias and Specification Tests for Quantile Regression

by

Ziad H. Nejmeldeen

Submitted to the Department of Economics on August 8, 2003 in Partial Fulfillment of the Requirements of the Degree of Doctor of Philosophy in Economics.

Abstract

Chapter 1 This chapter looks at a dynamic panel data model with fixed effects. Estimating the model with GMM is consistent but suffers from small sample bias. We apply Helmert's transformation to the model, assume that error terms and nuisance parameters are homoskedastic and independent across observations and of one another, and utilize the GMM bias calculation of Newey & Smith (2001). This leads to a closed form expression for the GMM bias applied to AR(1) model.

Chapter 2 This chapter develops specification tests for quantile regression under various data types. We consider what happens to the quantile regression estimator under local and global misspecification and design specification tests that handle a wide range of data types. We consider how to carry out such tests in practice and present Monte Carlo results to show the effectiveness of such tests.

Chapter 3 Through a Taylor expansion, We compute the bias of a general GMM model where the weighting matrix $\hat{A}$ of the moment conditions $g(z, \beta)$ is left unspecified, except for some general conditions. Our bias results are compared to those of Newey and West (2003). An important case of GMM estimation with a general weighting matrix $\hat{A}$ is when $\hat{A}$ is a function of a vector of parameters with fixed dimension. Arellano's IVE estimator is an example of this type of estimator - we consider the bias properties of Arellano's IVE estimator in the AR(1) setting and compare them to our results from Chapter 1.

Primary Thesis Supervisor: Whitney Newey
Title: Professor of Economics
Secondary Thesis Supervisor: Victor Chernozhukov
Title: Assistant Professor of Economics

Contents

Part I

Small Sample GMM Bias in the AR(1) Model

1 Introduction

Consider the dynamic panel model with fixed effects:

$$y_{i,t} = \alpha y_{i,t-1} + \eta_i + v_{i,t}, \;\; t = 2, \ldots, T, \;\; i = 1, \ldots, N$$

For large n, fixed T, Nickell (1981) showed that the standard fixed effects MLE suffers from the incidental parameter problem, which leads to inconsistency. To avoid the inconsistency of MLE, the literature has focused on applying GMM to first differences. In carrying out GMM, Ahn & Schmidt (1995) and Hahn (1999) have shown that the orthogonality of the lagged levels with the 1st differences provides the largest source of information.

The problem is that even though GMM is consistent, the estimator suffers from substantial finite sample bias (Alonso-Borrego & Arellano (1996)), which leads to inference problems. Our goal will be to find a closed-form expression for this bias and to understand how the bias grows with T. This problem is more approachable by first applying Helmert's Transformation.

Arellano & Bover (1995) note that the efficiency of the resultant GMM is not affected whether we use first differencing or Helmert's Transformation. This transformation is also used in Hahn, Hausman, and Kuersteiner (2002), where a second-order approach was taken to compute the bias as a function of the data. The resulting bias formula was used to construct a bias-corrected GMM estimator.

Our paper takes a different approach at the bias calculation. Our derivation of the bias will follow from Newey & Smith (2001). They give a general form of the bias of GMM that can be applied to the AR(1) panel data model. We will assume that error terms and nuisance parameters are homoskedastic and independent across observations and of one-another. This yields a simplification of the part of GMM bias which grows with the number of moment restrictions. Using this simplification, we will compute a closed-form expression of the bias as a function of the underlying parameters. The resulting bias calculation will be used to construct a bias-corrected GMM estimator.

2 GMM Bias

We consider a standard GMM model with a fixed number of moment restrictions. Let z_i $(i = 1, \ldots, n)$ be i.i.d. observations on a data vector z. Let β be a $KX1$ parameter vector and let our moment conditions be given by $g(z, \beta)$, a $JX1$ vector. At the true parameter β_0, we require that $g(z, \beta_0) = 0$.

The two-step GMM estimator of Hansen (1982) is given by:

$$\hat{\beta}_{GMM} = argmin_{\beta}\hat{g}(\beta)'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\beta) \tag{1}$$

where $\Omega(\beta) = (1/n)\sum_{i=1}^{n} g_i(\beta)g_i(\beta)'$ and $\tilde{\beta} = argmin_{\beta}\hat{g}(\beta)\hat{W}^{-1}\hat{g}(\beta)$, where $\hat{W}$ is an initial weighting matrix.

Newey and Smith (2003) derive stochastic expansions for this two-step GMM estimator. Under identification and regularity assumptions, as well as conditions on the initial weighting matrix $\hat{W}$, they find the asymptotic bias of GMM to be given by:

$$\begin{aligned} Bias(\hat{\beta}_{GMM}) &= B_I + B_G + B_\Omega + B_W \\ B_I &= H(-a + E[G_i H g_i])/n \\ B_G &= -\Sigma E[G_i{}' P g_i]/n \\ B_\Omega &= H E[g_i g_i{}' P g_i]/n \\ B_W &= H \textstyle\sum_{j=1}^{K} \tilde{\Omega}_{\beta_j}(H_W - H)' e_j/n \end{aligned}$$

where $H = \Sigma G'\Omega^{-1}$, $H_W = (G'W^{-1}G)^{-1}G'W^{-1}$, $G = E(G_i)$, $P = \Omega^{-1} - \Omega^{-1}G\Sigma G'\Omega^{-1}$, $\tilde{\Omega}_{\beta_j} = E\left[\partial\{g_i(\beta_0)g_i(\beta_0)'\}/\partial\beta_j\right]$, $\Sigma = (G'\Omega^{-1}G)^{-1}$, and a is an m-vector such that:

$$a_j = tr(\sum E\left[\partial^2 g_{ij}(\beta_0)/\partial\beta\partial\beta'\right])/2, \;\; j = 1, \ldots, m$$

Newey and Smith (2003) give interesting interpretations of the various parts of the bias. B_I is the asymptotic bias for a GMM estimator with the optimal linear combination $G'\Omega^{-1}g(z, \beta)$. B_G arises from estimating $G = E(G_i)$. This is zero if G_i is constant, but is generally non-zero if there is endogeneity. B_Ω arises from estimating Ω; this is zero if the third moments are zero, but is generally non-zero. B_W arises from the choice of $\hat{W}$, the first step weighting matrix. It is zero if W is a scalar multiple of Ω.

We will apply this bias calculation to the AR(1) model, assuming that the error terms and nuisance parameters are homoskedastic and independent across observations and of one-another. We next present the AR(1) model.

3 The AR(1) Model

Consider the dynamic panel data model with fixed effects:

$$y_{i,t} = \alpha y_{i,t-1} + \eta_i + v_{i,t}, \quad i = 1, \ldots, n \quad t = 2, \ldots, T \tag{2}$$

We make the following assumptions:

Assumption 1 $E(v_{i,t}v_{j,s}) = 0 \;\forall\; i \neq j \;\; or \; t \neq s, \; E(\eta_i \eta_j) = 0 \;\forall\; i \neq j$

Assumption 2 $E(\eta_i v_{j,t}) = 0 \;\forall\; i, j, t$

Assumption 3 $Var(v_{i,t}) = \sigma^2 \;\forall\; i, t, \; Var(\eta_i) = \sigma_\eta^2 \;\forall\; i.$

That is to say, $v_{i,t}$ and η_i are independent across i, t, are independent of one another, and each has constant variance.

GMM estimation is based on the orthogonality of the observations and some function of the error terms. The usual way to achieve this is through the first-difference form of the model.

Applying first-differences, we are able to exclude the η_i term:

$$(y_{i,t} - y_{i,t-1}) = \alpha(y_{i,t-1} - y_{i,t-2}) + (v_{i,t} - v_{i,t-1})$$

The transformed error terms $v_{i,t} - v_{i,t-1}$ are independent of $y_{i,s}$ for $s = 0, \ldots, t-2$ and GMM is based on this orthogonality:

$$E\left(y_{i,s-1}(v_{i,t} - v_{i,t-1})\right) = 0, \quad s \leq t, \; t = 2, \cdots, T-1$$

Instead of first-differencing, we follow the work of Arellano and Bover (1995) and apply Helmert's Transformation. This will greatly simplify the calculation of the GMM bias. The efficiency of the resultant GMM estimator is the same whether first-differencing or Helmert's Transformation is used. Applying Helmert's transformation to the observations, we have:

$$y^*_{i,t} = c_t \left(y_{i,t} - \frac{1}{T-t} \sum_{k=t+1}^{T} y_{i,k} \right), \qquad c_t = \sqrt{\frac{T-t}{T-t+1}}$$

This transformation gives:

$$y^*_{i,t} = \alpha c_t \left(y_{i,t-1} - \frac{1}{T-t} \sum_{k=t}^{T-1} y_{i,k} \right) + v^*_{i,t} \tag{3}$$

Allowing $x_t = y_{t-1}$, Eq. (2) can be written as:

$$y^*_{i,t} = \alpha x^*_t + v^*_{i,t}$$

If the $v_{i,t}$ are uncorrelated across time, then the transformed errors $v^*_{i,t}$ will retain this property: $E\left(v^*_{i,t} v^*_{i,t+j}\right) = 0$ for $j \neq 0$. This implies $E\left(v^*_{i,t} y_{i,s-1}\right) = 0$ for $s \leq t$, allowing us to use $y_{i,1}, \ldots, y_{i,t-1}$ as instruments.

HHK(2001) show that, with the homoskedasticity assumption on the $v_{i,t}$, the GMM estimator based on this orthogonality is actually a linear combination of the 2SLS estimators $\hat{b}_{2SLS,1}, \ldots, \hat{b}_{2SLS,T-1}$, where $\hat{b}_{2SLS,t}$ is the 2SLS of y^*_t on x^*_t. Since 2SLS is subject to substantial finite sample bias, it is reasonable to believe that our GMM estimator will have the same deficiency.

4 Closed-Form Bias for AR(1) Model

The computation of the bias will involve a number of steps. In Section 4.1, the general form of GMM bias for the AR(1) model will be given. In Section 4.2, the moment conditions will be addressed. Section 4.3 will address computation of the GMM weighting matrix Ω^{-1}. Sections 4.4 and 4.5 will focus on computing the two parts of the bias, $(E(G_i(\alpha))'\Omega^{-1}E(G_i(\alpha)))^{-1}$ and $E(G_i'\Omega^{-1}g_i)$, respectively. Section 4.6 will put all the pieces together to give a closed form expression for GMM bias for the AR(1) model.

4.1 General GMM Bias for AR(1)

We wish to apply the GMM bias results of Newey and West (2003) to the AR(1) model.

Lemma 1 *Consider our AR(1) model under assumptions 1-3. The finite sample GMM bias which grows with the number of moment restrictions is:*

$$
\begin{aligned}
BIAS &= -(E(G_i(\alpha))'\Omega^{-1}E(G_i(\alpha)))^{-1}E(G_i'\Omega^{-1}g_i)/n, \quad where \\
\Omega^{-1} &= E\left(g_i\left(\alpha\right)g_i\left(\alpha\right)'\right)^{-1} \\
G_i(\alpha) &= \tfrac{d}{d\alpha}g_i(\alpha)
\end{aligned}
$$

Proof: See Appendix A.1

Our goal in the following sections will be to compute the two components of the bias $(E(G_i(\alpha))'\Omega^{-1}E(G_i(\alpha)))^{-1}$ and $E(G_i'\Omega^{-1}g_i)$ for the AR(1) model under Assumptions 1-3. This will give us an understanding of how tGMM bias is growing with the number of moment restrictions.

4.2 Moment Conditions

Helmert's Transformation has given us the orthogonality conditions needed to conduct GMM. For each of the transformed equations, we have $t-1$ lagged values of y that we can use as instruments. The moment conditions are: $E\left(v_{i,t}^{*}y_{i,s-1}\right) = 0$ for $s \leq t$, $t = 2, \cdots, T-1$, where $v_{it}^{*} = y_{it}^{*} - \alpha x_{it}^{*}$. Define $\tilde{y}_t = (y_{i,1}y_{i,2}\cdots y_{i,t})'$ and $g_{i,t} = v_{i,t+1}^{*}\tilde{y}_{i,t}$. Then our moment conditions are given by $E(g_i(\alpha)) = 0$ where the vector $g_i(\alpha) = (g_{i,1}'g_{i,2}'\cdots g_{i,T-2}')$ is $\frac{(T-2)(T-1)}{2}$ X 1.

4.3 Computing $\Omega^{-1} = E\left(g_i\left(\alpha\right)g_i\left(\alpha\right)'\right)^{-1}$

We begin with the definition of the cross product of the g_i's: Let $F_i \equiv g_i\left(\alpha\right)g_i\left(\alpha\right)'$ and $f_i\left(a,b\right) \equiv \left[F_i\right]_{a,b}$. Note that $\Omega = E(F_i)$. We will first solve for Ω and then invert the matrix.

4.3.1 Solving for Ω

We must solve $E\left(f_i\left(a,b\right)\right)$ for various a, b. The following definition will allow us to solve $E\left(f_i\left(a,b\right)\right)$ under three cases.

a, b are in the same *family* if $j(a) = j(b)$. To simplify notation, let $p = \frac{\sigma^2}{1-\alpha^2}$, $q = \frac{\sigma_\eta^2}{(1-\alpha)^2}$.

Theorem 1 *We have the following results:*

- *If* $a = b$, *then* $E\left[f_i(a,a)\right] = \sigma^2(p+q)$

- *If $a \neq b$ but a and b are in the same family, then* $E\left[f_i(a,b)\right] = \sigma^2\left(\alpha^{|b-a|}p + q\right)$
- *If $a \neq b$ and a and b are in different families, then* $E\left[f_i(a,b)\right] = 0$

Proof: See Appendix A.3.

Theorem 1 gives us the following representation of Ω, which is a block diagonal matrix:

$$\Omega_i = \begin{pmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_{T-2} \end{pmatrix}, \quad \underbrace{A_t}_{t\times t} = \sigma^2 \begin{pmatrix} p+q & \alpha p+q & \alpha^2 p+q & \cdots & \alpha^{t-1}p+q \\ \alpha p+q & p+q & \alpha p+q & \cdots & \alpha^{t-2}p+q \\ \alpha^2 p+q & \alpha p+q & p+q & \cdots & \alpha^{t-3}p+q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{t-1}p+q & \alpha^{t-2}p+q & \alpha^{t-3}p+q & \cdots & p+q \end{pmatrix}$$

4.3.2 Computing Ω^{-1}

We can rewrite A_t as $A_t = \sigma^2\left(pM_t + qe_te_t'\right)$, where e_t is a $t \times 1$ column of 1's and M_t is a t x t matrix. To invert Ω_i, we must first invert the A_t matrices. This requires an application of the partitioned matrix inverse formula (Linear Statistical Inference by Rao). The formula states that if B is a nonsingular matrix and U, V are column vectors, then

$$(B+UV')^{-1} = B^{-1} - \frac{\left(B^{-1}U\right)\left(V'B^{-1}\right)}{1+V'B^{-1}U}$$

Theorem 2 *Applying the partitioned matrix inverse formula gives us the following representation of Ω_i^{-1}:*

$$\Omega_i^{-1} = \begin{pmatrix} A_1^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_{T-2}^{-1} \end{pmatrix}, \quad A_t^{-1} = \frac{1}{\sigma^4}\left(N_t - h_tR_t\right)\forall\, t$$

$$h_t = \left(\frac{\sigma_\eta^2}{\sigma^2+\sigma_\eta^2\left(t+\frac{2\alpha}{1-\alpha}\right)}\right)\forall t$$

$$N_1 = (1-\alpha^2), \quad N_2 = \begin{pmatrix} 1 & -\alpha \\ -\alpha & 1 \end{pmatrix}, \quad R_1 = (1+\alpha)^2, \quad R_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$ *For* $t \geq 3:$

$$N_t = \begin{pmatrix} 1 & -\alpha & 0 & \cdots & 0 \\ -\alpha & 1+\alpha^2 & -\alpha & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\alpha & 1+\alpha^2 & -\alpha \\ 0 & \cdots & 0 & -\alpha & 1 \end{pmatrix}, \; R_t = \begin{pmatrix} 1 & 1-\alpha & \cdots & 1-\alpha & 1 \\ 1-\alpha & (1-\alpha)^2 & \cdots & (1-\alpha)^2 & 1-\alpha \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-\alpha & (1-\alpha)^2 & \cdots & (1-\alpha)^2 & 1-\alpha \\ 1 & 1-\alpha & \cdots & 1-\alpha & 1 \end{pmatrix}$$

Proof: See Appendix A.4.

With Ω^{-1} computed, we can concentrate on computing the two parts of the bias: $E(G_i'\Omega^{-1}g_i)$ and $(E(G_i(\alpha))'\Omega^{-1}E(G_i(\alpha)))^{-1}$.

4.4 Computation of $E(G_i'\Omega^{-1}g_i)$

This part of the computation is by far the most intensive. Recall $g_{i,t} = v^*_{i,t+1}\tilde{y}_t$ where $v^*_{i,t+1}$ is the error term in the transformed equation (Eq. 2), $\tilde{y}_t = \begin{pmatrix} y_{i,1} & y_{i,2} & \cdots & y_{i,t} \end{pmatrix}'$, $G_i(\alpha) = \frac{d}{d\alpha}g_i(\alpha)$. Then we have:

$$\begin{aligned} G_i(\alpha) &= \begin{pmatrix} G'_{i,1} & G'_{i,2} & \cdots & G'_{i,T-2} \end{pmatrix}' \\ G_t &= -c_{t+1}\left[y_{i,t} - \frac{1}{T-t-1}\sum_{k=t+1}^{T-1} y_{i,k}\right]\tilde{y}_t \end{aligned}$$

We are now able to write $E\left(G_i'\Omega^{-1}g_i\right) = \sum_{t=1}^{T-2} E\left(G_t'A_t^{-1}g_t\right)$. The right-hand side term separates into parts, given in Theorem 1 (Apppendix B.1). The expectations are taken in Theorem 2 (Apppendix B.2). Theorem 2 gives the final closed-form expression for $E\left(G_i'\Omega^{-1}g_i\right)$.

Theorem 3

$$E\left(G_i'\Omega^{-1}g_i\right) = \sum_{t=1}^{T-2} B(\alpha,\sigma,T,t) * Term(t)$$

$$B(\alpha,\sigma,T,t) = -\left[1+\alpha^{T-t-1}[\alpha(T-t-1)-(T-t)]\right] / \left[\sigma^2(1-\alpha)^4(1+\alpha)(T-t)(T-t-1)\right]$$

$$Term(t) = \begin{cases} 2\sigma_\eta^2\alpha(1-\alpha^2) - \sigma^2(1-\alpha)^3 & t=1 \\ \left[\sigma_\eta^4(1+\alpha)^2 + 2\sigma^2\sigma_\eta^2\alpha(1-\alpha^2) - \sigma^4(1-\alpha)^2(1-2\alpha)\right] / \left[2\sigma_\eta^2+\sigma^2(1-\alpha)\right] & t=2 \\ \left[\sigma^4 C(\alpha,t) + \sigma^2\sigma_\eta^2 D(\alpha,t) + \sigma_\eta^4 E(\alpha,t)\right] / \left[\sigma^2(1-\alpha)+\sigma_\eta^2(t(1-\alpha)+2\alpha)\right] & t\geq 3 \end{cases}$$

$$\begin{aligned} C(\alpha,t) &= (1-\alpha)^2(1-t(1-\alpha^2)) \\ D(\alpha,t) &= (1-\alpha)\left[(1-\alpha)[2\alpha(\alpha^t+\alpha^{t-2}-2)-1-t(t(1-\alpha^2)+2\alpha(1+\alpha)-1)]+1+3\alpha\right] \\ E(\alpha,t) &= (1+\alpha)\left[(1-\alpha)[5-3t+2\alpha^{t-1}+4\alpha(t-2)]+2\alpha\right] \end{aligned}$$

Proof: See Appendix A.7.

4.5 Solving for $\Sigma = E(G_i(\alpha))'\Omega^{-1}E(G_i(\alpha)))^{-1}$

Using Ω^{-1} and $E(G_i)$ (Lemma 4, Appendix C.1), we have a closed-form expression for Σ:

Theorem 4 *Plugging in for Ω^{-1} and $E(G_i(\alpha))$ (Lemma 3 - See Appendix A.8) gives:*

$$\Sigma = \frac{(1+\alpha)}{\sigma^4}\left(\sum_{t=1}^{T-2}\frac{T-t}{T-t-1}\left[1-\frac{1-\alpha^{T-t}}{(1-\alpha)(T-t)}\right]^2\left(\frac{\sigma^2+\sigma_\eta(t-1)}{\sigma^2(1-\alpha)+\sigma_\eta[t(1-\alpha)+2\alpha]}\right)\right)^{-1}$$

Proof: See Appendix A.9.

4.6 The Bias

Combining Σ and $E(G_i'\Omega^{-1}g_i)$, we have:

$$BIAS = -\Sigma E(G_i'\Omega^{-1}g_i)/n \tag{4}$$

$$\Sigma = \frac{(1+\alpha)}{\sigma^4}\left(\sum_{t=1}^{T-2}\frac{T-t}{T-t-1}\left[1-\frac{1-\alpha^{T-t}}{(1-\alpha)(T-t)}\right]^2\left(\frac{\sigma^2+\sigma_\eta(t-1)}{\sigma^2(1-\alpha)+\sigma_\eta[t(1-\alpha)+2\alpha]}\right)\right)^{-1}$$

$$E\left(G_i'\Omega^{-1}g_i\right) = \sum_{t=1}^{T-2} B(\alpha,\sigma,T,t) * Term(t)$$

$$B(\alpha,\sigma,T,t) = -\left[1+\alpha^{T-t-1}[\alpha(T-t-1)-(T-t)]\right] / \left[\sigma^2(1-\alpha)^4(1+\alpha)(T-t)(T-t-1)\right]$$

$$Term(t) = \begin{cases} 2\sigma_\eta^2\alpha(1-\alpha^2)-\sigma^2(1-\alpha)^3 & t=1 \\ \left[\sigma_\eta^4(1+\alpha)^2+2\sigma^2\sigma_\eta^2\alpha(1-\alpha^2)-\sigma^4(1-\alpha)^2(1-2\alpha)\right] / \left[2\sigma_\eta^2+\sigma^2(1-\alpha)\right] & t=2 \\ \left[\sigma^4C(\alpha,t)+\sigma^2\sigma_\eta^2D(\alpha,t)+\sigma_\eta^4E(\alpha,t)\right] / \left[\sigma^2(1-\alpha)+\sigma_\eta^2(t(1-\alpha)+2\alpha)\right] & t\geq 3 \end{cases}$$

$$\begin{aligned} C(\alpha,t) &= (1-\alpha)^2(1-t(1-\alpha^2)) \\ D(\alpha,t) &= (1-\alpha)\left[(1-\alpha)[2\alpha(\alpha^t+\alpha^{t-2}-2)-1-t(t(1-\alpha^2)+2\alpha(1+\alpha)-1)]+1+3\alpha\right] \\ E(\alpha,t) &= (1+\alpha)\left[(1-\alpha)[5-3t+2\alpha^{t-1}+4\alpha(t-2)]+2\alpha\right] \end{aligned}$$

5 Monte Carlo

Equation (3) gives us a closed form calculation for the bias of GMM under a dynamic panel data model with fixed effects. Using R, we have constructed a function GMMBIAS which takes α,σ_η^2,σ^2,n, and T as its arguments.

With $\sigma_{eta} = \sigma = 1$, Table 1 tabulates both the actual bias approximated by 10000 Monte Carlo runs and the bias predicted by Second-Order Theory - these results are taken directly from HHK, Table 1. Our own calculations for the bias are calculated in R using the function GMMBIAS, and are included in Table 1.

Of course, when carrying out GMM estimation for the dynamic panel data model with fixed effects, α will not be known and σ_η^2 and σ^2 are generally unknown. HHK uses an estimator of the Second-Order Bias to create a bias corrected estimator, $\hat{\alpha}_{BC2}$. In the same way, we can construct an estimator $\hat{\alpha}_Z$ by plugging in estimates for the unknowns in our own bias calculator:

$$\hat{\alpha}_Z = \hat{\alpha}_{GMM} - GMMBIAS(\hat{\alpha}_{GMM}, \hat{\sigma}_\eta^2, \hat{\sigma}^2, n, T)$$

In computing $\hat{\alpha}_{BC2}$, HHK ran a Monte Carlo with 5000 iterations. We ran a separate Monte Carlo with 5000 iterations and have included our results with theirs in Table 2.

As we saw in Table 1, our calculation of the bias is poor for values of α close to 1. In running a Monte Carlo with 5000 iterations, there will be a handful of iterations where GMM will estimate α as close to 1 even when it is not. Those particular estimates, when run through GMMBIAS, will lead to an extremely biased $\hat{\alpha}_Z$.

Table 2 does not include the mean of $\hat{\alpha}_Z$ for the 5000 iterations. Instead, we have included two variants: $\hat{\alpha}_{Z,trunc}$, and $\hat{\alpha}_{Z,med}$. $\hat{\alpha}_{Z,trunc}$ throws out the maximum 100 and minimum 100 $\hat{\alpha}_Z$ estimators and then takes the mean. $\hat{\alpha}_{Z,med}$ simply looks at the median $\hat{\alpha}_Z$ among the 5000 estimators. Both methods are meant to minimize the impact of extreme estimators stemming from a GMM estimate of α close to 1.

In the last two columns of Table 2, two other variants of $\hat{\alpha}_Z$ are included: $\tilde{\alpha}_{Z,trunc}$ and $\tilde{\alpha}_{Z,med}$. These estimators are computed with σ_η^2 and σ^2 assumed known and set to 1. There does seem to be some improvement in our estimator with the variances assumed known, but the empirical evidence is not overwhelming.

6 Conclusion

The focus of the research here was to find an expression for the part of GMM bias that grows with the number of moment restrictions in the AR(1) model. Applying Helmert's transformation to the panel data model, and making simplifying assumptions on co-variance of the error terms (Assumptions 1-3) allows us to find such an expression. Section 4.6 presents the GMM bias which is easily programmable, and therefore, applicable. We have constructed an estimator $\hat{\alpha}_Z$ which carries out GMM as a first step, estimates the bias based on the first-order GMM calculation, and yields a bias-corrected estimator in the second-step. The estimator does not do as well as $\hat{a}_{BC2}$, but there is hope that this can be improved by including the parts of the bias which do not grow with the number of moment restrictions.

Part II

Specification Test Processes for Quantile Regression

1 Introduction

This paper is concerned with providing simple, attractive tests - both computationally and theoretically - with regards to the validity of quantile regression models as global descriptions of the conditional distribution.

Quantile regression models allow us to focus on local slices of the conditional distribution and isolate factors that influence particular quantiles without imposing the restriction that these factors affect other quantiles in the same way.

Section 2 will present various data types which will be covered by our specification tests. Section 3 will consider what happens to quantile regression under misspecification. Section 4 will outline specification tests. Monte Carlo results will be presented in Section 5.

2 Data Types

2.1 Quantile Regression Specification

Our target is the conditional quantile model of the dependent real variable Y given covariates X in $\mathbb{R}^d$, $Q_{Y|X}$. $Q_{Y|X}$ is the inverse of the conditional distribution function $F_{Y|X}$:

$$Q_{Y|X}(\tau) = \inf_{v \in \mathbb{R}} \{v : F_{Y|X}(v) \geq \tau\};$$

therefore $Q_{Y|X}$ is a complete description of the stochastic relation of Y to X.

The linear model of $Q_{Y|X}$ is of fundamental importance, convenience, conceptual appeal and computational simplicity, incorporating the classical linear model and linear location-scale models as special cases,

$$Q_{Y|X}(\tau) = X'\beta(\tau),$$

or, equivalently, in terms of a random coefficient's model,

$$Y = Q_{Y|X}(U) = X'\beta(U),$$

where $U \sim Uniform(0,1)$ and independent of X. And, indeed, such a model is central to a substantial number of empirical studies. Nonlinear models have also been used but we shall focus only on the analysis of the linear model with no endogeniety. (Incorporating endogeneity is straightforward in GMM framework provided there are instruments Z that are independent of the error U, correlated with the endogenous variables X).

This paper focuses on determining whether $X'\beta(\cdot)$ is an accurate description of $Q_{Y|X}$ - this is a goodness-of-fit problem. The tests offered here apply to a wide range of data types encountered in empirical research. The next section presents some basic reference models.

2.2 The Data Models

We will design tests to cover the following general data types. The 4 models given here (**M1**,**M2a**,**M2b**,**M3**,) are intended to cover a wide variety of applications.

2.2.1 iid Data

Model 1 (M1) $\{W_t = (Y_t, X_t), t \leq n\}$ is an i.i.d. triangular sequence.

iid sampling is a general mechanism that applies to a variety of situations.

2.2.2 Panel Data Models with Random Effects

Panel data consists of cross-section and time dimensions. The cross-section dimension of the panel is indexed by $i \leq n_0$ and the time dimension by $j \leq J$, with the total number of observations denoted as n. We assume that $n_0 \to \infty$ and J is fixed.

The notion of "time" is defined broadly. E.g., in twins studies, i may denote the twin pair, and "time" j denotes 1 or 2. In education studies, where it is important to account for peer effects, i may denote a particular class or group under consideration and j may denote a pupil in this class.

Models of panel data with random effects incorporate dependence across the "time" dimension in order to correctly conduct inference. For example, consider a mean regression model:

$$\begin{aligned} Y_{ij} &= v_i + X_{ij}'\alpha + \mathcal{E}_{ij}, \\ E(\mathcal{E}_{ij} + v_i) &= 0, \end{aligned} \tag{1}$$

where $\mathcal{E}_{ij}, v_i$ are independent of X_{ij}, $\mathcal{E}_{ij}$ are i.i.d. across i and j, and v_i are i.i.d. across i. The random effect v_i induces stochastic dependence of the errors $\mathcal{E}_{ij} + v_i$ across j, for a given i. Note that v_i is introduced in a way that does not change the main structural part of the model – the conditional mean function $x'\alpha$.

We next describe a simple model of random effects for quantile regression. This model captures the dependence of response about the individual i, without affecting the conditional quantile function $Q_{Y_{ij}|X_{ij}}$.

Model 2 (M2a) For an unspecified function Φ

$$\begin{aligned} &Y_{ij} = Q_{Y|X_{ij}}(U_{ij}), \quad U_{ij} = \Phi(v_i, \mathcal{E}_{ij}), \\ &U_{ij} \overset{d}{=} U(0,1) \text{ conditional on } \{X_{ij}, j = 1, ..J\}, \end{aligned} \tag{2}$$

v_i are independent across i, $\mathcal{E}_{ij}$ are independent across i and j. For $W_i \equiv \{(Y_{ij}, X_{ij}), j \le J\}$, $\{W_i, i \le n_0\}$ is an i.i.d. triangular sequence, and $\{Y_t, X_t, t \le n\}$ is a stationary triangular sequence.

Here, v_i is the "random effect" that reflects an interdependence of data about "individual" i. **M2a** easily nests the mean regression model (1). Indeed, denoting by F the distribution function of $v_i + \mathcal{E}_{ij}$,

$$\begin{aligned} Q_{Y|X}(\tau) &= X'\alpha + F^{-1}(\tau), \\ U_{ij} &\equiv F(v_j + \mathcal{E}_{ij}), \end{aligned}$$

By taking e.g. $U_{ij} \equiv F(v_{1i} + v_{2i} \cdot \mathcal{E}_{ij})$, $v_i = (v_{1i}, v_{2i})$ one generates even more complicated random effects forms. The factors v_i influence location and scale of individual errors $\mathcal{E}_{ij} \equiv F^{-1}(U_{ij})$. One can proceed with further examples to show that **M2a** is capable of capturing a wide variety of dependence.

M2a leads to a simple variance of estimators of $Q_{Y|X}$, since it implies that the "errors" $1(Y_{ij} \le Q_{Y|X_{ij}}(\tau)) \equiv 1(U_{ij} \le \tau)$ are mutually correlated across j with a constant correlation $r(\tau)$, and uncorrelated across i. Parameter $r(\tau)$ has interesting structural meanings as well.

2.2.3 Panel Data with General Dependence

More general models for panel data were developed. See, for example MaCurdy (1982), (2001). Many of them intend to capture various forms of temporal dependence in the error structure. To this end, consider the model

Model 3 (M2b)

$$\begin{aligned} &Y_{ij} = Q_{Y|X_{ij}}(U_{ij}) \\ &U_{ij} \text{ is independent of } X_{ij}. \end{aligned} \tag{3}$$

For a given i, (U_{ij}, X_{ij}) are possibly dependent but stationary across j (This is to simplify notation). For a given j, (U_{ij}, X_{ij}) are assumed to be independent across i. Furthermore, write $\mathcal{W}_T \equiv \{W_i \equiv \{(Y_{ij}, X_{ij}), j \leq J\}, i \leq n\}$. Note that by construction, $\{W_i\}$ is an i.i.d sequence and $\{Y_t, X_t\}$ is stationary. This assumption is often made in the applied regression analysis of panel data (see MaCurdy, 2001)[1]. Clearly such a model nests **M2a** as a special case.

2.2.4 Weakly Dependent Data

Denoting $W_t = (Y_t, X_t)$, sequence $\{W_t\}$, is assumed to satisfy a certain weak-dependence condition, called strong mixing (see Doukhan (1994) for definition). Mixing insures that once two events are separated far enough in time, they are almost independent.

Model 4 (M3) Suppose for every n, we observe $\{W_t, 1 \leq t \leq n\}$, a chain in the sequence $\{W_t\}$ defined on probability space $(\Omega_n, \mathcal{F}_n, P_n)$. Let $\mathcal{F}_n^{l,m}$ be the σ -algebra generated by $\{W_t, l \leq t \leq m\}$, and define the corresponding α-mixing sequence by

$$\alpha(k) = \lim_n \sup \sup_{A,B} |P_n(A \cap B) - P_n(A)P_n(B)|,$$

where A and B vary over the σ-fields $\mathcal{F}_n^{-\infty,l}$ and $\mathcal{F}_n^{l+k,\infty}$, respectively. $\{W_t\}$ is said to be α-mixing or strong mixing if $\alpha(k) \to 0$ as $k \to 0$.

Such data sequences lend themselves to stochastic limit laws under conditions mentioned in the appendix. The class of mixing processes is fairly broad.

[1] Typically, the data (such as wages) are differenced to obtain stable series.

For example, ARMA and GARCH processes with continuously distributed stationary innovations and bounded variances are strongly mixing (with geometrically decaying coefficients) (Doukhan, 1994). It is possible to go beyond these mixing results to the degree the key stochastic limit laws hold, as posed in the appendix.

3 Quantile Regression under Misspecification

Section 3.1 discusses the quantile regression estimator of Koenker and Bassett (1978). Section 3.2 presents the assumptions we need to obtain large sample results of Section 3.3. In Section 3.3, we show an asymptotic Gaussianity of the regression quantile coefficient process $(\hat{\beta}(\tau), \tau \in [\epsilon, 1-\epsilon])$.

3.1 The Quantile Estimator

For simplicity, the data $\{W_t = (Y_t, X_t)\}_{t=1}^n$ will be treated as an ergodic stationary sequence defined on the complete probability space $(\Omega, \mathcal{F}, P)$. E_P denotes the expectation with respect to P, and E_n denotes the expectation with respect to the empirical measure: $E_n f \equiv \frac{1}{n}\sum_{t=1}^n f(W_t)$. The set of quantile indices of interest is given by $\mathcal{T} \equiv [\epsilon, 1-\epsilon]$, $\epsilon \in (0, 1/2)$.

This paper focuses on the quantile regression estimator of Koenker and Bassett (1978), which minimizes the asymmetric least absolute deviation criterion:

$$\hat{\beta}(\tau) \in \underset{\beta \in \mathbb{R}^d}{\operatorname{argmin}}\ E_n \rho_\tau(Y - X'\beta), \tag{4}$$

where

$$\rho_\tau(u) = (\tau - 1(u \leq 0))u.$$

$\hat{\beta}(\tau)$ could be interpreted as an analog estimator. Under stated conditions, $\hat{\beta}$ converges in probability to $\beta(\tau)$ such that

$$\beta(\tau) = \underset{\beta \in \mathbb{R}^d}{\operatorname{argmin}}\ E_P \rho_\tau(Y - X'\beta), \tag{5}$$

if the latter is uniquely defined. The true conditional quantile function $Q_{Y|X}$ satisfies an analogous condition

$$Q_{Y|X} \in \operatorname{argmin}_f\ E_P \rho_\tau(Y - f(X)), \tag{6}$$

where the minimum is taken over the set of the measurable functions of X. Thus $X'\beta(\tau)$ is a convenient approximation of $Q_{Y|X}$, and $\hat{\beta}(\tau)$ is the estimate of the parameters of this approximation. The convenience ranges from such practical considerations as computability to a number of useful equivariance and robustness properties, as studied in Koenker and Bassett (1978).

3.2 Assumptions

(Y_t, X_t) is a stationary, ergodic triangular sequence on the probability space $(\Omega_n, \mathcal{F}_n, P_n)$. Define $E_M f \equiv \lim_{T\to\infty} E_P E_n f(Y_t, X_t)$. The parameter $\beta_n(\tau)$ is defined to solve the equation

$$E_{P_n}(\varphi_\tau(Y - X'\beta_n(\tau))X = 0$$

for each n, where $\varphi_\tau(u) \equiv (\tau - 1(u \le 0))$. Under conditions **M** and **L** (see Assumption 1 below), this is equivalent to $\beta_n(\tau)$ solving the population prediction problem: $\beta_n(\tau) = \text{argmin}_{\beta_n \in \mathbb{R}^d} E_{P_n}\rho_\tau(Y - X'\beta)$. $\beta(\tau)$ denotes the solution under measure M, equivalently denoted here as P_∞. Index τ belongs to $\mathcal{T} \equiv (\epsilon, 1-\epsilon)$, for $\epsilon > 0$.

Outer/inner probabilities P^*, P_* and stochastic equicontinuity are as in VanDerVaart. We will say that the process $\{l \mapsto v_n(l), l \in \mathcal{L}\}$ is *stochastically equi-continuous* (s.e) in $\ell^\infty(\mathcal{L})$ if $\forall \epsilon > 0$ and $\eta > 0, \exists \delta > 0$:

$$\limsup_{T\to\infty} P_n^* \left(\sup_{\rho(l,l')<\delta} |v_n(l) - v_n(l')| > \eta \right) < \epsilon$$

for some semi-metric ρ on $\mathcal{L}$, s.t. $(\mathcal{L}, \rho)$ is totally bounded.

Assumption 1

(**M**) *Uniformly in $n > n_o$, as well as $n = \infty$*

(i) $\beta_n(\tau)$ uniquely solves $E_{P_n}\varphi_\tau(Y - X'\beta)X = 0$, $\beta_n(\tau) \in$ interior Θ, $\forall \tau \in \mathcal{T}$, Θ is compact set of $\mathbb{R}^d$. $\beta_n(\tau)$ is continuous in τ on $\mathcal{T}$ (uniformly in n)

(ii) X_t is supported on $\mathbf{X}$, a compact set in $\mathbb{R}^d$, $Var(X)$ is positive definite.

(iii) $f_{Y|X}(y)$ is uniformly continuous and bounded in y, and $f_{Y|x}(x'\beta(\tau)) > 0$, uniformly in τ and $x \in \mathbf{X}$

(iv) $P_n^{[n]}$, the law of $(Y_t, X_t, t \le n)$ under P_n, is contiguous to $M^{[n]}$.

(**L**) *For all uniformly in n bounded measurable maps $(X,Y) \mapsto h_n(X,Y)$,*

$$\sqrt{n}(E_n h_n - E h_n) \xrightarrow{d} N(0, V_M(h_n, h_n)), \text{ where } V_M(h_n,h_n) = \lim_{n\to\infty} Var_M[\sqrt{n}E_n h_n]$$

(E) $\sqrt{n}\,[E_n 1(Y \leq X'\beta)g(X) - E_{P_n} 1(Y \leq X'\beta)g(X)]$ *is s.e. in* $\ell^\infty(\Theta \times \mathcal{G})$. $X \in \mathcal{G}$, *where* $\mathcal{G}$ *is a collection of test functions.*

M is the basic model setup. Assumption **L** requires applicability of CLTs and LLNs. **E** is easy to verify in applications. The class of functions $\mathcal{G}$ is used to form the over-identifying restrictions that serve as basis for the specification tests. All examples given here are considered to satisfy **E**.

3.3 Quantile Regression Large Sample Properties

Theorem 1 (Gaussianity of $\hat{\beta}_n(\cdot)$) *Under conditions (M) (L) (E), in* $\ell^\infty(\mathcal{T})$,

$$\sqrt{n}\left(\hat{\beta}_n(\cdot) - \beta_n(\cdot)\right) \Rightarrow z(\cdot),$$

where $z(\cdot)$ *is a Gaussian process with covariance function*

$$w_z(\tau, \tau') \equiv V(l(\tau), l(\tau')),$$

with $l(\tau) \equiv J_x^{-1}(\tau)\varphi_\tau(Y - X'\beta(\tau))X$, $J_x(\tau) \equiv E_{P_n} f_{Y|X}(X'\beta_n(\tau))XX'$.

Proof: See Appendix B.1.

The theorem requires an availability of suitable CLT and LLN, and a simple equicontinuity condition, which is extremely easy to verify in all applications which we consider. Hence it has a very wide applicability.

Theorem 1 holds under local or global misspecification. Suppose that the data is iid. Then, the asymptoic variance is:

$$w_z(\tau, \tau') \equiv J_x^{-1}(\tau)E_M XX'\varphi_\tau(Y - X'\beta(\tau))^2 J_x^{-1}(\tau)$$

If it further happens that either 1) $Q_{Y|X}(\tau) = X'\beta(\tau)$ P- a.s. (correct specification) or 2) We have local misspecification (see equation (7) below), then:

$$w_{0z}(\tau, \tau') \equiv J_x^{-1}(\tau)E_M XX'\tau(1-\tau)J_x^{-1}(\tau).$$

It is common practice to report a consistent estimate of $J_x^{-1}(\tau)EXX\tau(1-\tau)J_x^{-1}(\tau)$ as an estimate of the asymptotic variance of $\hat{\beta}(\tau)$. Unfortunately, it is typically the case that

$$\left[E_M XX'\varphi_\tau(Y - X'\beta(\tau))^2\right]_{ii} \neq \left[E_M XX\tau(1-\tau)\right]_{ii}$$

Therefore, the common practice may incorrectly state the variance of the estimator under (large) misspecification. Moreover, the pivotal re-sampling approaches that use an incorrect estimate of variance may provide invalid confidence intervals.

Limit variance under local misspecification. Under the regularity conditions posed, consider a sequence of models where

$$Q_{Y|X}(\tau) = X'\beta(\tau) + g(X,\tau)/\sqrt{n}, P_n \text{ - a.s.} \tag{7}$$

such that $P_n^{[n]}$, describing the law of the sequence $\{W_t, 1 \le t \le n\}$, is contiguous to a probability measure $P^{[n]}$, under which $Q_{Y|X}(\tau) = X_t'\beta(\tau)$. Additionally, assume $\sqrt{n}(\beta_n(\tau) - \beta(\tau)) \to c_0(\cdot)$, uniformly in $\mathcal{T}$, where $c_0(\cdot)$ is a fixed continuous function.

Then the variance operator simplifies to

$$w_{0z}(\tau,\tau') \equiv V_P(l_0(\tau), l_0(\tau')),$$

where $l_0(\tau) \equiv J_{0x}^{-1}(\tau)\varphi_\tau(Y - X'\beta(\tau))X$, $J_{0x}(\tau) \equiv E_P f_{Y|X}(X'\beta(\tau))XX'$. Although asymptotically $\sqrt{n}(\hat{\beta}_n(\tau) - \beta_n(\tau))$ is centered at zero, it is the case that

$$\sqrt{n}(\hat{\beta}_n(\cdot) - \beta(\cdot)) \Rightarrow c_0(\cdot) + z_0(\cdot),$$

where $z_0(\cdot)$ is the centered Gaussian process with covariance function $w_{0z}(\tau,\tau')$.

Limit variance in panel models. For the case of the global misspecification, covariance $w_z(\tau,\tau')$ takes the following form in the panel data model **M2b** (and **M2a**):

$$w_z(\tau,\tau') \equiv J_x(\tau)\Bigg\{\sum_{j=1}^{J}\Bigg[E_M(l_{ij}(\tau)l_{ij}(\tau')') + \sum_{k>j}^{J}\Big[E_M l_{ij}(\tau)l_{ik}(\tau')' + E_M l_{ik}(\tau')l_{ij}(\tau)'\Big]\Bigg]\Bigg\}J_x(\tau')'.$$

For the case of local misspecification, replace $l_{ij}(\tau) \equiv \varphi_\tau(Y_{ij} - X_{ij}'\beta_n(\tau))\, X_{ij}$, by $\varphi_\tau(Y - X'\beta(\tau))X_{ij}$, and $J_x^{-1}(\tau)$ by $J_{0x}^{-1}(\tau)$.

In the locally misspecified case, for **M2a** it is further the case:

$$w_{0z}(\tau,\tau') \equiv J_{0x}^{-1}(\tau)\Bigg\{\sum_{j=1}^{J}\Bigg[\{\min(\tau,\tau') - \tau\tau'\}E_P(X_{ij}X_{ij}') + \{\rho(\tau,\tau') - \tau\tau'\}\sum_{k>j}^{J}\Big[E_P X_{ij}X_{ik}' + E_P X_{ik}X_{ij}'\Big]\Bigg]\Bigg\}J_{0x}^{-1}(\tau')',$$

where for $j \neq j'$

$$\rho(\tau, \tau') \equiv E1(U_{ij} \leq \tau) \cdot 1(U_{ij'} \leq \tau').$$

$\rho(\tau, \tau')$ could be estimated by computing sample covariances between $1(Y_{ij} \leq X_{ij}'\hat{\beta}_n(\tau))$ and $1(Y_{ij} \leq X_{ij'}'\hat{\beta}_n(\tau))$.

If the panel is unbalanced (i.e. for a given i, $j \leq J_i \leq K$), insert J_i in place of J and put $\lim_{n_0 \to \infty} n_0^{-1} \sum_{i \leq n_0}$ in front of $\{\cdot\}$.

3.4 On the Need for Specification Tests

The above discussion provides a strong motivation for specification tests that have power against $\sqrt{n}$ alternatives. They are needed to construct models that can be thought of as locally misspecified in the usual Pitman sense. Pitman and Le Cam proximity notions or statistical experiments formalize the concept of small misspecification and allow the decision-maker to proceed using his model as local to the true one. Therefore, (1) if the conditional quantile model is locally misspecified, then it is in the proximity of a genuine quantile model; (2) Local misspecification simplifies variance operators and rationalizes the otherwise unrobust conventional inference; (3) Model reduction hypotheses are often considered, e.g. in Koenker and Xiao (2000), assuming that the model is correctly specified; such inference is validated also when the misspecification of the model is $\sqrt{n}$ local.

4 Specification or Goodness-of-Fit Tests

4.1 Tests Based on Instrumental Variables

By Theorem 1, under correct specification or global misspecification, $\hat{\beta}_n(\tau)$ converges uniformly in $\mathcal{T}$ in probability to the $\beta_n(\tau)$ that solves the equation

$$E_{P_n}\ \varphi_\tau(Y - X'\beta_n(\tau)) \cdot X = 0, \text{ for each } \tau \text{ in } \mathcal{T}.$$

The hypothesis of correct specification,

$$H_0 : Q_{Y|X}(\tau) = X'\beta_n(\tau), \text{ for each } \tau \text{ in } \mathcal{T}, \text{ with prob. } P_n\ 1,$$

is equivalent to the condition

$$E_P\ \varphi_\tau(Y - X'\beta_n(\tau)) \cdot Z = 0, \text{ for each } \tau \text{ in } \mathcal{T}, \tag{8}$$

for any measurable function $Z = g(X)$. Indeed, (8) will not hold for some $g(X)$, if $X'\beta_n(\tau)$ differs from $Q_{Y|X}(\tau|X)$ with positive probability for some τ. Since we assumed that $Q_{Y|X}(\tau)$ is a strictly monotone continuous function of τ, $P[Y < v|X] - \tau = 0$ iff $Q_{Y|X}(\tau) = v$ P-a.s. Thus, under the alternative to H_0, $Var(P[Y < X'\beta_n(\tau)|X] - \tau) > 0$ for some τ. Hence for $Z = P[Y < X'\beta_n(\tau)|X] - \tau$, $E\varphi_\tau(Y - X'\beta_n(\tau))Z = Var(Z) \neq 0$, for some τ.

Therefore, a basic specification test can examine the validity of (8) by considering the finite sample approximation of (8):

$$\hat{\mu}(\tau, g) = E_n\varphi_\tau(Y - X'\hat{\beta}_n(\tau))g(X),$$

for a collection of test functions $g(X)$. These test functions g should be carefully chosen to reveal that $P[Y < X'\beta_n(\tau)|X] \neq \tau$, when H_0 fails. The discussion of this choice is postponed until we get through the basic material.

Next consider a very simple statistic that will be a basic building block for more complex statistics that examine the approximating ability of quantile models:

$$\hat{S}(\tau, g) = n\hat{\mu}(\tau, g)'\hat{W}(\tau, g)\hat{\mu}(\tau, g),$$

where $\hat{W}(\tau, g)$ is chosen to yield a standard distribution under H_0. Given regularity conditions posed later, under H_0,

$$\sqrt{n}\hat{\mu}(\tau, g) \xrightarrow{d} N(0, \Omega(\tau, g)).$$

Choose $\hat{W}(\tau, g) \xrightarrow{p_n} \Omega^{-}(\tau, g)$, so that under H_0

$$\hat{S}(\tau, g) \xrightarrow{d} \chi^2_k, \quad k \equiv \text{ rank } \Omega(\tau, g),$$

for a given τ and g. This leads to a critical region of the form $\{\hat{S}(\tau, g) > c\}$, where c is the $1 - \alpha$-th quantile of the χ^2_k variable. For a given quantile index τ, this is the simplest form of a goodness-of-fit test.

Three degrees of complexity will be added to this basic formulation. First, we wish to examine whether (8) is valid for all τ in $\mathcal{T}$, which demands a Kolmogorov-Smirnov (or Cramer-von-Misses) formulation of the test statistic. For instance,

$$\sup_{\tau \in \mathcal{T}} \hat{S}(\tau, g).$$

Second, the choice of g will be addressed in two different manners, allowing g to vary among finite and infinite collections of functions $\mathcal{G}$, leading to statistics such as

$$\sup_{\tau \in \mathcal{T}, g \in \mathcal{G}} \hat{S}(\tau, g).$$

The following three sections introduce the formal assumptions, formulate the statistics of the above kind, and describe their sampling theory and simple computational methods. The last subsection introduces a method of estimating the critical values for the tests using subsample bootstrap,and studies the properties of the resulting test.

4.2 General Specification Tests.

First, we consider cases when the number of test functions g is fixed - i.e., when the number of test functions is small relative to the sample size.

4.2.1 Finite Number of Test Functions g.

Denote by g a function that maps $g : \mathbf{X} \mapsto \mathbb{R}^k$. If we have several vector valued functions $g_m, m \leq h$, they can be arranged into one function $\mathcal{G} = (g_i, i \leq h)$.

We call

$$\hat{S} \equiv \Big(\hat{S}(\tau, g), \ \tau \in \mathcal{T} \Big)$$

the *specification test process.* $\hat{S}$ is a stochastic process or function in $\ell^\infty(\mathcal{T})$, the metric space of bounded functions, equipped with the sup metric. Similarly, we call

$$\hat{\mu} \equiv (\ \hat{\mu}(\tau, g), \tau \in \mathcal{T} \ )$$

the *score process*; $\hat{\mu}$ also takes its values in $\ell^\infty(\mathcal{T})$. To avoid measurability problems caused by the discontinuities in the sample paths of $\hat{\mu}$, and $\hat{S}$, we use stochastic convergence in the sense of Hoffman-Jorgenson (See Vandervaart).

Define

$$\mu(\tau, g) \equiv E_P \varphi_\tau (Y - X'\beta_n(\tau)) g(X).$$

In reference to the general hypothesis H_0, consider the following null, alternative, and local alternative hypotheses

$$\begin{aligned}
&\mathbf{H}_0 : \mu(\tau, g) = 0, \quad \forall \tau \in \mathcal{T}, \forall g \in \mathcal{G} \\
&\mathbf{H}_A : \mu(\tau, g) \neq 0, \quad \exists \tau \in \mathcal{T}, \exists g \in \mathcal{G} \\
&\mathbf{H}_{An} : \mu(\tau, g) = \eta(\tau, g)/\sqrt{n}, \quad \forall \tau \in \mathcal{T}.
\end{aligned}$$

$\eta(\tau, g)$ is a continuous function of τ. Generally, $\mathbf{H}_0$ is less general than H_0 in the sense that $\mathbf{H}_0 \subset H_0$, but rejection of $\mathbf{H}_0$ entails rejection of H_0. By picking a suitable g, it is possible to insure equivalence, as discussed later.

We focus on the Kolmogorov-Smirnov test statistic of the form:

$$KS_n \equiv \sup_{\tau \in \mathcal{T}, g \in \mathcal{G}} S(\tau, g),$$

but any other test statistics, which are continuous functionals $\mathbf{f} : \ell^\infty(\mathcal{T}) \mapsto \mathbb{R}_+$ s.t. $\mathbf{f}(s) = 0$ iff $s = 0$, are allowed, particularly the Cramer-Von-Misses test. Since we are primarily interested in KS statistics, and for brevity sake, we restrict attention to the cases when $\mathbf{f}$ has the property: $\mathbf{f}(\alpha \cdot s + \beta - \alpha_1^c \mathbf{f}(s)| \leq c_2 \beta$, for scalars $\alpha, \beta, c > 0$.

The following theorem states the behaviour of the test process $\hat{S}$ that determines that of $\sup_{\tau \in \mathcal{T}} S(\tau, g)$ and other functionals $\mathbf{f}(S)$. We also define an important process

$$\hat{\mathcal{S}}(\tau, g) = n\left(\hat{\mu}(\tau, g) - \mu(\tau, g)\right)' \hat{W}(\tau, g)\left(\hat{\mu}(\tau, g) - \mu(\tau, g)\right),$$

$\hat{\mathcal{S}} \equiv \{\hat{\mathcal{S}}(\tau, g)\}$, $\hat{W} = \{\hat{W}(\tau, g)\}$. We will need the statistic $\hat{\mathcal{S}}(\tau, g)$ later when we discuss bootstrapping. This statistic will mimic the null behaviour of $\hat{S}$, even when the null is false.

Theorem 2 *Suppose $\hat{W}(\cdot) \xrightarrow{p_n} W(\cdot)$, $W(\tau, g)$ is uniformly p.d.*
*1. Under conditions **M L E** and $\mathbf{H}_{A_n}$, in $\ell^\infty(\mathcal{T} \times \mathcal{G})$*

$$\begin{aligned} \sqrt{n}\hat{\mu} &\Rightarrow \xi_0 + \eta, \\ \hat{S} &\Rightarrow S_\infty \equiv (\xi_0 + \eta)' \cdot W \cdot (\xi_0 + \eta), \end{aligned}$$

ξ_0 is a zero-mean P-Brownian Bridge with covariance function: $\omega_{0\mu}(\tau, g; \tau', g') \equiv V_P(j_{0g}(\tau), j_{0g'}(\tau))$, where $j_{0g}(\tau) \equiv \varphi_\tau(Y - Q_{Y|X}(\tau)) \cdot \left[g(X) - J_{0g}(\tau) J_{0x}^{-1}(\tau) X\right]$, $J_{0g}(\tau) \equiv E_P f_{Y|X}(Q_{Y|X}(\tau)) g(X) X'$, $J_{0x}(\tau) \equiv E_P f_{Y|X}(Q_{Y|X}(\tau)) X X'$, where $Q_{Y|X}(\tau) = X'\beta(\tau)$ P-a.s.
*2. Under conditions **M L E** and $\mathbf{H}_A$, in $\ell^\infty(\mathcal{T} \times \mathcal{G})$, $\frac{1}{n}\hat{S} \xrightarrow{p_n} \mu \cdot W \cdot \mu$*

$$\begin{aligned} \sqrt{n}(\hat{\mu} - \mu) &\Rightarrow \xi, \\ \hat{\mathcal{S}} &\Rightarrow \mathcal{S}_\infty \equiv (\xi)' \cdot W \cdot (\xi), \end{aligned}$$

ξ is a zero-mean P-Brownian Bridge with covariance f-n: $\omega_\mu(\tau, g, \tau', g') \equiv V_M(j_g(\tau), j_{g'}(\tau'))$, where $j_g(\tau) \equiv \varphi_\tau(Y - X'\beta_n(\tau)) \cdot \left[g(X) - J_g(\tau) J_x^{-1}(\tau) X\right]$, $J_g(\tau) \equiv E_M f_{Y|X}(X'\beta_n(\tau)) g(X) X'$. Under $\mathbf{H}_{A_n}$, $\xi \stackrel{d}{=} \xi_0$.

Proof: See Appendix B.2.

By the continuous mapping theorem, a trivial corollary of this theorem is that the KS test is consistent and has power against $\sqrt{n}$ alternatives.

4.2.2 Choosing g? Large Number of Functions g

In order to approximate H_0 with $\mathbf{H}_0$, one needs to increase the number of functions with the sample size. Potentially, this is a caveat since we assumed so far that the number of test functions g in $\mathcal{G}$ is fixed relative to the sample size n. However, if the class of functions $\mathcal{G}$ has good complexity properties, Theorem 2 will remain valid. Indeed, if the functional class

$$\left\{1(Y \leq X'\beta_n(\tau))g(X), \quad \tau \in \mathcal{T}, g \in \mathcal{G}\right\}$$

is Donsker, Theorem 2 will apply. Thus we need a class of functions $\mathcal{G}$ that is both Donsker *and* is able to reveal that $P(Y \leq X'\beta_n(\tau)|X) \neq \tau$.

The search for such functions leads to the remarkable work of Bierens (1990), Sithicombe and White (1998), and Bierens and Ploberger (1997). Particularly, consider

$$\mathcal{G} \equiv \{g_\xi(x) = w(\xi'\Phi(x)), \quad \xi \in \Xi\}$$

where w is an analytic function, except a polynomial, Ξ is an bounded subset of Euclidian space with positive Lebesgue measure, e.g. $(0,1)^d$, and Φ is bounded bijective mapping. Allowed $w(\cdot)$ include $\exp(\cdot)$, $\cos(\cdot)$, $\sin(\cdot)$, and others.

Lemma 1 *Assume condition* ***M****. Let $\Xi' \equiv \cup_{\tau\in\mathcal{T}}\Xi(\tau)'$ have Lebesgue measure zero and is nowhere dense in Ξ. Then, under H_0, $E(1(Y \leq X'\beta_n(\tau)) - \tau)w_\xi(X) = 0$ for all $\xi \in \Xi, \forall\tau \in \mathcal{T}$; otherwise, $E(1(Y \leq X'\beta_n(\tau)) - \tau)w_\xi(X) = 0$ for all $\xi \in \Xi(\tau)', \forall\tau \in \mathcal{T}$.*

For a fixed τ and $\Xi(\tau)$, this is a result of Bierens (1990) and Stinchcombe and White (1998), upon changing notation. To show that $\Xi \equiv \cup_{\tau\in\mathcal{T}}\Xi(\tau)'$ satisfies the above properties, in any nonempty open ball in $\mathcal{T} \times \Xi$, there is $\xi' \notin \Xi(\tau')$ which implies that there exits a small open neigborhood $\mathcal{N}$ of (ξ', τ') of positive Lebesgue measure s .t. for all $(\xi, \tau) \in \mathcal{N}$, $E(1(Y \leq X'\beta_n(\tau)) - \tau)w_\xi(X) \neq 0$. This is by the uniform continuity of $E(1(Y \leq X'\beta_n(\tau)) - \tau)w_\xi(X)$ in ξ and τ in $\mathcal{T} \times \Xi$.

A drastic conclusion of this is that, in principle, we could randomly pick one test function and have a consistent test, as in Bierens (1990). Obviously, such a test may not have good power in smaller samples. This also leaves a researcher too much flexibility that can be enjoyed for the sake of "snooping," i.e. searching for a test that "accepts" the desired hypothesis. A good test should prevent such a possibility. These considerations lead to Kolmogorov-Smirnov and Cramer-Von Misses type tests or their approximations. Thus, we need a theory that explicitly accounts for the large number of test functions.

Theorem 3 *Assume $\eta(\tau, g)$ is uniformly continuous in $\ell^\infty(\mathcal{T} \times \mathcal{G})$. Then under conditions **(M)**, **(L)**, **(E)**, Theorem 2 holds for the case when $\mathcal{G}$ is possibly infinite.*

Proof: See Appendix B.3.

It also makes sense to include the functions in $\mathcal{G}$ that may increase the finite sample power of the test. The finite set of polynomials of X is one such example.

4.3 Resampling the Specification Tests

Here we demonstrate how one can obtain the critical values of the subsample bootstrap (see Politis et. al (1998)). This method overcomes the Durbin problem.

The basic idea is to use a mimicking process, which is $\hat{S}$, to correctly mimic the distribution of the actual test process S under the null, when the actual model is in the $1/\sqrt{n}$ neighborhood of the null model. The subsampling bootstrap is used to estimate the distribution of S to construct the critical value. This critical value leads to the test of correct size and entails no loss of power and does not require estimating non-parametric nuisance parameters (the common drawbacks of Khamaladization). Under large deviations (greater than $1/\sqrt{n}$) of the model from the null model, $\bar{S}$ no longer correctly mimics the distribution of S under the null, but the critical value remains bounded while S tends to infinity, insuring consistency of the test.

The particular resampling method employed here is the subsampling bootstrap. This method has several advantages – both conceptual and theoretical, as has been recently emphasized in Sakov and Bickel(2000) for another version of the sub-sample bootstrap ("m out of n").

The basic idea of the subsample bootstrap is to approximate the sampling distribution of a statistic based on the values of the statistic computed over smaller subsets of the data . After suitable normalization, recomputed values of the statistic are used to approximate the sampling distribution.

Now we operate with the data models **M1** - **M3** introduced in section 2. For all these data models, we have defined "units" of data $\{W_t, t \leq n\}$ that suitably combine one or several pairs of (Y_t, X_t). In particular, for panel models **M2a** -**M2b**, they are combined in the way to make $\{W_t\}$ an i.i.d. sequence. For the basic i.i.d model **M1** and stationary time series model **M3**, there is no re-combining: $W_t = (Y_t, X_t)$.

For cases when $\{W_t\}$ is i.i.d., the test statistic is computed over all subsamples of size b; the number of such subsets B_n, which we index by i, is "n choose b". For cases when $\{W_t\}$ is stationary, the statistic is computed only over $B_n = n - b + 1$ subsets of size b of the form $\{W_i, ..., W_{i+b-1}\}$. To save computational time, one can compute the statistic over a smaller number of subsets (or bootstrap draws), as long as $B_n \to \infty$ as $n \to \infty$. (See section 2.5 in Politis et al.).

Denote by θ_n the statistic $\mathbf{f}(\hat{S})$, computed over the whole sample; and by $\theta_{b,n,i}$ the statistic computed over the i-th subset of data, using $r_b = b$ in place of n :

$$\hat{S}_{n,b,i}(\tau, g) \equiv r_b \left(m_{n,b,i}(\tau, g)\right)' \hat{W}(\tau, g) \left(m_{n,b,i}(\tau, g)\right),$$
$$m_{n,b,i}(\tau, g) \equiv \left(\hat{\mu}_{n,b,i}(\tau, g) - \hat{\mu}(\tau, g)\right)$$

Let

$$G_n(x, P) \equiv Prob_P\{\mathbf{f}(\mathcal{S}) \leq x\},$$
$$H_n(x, P) \equiv Prob_P\{\mathbf{f}(\hat{S}) \leq x\}.$$

From Theorem 2-3, $H_n(x, P)$ converges in law to $H_0(x)$ under $\mathbf{H}_0$ and $H_A(x)$ under $\mathbf{H}_{AT}$. $H_0(x)$ and $H_A(x)$ differ if the test has power, as in the case of the KS test.

Also we note from Theorem 2-3 that $G_n(\cdot, P)$ converges in distribution to $G(\cdot)$, where $G(\cdot) = H_0(\cdot)$ under $\mathbf{H}_0$ and under $\mathbf{H}_{AT}$. Therefore, we need to consistently approximate $G_n(\cdot, P)$ in order to approximate $H_0(\cdot)$, at least under local alternatives. The subsample bootstrap accomplishes this. Estimate $G_n(x, P)$ by

$$\hat{G}_{n,b}(x) = B_n^{-1} \sum_{i=1}^{B_n} 1\{\theta_{n,b,i} \leq x\}$$

Using this estimated sampling distribution, the critical value for the test is obtained as the $1-a$-th quantile of $G_{n,b}(\cdot)$:

$$c_{n,b}(1-\alpha) = \hat{G}_{n,b}^{-1}(1-\alpha).$$

Finally, the size α test rejects $\mathbf{H}_0$ if $\mathbf{f}(\hat{S}) > c_{n,b}(1-\alpha)$.

The following theorem shows that this test will have the same power as a test where the critical value is known.

Theorem 4 *For models **M1-M3**, under the conditions of Theorems 1-3 and assumptions **(M)**, **(L)**, **(E)**, as* $b/n \to 0, b \to \infty, n \to \infty$, $B_n \to \infty$,

(i) Under $\mathbf{H}_0$, *if* H *is continuous at* $H_0^{-1}(1-\alpha)$:

$$c_{n,b}(1-\alpha) \xrightarrow{P} H_0^{-1}(1-\alpha)$$

$$Prob_P(\mathbf{f}(\hat{S}) > c_{n,b}(1-\alpha)) \to \alpha.$$

(ii) under $\mathbf{H}_A$, *the test is consistent when* $\mathbf{f}(\hat{S})$ *is KS or CM and*

$$c_{n,b}(1-\alpha) \xrightarrow{p_n} G^{-1}(1-\alpha)$$

$$Prob_{P_n}(\mathbf{f}(\hat{S}) > c_{n,b}(1-\alpha)) \to 1.$$

(iii) Under $\mathbf{H}_{An}$, *if* $H_A(x)$ *is continuous at* $H_0^{-1}(1-\alpha)$,

$$c_{n,b}(1-\alpha) \xrightarrow{p_n} H_0^{-1}(1-\alpha)$$

$$Prob_{P_n}(\mathbf{f}(\hat{S}) > c_{n,b}(1-\alpha)) \to Prob_P(\mathbf{f}(S_\infty) > H_0^{-1}(1-\alpha)).$$

(iv) $H_0(x)$, $H_A(x)$ *and* $G(x)$ *are absolutely continuous when the covariance functions* ω_m, $\omega_{\mu 0}$ *in Theorem 2 are nondegenerate.*

Proof: See Appendix B.4.

Thus the KS test with the sub-sampled critical value is consistent, asymptotically unbiased, and has the *same power* as the KS test with the known critical value.

5 Monte Carlo

It is both impractical and infeasible to compute $sup_{\tau \in \mathcal{T}, g \in \mathcal{G}} \hat{S}(\tau, g)$. Therefore, we compute $\hat{S}(\tau, g)$ over grids $\mathcal{T}_n$, $\mathcal{G}_n$. Our Monte Carlo looks at three model specifications:

- **Model 1** $Y = X_1 + X_2 + \cdots + X_k + \epsilon$.
- **Model 2** $Y = X_1 + X_2^2 + \cdots + X_k^k + \epsilon$.
- **Model 3** $Y = sin(X_1) sin(X_2) \ldots sin(X_k)\epsilon$.

In all three models, ϵ and $X_j, j = 1, \ldots, k$ are $N(0, I_{nxn})$. For each model, we run a Monte Carlo for 8 sets of parameters (see Table 1). For each set of parameters, we have 500 realizations. For each realization, we draw $r = n/5$ sub-samples, each of size $b = n/3$. In all cases, the set of functions g consists of $(cos(X'\psi_1), \ldots, cos(X'\psi_d))$, where $\psi_h, h = 1, \ldots, d$ is a vector consisting of k independent U(0,1) draws. We consider $\tau \in \mathcal{T} = (.1, .3, .5, .7, .9)$.

Each realization consists of the following steps:

- We draw $X_j, j = 1, \ldots, k$ and ϵ, and create Y as specified by the model. $W = \{X, Y\}$.
- We calculate $\hat{\mu}(\tau) \forall \tau \in \mathcal{T}$. We compute $\hat{S}_{fs} = sup_{\tau \in \mathcal{T}} \hat{S}(\tau)$, where $\hat{S}(\tau) = n\hat{\mu}(\tau)'\hat{W}\hat{\mu}(\tau)$, with $\hat{W} = I$.
- We draw r sub-samples, each of size b, from W. For each of the r sub-samples, we calculate $\hat{\mu}(\tau)_r \forall \tau \in \mathcal{T}$. We compute $\hat{S}(\tau)_r = b(\hat{\mu}(\tau)_r - \hat{\mu}(\tau))'\hat{W}(\hat{\mu}(\tau)_r - \hat{\mu}(\tau))$, with $\hat{W} = I$. The set of statistics $\{\hat{S}(\tau)_r, \tau \in \mathcal{T}, \forall r\}$ is used to construct the critical values for $\hat{S}_{fs}$.

Table 3 tabulates how often our full-sample statistic is rejected in the 500 realizations under the null hypothesis $Q_{Y|X}(\tau) = X'\beta(\tau)$. We see higher rejection levels for Models 2-3 over Model 1. This is precisely what we expect as Model 1 has the correct formulation $Q_{Y|X}(\tau) = X'\beta(\tau)$. We also find the rejection levels increasing with n, k across models.

6 Conclusion

We have developed a means of testing the specification of quantile regression, under the null hypothesis that the true model is a linear location-scale model.

The specification tests designed handle a wide range of data types - we have considered the simple case of i.i.d. in our Monte Carlo example. Our Monte Carlo gives a step-by-step procedure for carrying out a specification test in practice, and finds that our test does well at rejecting models which diverge from the $Q_{Y|X}(\tau) = X'\beta(\tau)$ formulation.

Part III

Bias of GMM with a General Moment Weighting Matrix

1 Introduction

GMM estimation is a popular method of estimation due to its consistency and well-defined notion of optimality. However, the GMM estimator suffers from substantial finite sample bias (Alonso-Borrego & Arellano (1996)). The exact form of this bias is calculated in Newey and Smith (2003).

We focus on computing the bias of GMM estimation when the vector of moment conditions is weighted by a matrix A which is unspecified, except for some general conditions. The bias is computed by taking a higher-order Taylor expansion of the moment conditions. We show that the GMM bias depends on the influence function $\hat{\Psi}(Z)$ of $\hat{A}$, and that one means of reducing the bias is to restrict $\hat{A}$. We use our results to look at some alternative GMM estimators which fit into our framework. One such estimator is Arellano's IVE estimator (Arellano (2003)).

Arellano's IVE estimator is designed for dynamic panel data with fixed effects. The bias of GMM in this setting has been recently studied. Hahn, Hausman, and Kuersteiner (2002) computed the bias of GMM in this setting using a second order approach. Chapter 1 also looked at GMM bias in this setting.

Arellano's IVE estimator differs from the standard GMM estimator considered in Hahn, Hausman, and Kuersteiner (2002) and in Chapter 1. Both estimators will have similar robustness properties for fixed T and large N. However, the IVE estimator is immune to the asymptotic biases that result when T is not fixed. We will look at the bias of Arellano's IVE estimator and show that, as T grows, it does better than the standard GMM estimator.

2 The GMM Model and its Bias

Consider a standard GMM model with a fixed number of moment restrictions. Let z_i $(i = 1, \ldots, n)$ be i.i.d. observations on a data vector z. Let β be a $KX1$

parameter vector and let our moment conditions be given by $g(z,\beta)$, a $JX1$ vector. At the true parameter β_0, we require that $g(z,\beta_0)=0$.

The two-step GMM estimator of Hansen (1982) is given by:

$$\hat{\beta}_{GMM} = argmin_{\beta}\hat{g}(\beta)'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\beta) \tag{1}$$

where $\Omega(\beta) = (1/n)\sum_{i=1}^{n} g_i(\beta)g_i(\beta)'$ and $\tilde{\beta} = argmin_{\beta}\hat{g}(\beta)\hat{W}^{-1}\hat{g}(\beta)$, where $\hat{W}$ is an initial weighting matrix.

Recall the results of Newey and Smith (2003) introduced in Part 1, Section 2. They are included here again for convenience. Newey and Smith derived stochastic expansions for this two-step GMM estimator. Under identification and regularity assumptions, as well as conditions on the initial weighting matrix $\hat{W}$, they find the asymptotic bias of GMM to be given by:

$$\begin{aligned}
Bias(\hat{\beta}_{GMM}) &= B_I + B_G + B_{\Omega} + B_W \\
B_I &= H(-a + E[G_i H g_i])/n \\
B_G &= -\Sigma E[G_i' P g_i]/n \\
B_{\Omega} &= HE[g_i g_i' P g_i]/n \\
B_W &= H\sum_{j=1}^{K}\tilde{\Omega}_{\beta_j}(H_W - H)'e_j/n
\end{aligned}$$

where $H=\Sigma G'\Omega^{-1}$, $H_W=(G'W^{-1}G)^{-1}G'W^{-1}$, $G=E(G_i)$, $P=\Omega^{-1}-\Omega^{-1}G\Sigma G'\Omega^{-1}$, $\tilde{\Omega}_{\beta_j}=E\left[\partial\{g_i(\beta_0)g_i(\beta_0)'\}/\partial\beta_j\right]$, $\Sigma=(G'\Omega^{-1}G)^{-1}$, and a is an m-vector such that:

$$a_j = tr(\sum E\left[\partial^2 g_{ij}(\beta_0)/\partial\beta\partial\beta'\right])/2, \;\; j=1,\ldots,m$$

B_I is the asymptotic bias for a GMM estimator with the optimal linear combination $G'\Omega^{-1}g(z,\beta)$. B_G arises from estimating $G=E(G_i)$. This is zero if G_i is constant, but is generally non-zero if there is endogeneity. B_{Ω} arises from estimating Ω; this is zero if the third moments are zero, but is generally non-zero. B_W arises from the choice of $\hat{W}$, the first step weighting matrix. It is zero if W is a scalar multiple of Ω.

3 GMM With a General Weight Matrix

The first order condition for the GMM estimator introduced above is given by:

$$[\sum_{i=1}^{n} G_i(\hat{\beta}_{GMM})/n]'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\hat{\beta}_{GMM}) = 0 \tag{2}$$

The $\hat{\beta}_{GMM}$ estimator can then be thought of as the vector β that solves $[\sum_{i=1}^{n} G_i(\beta)/n]'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\beta) = 0$. Here, $[\sum_{i=1}^{n} G_i(\beta)/n]'\hat{\Omega}(\tilde{\beta})^{-1}$ is a KXJ matrix that provides optimal weights for the $\hat{g}(\beta)$ moment vector.

What if we consider a more general KXJ matrix? Suppose $\hat{A}$ is KXJ such that $\hat{A} = A + \frac{1}{n}\sum_{i=1}^{n}\psi(z_i) + O_p(n^{-1})$, $E(\psi(z_i)) = 0$. Then we have a GMM estimator $\hat{\alpha}_{GMM}$ which is the α vector that satisfies $\hat{A}\hat{g}(\alpha) = 0$. Now we can see that the GMM estimator defined in Eq.(1) is a special case of this GMM estimator. Setting $\hat{A} = \hat{G}'\hat{\Omega}^{-1}$ gives us $\hat{\alpha}_{GMM} = \hat{\beta}_{GMM}$.

Definition 1 Given a set of moment conditions $\hat{g}(\beta) = \frac{1}{n}\sum_{i=1}^{n} g(z_i, \beta)$ with $E(g(z_i, \beta_0)) = 0$, let $\hat{\beta}$ be chosen such that $\hat{A}\hat{g}(\hat{\beta}) = 0$ for some matrix $\hat{A}$ satisfying $\hat{A} = A + \frac{1}{n}\sum_{i=1}^{n}\psi(z_i) + O_p(n^{-1})$, $E(\psi(z_i)) = 0$. Then $\hat{\beta}$ is the $\hat{A}$-GMM estimator of β.

Note that the definition does not exclude $\hat{A}$ from depending on β. Indeed, for the two-step GMM estimator discussed above, this is exactly the case. We will be interested in the case where $\hat{A}$ depends on a parameter θ of fixed dimension, as in the following example:

Example 1 Let $\hat{A} = A(\hat{\theta})$, where θ has a fixed dimension. Suppose

$$\hat{\theta} = \theta_0 + \frac{1}{n}\sum_{i=1}^{n}\delta_i + O_p(n^{-1}), \qquad E(\delta_i) = 0$$

Then $\hat{A} = A(\hat{\theta}) = A(\theta_0) + \frac{\partial A(\theta_0)}{\partial \theta}(\theta - \theta_0) + O_p(n^{-1})$. This in turn implies $\psi(z_i) = \frac{\partial A(\theta_0)}{\partial \theta}\delta_i$.

4 Bias of General Moment Weighting GMM

We compute the bias of the GMM estimator with general moment wighting matrix $\hat{A}$ taking a Taylor Expansion of the moment conditions $\hat{A}\hat{g}(\hat{\beta}) = 0$. We require a few preliminary assumptions.

Assumption 1 **(a)** *$\beta_0 \in \mathcal{B}$ is the unique solution to $E[g(z, \beta)] = 0$;* **(b)** *$\mathcal{B}$ is compact;* **(c)** *$g(z, \beta)$ is continuous at each $\beta \in \mathcal{B}$ with probability one.*

Assumption 2 *There exists an A and $\psi(z)$ such that $\hat{A} = A + \frac{1}{n}\sum_{i=1}^{n}\psi(z_i) + O_p(n^{-1})$ with $E(\psi(z_i)) = 0$.*

Let $\hat{\beta}_{GMM}$ be the vector β that sets $\hat{A}\hat{g}(\beta) = 0$. The regularity and identification assumptions in Assumption 1 guarantee that $\hat{\beta}_{GMM}$ is consistent.

Theorem 1 *Under Assumptions 1 and 2, let $\hat{\beta}_{GMM}$ solve $\hat{A}\hat{g}(Z,\beta) = 0$ and let*

- $\hat{\beta}_j = \beta_{j0} + \hat{\Lambda}_j + O_p(n^{-1})$, *where* $\hat{\Lambda}_j = \frac{1}{n}\sum_{k=1}^{n}\lambda_{kj}$
- $\hat{\Psi}(Z) = \frac{1}{n}\sum_{i=1}^{n}\psi(z_i)$
- $\hat{g}(\beta) = \frac{1}{n}\sum_{i=1}^{n}g(z_i,\beta)$

Then our GMM estimator has the following bias:

$$E(\hat{\beta}_{GMM}-\beta) = -\Sigma\left[E(\hat{\Psi}(Z)\hat{g}(\beta_0)) - E(\hat{\Psi}(Z)G(\beta_0)Q) - AE(\hat{G}(\beta_0)Q) - \frac{1}{2}\sum_{i}^{k}H_i)\right],$$

$$\Sigma = [AG(\beta_0)]^{-1}, \qquad H_i = E(\hat{\Lambda}_i A G_i'(\beta_0)Q) \qquad Q = \Sigma A\hat{g}(\beta_0)$$

Proof: See Appendix C.1.

We are able to see how the formula in Theorem 1 relates to the bias results of Newey and Smith (2003) given in Section 2. First recall that their results are based on the two-step GMM estimator. For their calculations, we have $\hat{A} = \hat{G}'\hat{\Omega}(\tilde{\beta})^{-1}$ and $A = G'\Omega^{-1}$, where $\hat{G} = [\sum_{i=1}^{n}G_i(\hat{\beta}_{GMM})/n]$. We can now view $\hat{\Psi}(Z)$ as the influence function for $\hat{G}\hat{\Omega}^{-1}$. Plugging in $A = G'\Omega^{-1}$ shows that our third term is just B_I. B_G comes from both the first and second terms.

The portion of the bias given by $-[AG(\beta_0)]^{-1}E(\hat{\Psi}(Z)\hat{g}(\beta_0))$ in Theorem 1 is represented by $-[E(G_i(\alpha))'\Omega^{-1}E(G_i(\alpha))]^{-1}E(G_i'\Omega^{-1}g_i)/n$ in Newey and Smith. This is precisely the term which we applied to the AR(1) setting in Part 1 of this thesis, and for which we obtained a closed-form solution as a function of the underlying parameters of the model. We now look at its (more general) counterpart given by $-[AG(\beta_0)]^{-1}E(\hat{\Psi}(Z)\hat{g}(\beta_0))$.

Consider the $E(\hat{\Psi}(Z)\hat{g}(\beta_0))$ term. This depends on $\hat{\Psi}(Z)$, the influence function of $\hat{A}$. In Example 1, we saw that $\hat{\Psi}(Z) = \frac{\partial A(\theta_0)}{\partial \theta}\hat{\Delta}$ if $\hat{A}$ depends on a parameter of fixed dimension with influence function $\hat{\Delta}$. In the next example, we will continue with this assumption on $\hat{A}$ and show a means of simplifying the bias calculation.

Example 2 Let $\hat{A} = A(\hat{\theta})$, where θ has a fixed dimension. If θ is $JX1$, then (from Ex. 1), $\psi(z_i) = \sum_{j=1}^{J} \frac{\partial A(\theta_0)}{\partial \theta_j} \delta_{ij}$. Suppose our data Z is i.i.d.

$$\begin{aligned}
E(\hat{\Psi}(Z)\hat{g}(\beta_0)) &= E\left(\tfrac{1}{n}\sum_{i=1}^{n} \psi(z_i) * \tfrac{1}{n}\sum_{i=1}^{n} g(z_i, \beta_0)\right) \\
&= (1/n)^2 \sum_{i=1}^{n} E(\psi(z_i) g(z_i, \beta_0)) + (1/n)^2 \sum_{i \neq k} \underbrace{E(\psi(z_i) g(z_k, \beta_0))}_{0} \\
&= (1/n)^2 \sum_{i=1}^{n} E\left(\sum_{j=1}^{J} \frac{\partial A(\theta_0)}{\partial \theta_j} \delta_{ij} g(z_i, \beta_0)\right) \\
&= (1/n)^2 \sum_{i=1}^{n} \sum_{j=1}^{J} \frac{\partial A(\theta_0)}{\partial \theta_j} E(\delta_{ij} g(z_i, \beta_0)) \\
&= (1/n)^2 \sum_{i=1}^{n} \sum_{j=1}^{J} E\left(\delta_{ij} \frac{\partial A(\theta_0)}{\partial \theta_j} g(z_i, \beta_0)\right)
\end{aligned}$$

Studying the behavior of $\frac{\partial A(\theta_0)}{\partial \theta_j} g_i$ will give us an understanding of how the bias grows with T. We will carry out this exercise by looking at the Arellano IVE Estimator.

5 Arellano's IVE Estimator

One estimator which fits in well with our framework is Arellano's projection-restricted IVE Estimator (Arellano (2003)) for dynamic panel data models. In Section 5.1, we show that the estimator has a $\hat{A}$ that is a function of a parameter of fixed dimension. We also look at form the instrument takes in the AR(1) model. In Section 5.2, we apply the results of Example 2 to find the rate at which the Arellano IVE estimator grows with T in the AR(1) setting.

5.1 The Estimator

In dynamic panel data models, the GMM estimator takes on the following form:

$$\hat{\beta} = \left(\sum_i \sum_t \hat{h}_{it} x_{it}^{*\prime} \right)^{-1} \sum_i \sum_t \hat{h}_{it} y_{it}^{*}$$

GMM sets $\hat{h}_{it}{}' = z_i^{t\prime} \hat{\Pi}_t$ where $\hat{\Pi}_t$ is an OLS estimate. Arellano's IVE estimator sets $\hat{h}_{it}{}' = z_i^{t\prime} \Pi_t(\hat{\gamma})$ where $\hat{\gamma}$ is a pseudo-maximum likelihood (PML) estimate.

The drawback of GMM here is especially acute when T and N both tend to infinity and the right-hand side variables are endogenous. In that case, the GMM bias is of order T/N. When T is fixed and N is large, both GMM and Arellano's IVE estimator are consistent under the same assumptions. However, when T is not fixed, Arellano's IVE estimator is immune to asymptotic biases because the number of first-stage coefficients does not increase with T.

Arellano's IVE GMM estimator requires the construction of an auxiliary VAR model for the instruments. It is this auxiliary model which leads to the PML estimates of the components of $\hat{\gamma}$. We will consider Arellano's IVE estimator is in the framework of our GMM estimator with general weighting matrix $\hat{A}$. We will look specifically at the special case of an AR(1) model with a strictly stationary auxiliary model.

Example 3 Consider the dynamic panel data model:

$$y_{it} = x'_{i,t-1}\beta + \eta_i + \epsilon_{it} \quad t = 1, \dots, T; \;\; i = 1, \dots, N$$

Then Arellano's IVE estimator is a GMM estimator with weighting matrix A and moments $g(Z, \beta)$ defined below:

$$\begin{array}{rclrcl} \hat{A} & = & (\Pi_1(\hat{\gamma})' \cdots \Pi_T(\hat{\gamma})'), & \hat{g}(\beta) & = & \frac{1}{n}\sum_{i=1}^{n} g(z_i, \beta), \\ \Pi_t(\hat{\gamma}) & = & (\pi_t^0(\hat{\gamma})' \cdots \pi_t^t(\hat{\gamma}))' & g(z_i, \beta) & = & (e_{i1} z_i^{1\prime} \cdots e_{iT} z_i^{T\prime})' \\ e_{it} & = & (y_{it}^* - x_{it}^{*\prime}\beta) & z_i^t & = & (z_{i1}' \cdots z_{it}')' \end{array}$$

$$y_{it}^* = \left(\frac{T-t}{T-t+1}\right)^{1/2} \left[y_{it} - \frac{1}{T-t}(y_{i,t+1} + \cdots + y_{iT}) \right]$$

In Example 3, we do not specify the exact functional form of the π_t^s coefficients or the dimension of γ. Both of these are determined by the assumptions we make in the auxiliary VAR model. General formulas are found in Appendix B of Arellano (2003). We will look at the specific case of an AR(1) model with individual effects, and assume that the auxiliary VAR model is strictly stationary. The model is given by:

$$y_{it} = \alpha y_{i,t-1} + (1-\alpha)\mu_i + v_{it} \tag{3}$$

$$E(v_{it} | y_{i0}, y_{i1}, \dots, y_{i,t-1}) = 0$$

Let y_{it}^* be defined as in Example 3 and let $x_{it} = y_{i,t-1}$. Then we have:

$$y_{it}^* = \alpha x_{i,t}^* + v_{it}^*$$

The instrument is given by $h_{it} = E(x_{it}^* | y_i^{t-1})$.

Lemma 1 *Consider the AR(1) Model in Eq.(3). Assume that the auxiliary VAR model is strictly stationary. Then the Arellano IVE estimator's instrument takes the following form:*

$$
\begin{aligned}
h_t(y_i^{t-1},\gamma) &= c_t\left[1-\tfrac{\alpha}{1-\alpha}\left(\tfrac{1-\alpha^{T-t}}{T-t}\right)\right]\left[y_{i,t-1}-m_{t-1}(y_i^{t-1},c)\right] & t\geq 1\\
m_0(y_i^0,\gamma) &= \left[\mu+\phi(1-\alpha^2)y_{i0}\right]/\left[1+\phi(1-\alpha^2)\right] & t=1\\
m_{t-1}(y_i^{t-1},\gamma) &= \tfrac{\mu+\phi\left[(1-\alpha)\sum_{s=1}^{t-1}u_{is}+(1-\alpha^2)y_{i0}\right]}{1+\phi[(t-1)(1-\alpha)^2+1-\alpha^2]} & t\geq 2
\end{aligned}
$$

$$
\begin{aligned}
&\phi=\sigma_\mu^2/\sigma^2,\ \ \mu=E(\mu_i),\ \ u_{is}=y_{is}-\alpha y_{i,s-1},\\
&\gamma=(\alpha,\,\phi,\,\mu),\ \ c_t=(\tfrac{T-t}{T-t+1})^{1/2}.
\end{aligned}
$$

Proof: See Arellano (2003), Apppendix B.

The assumptions of a strictly stationarity auxiliary model need not be true. The coefficients should therefore be understood as pseudo true values for which we use the notation $c=(a,f,m)$. Our instrument is then $h_t(y_i^{t-1},c)$. As can be seen in Lemma 1, $E(x_{it}^*|y_i^{t-1})$ is a linear combination of y_i^{t-1}.

5.2 Bias of Arellano IVE Estimator

In Example 2, we showed that when $\hat{A}$ depends on a parameter vector of fixed dimension,

$$E(\hat{\Psi}(Z)\hat{g}(\beta_0))=(1/n)^2\sum_{i=1}^{n}\sum_{j=1}^{J}E(\delta_{ij}(\partial A(\theta_0)/\partial\theta_j)g(z_i,\beta_0))$$

For the Arellano IVE estimator, the order of $E(\hat{\Psi}(Z)\hat{g}(\beta_0))$ is determined by the order at which $\sum_{j=1}^{J}E(\delta_{ij}(\partial A(\theta_0)/\partial\theta_j)g(z_i,\beta_0))$ grows with T.

Theorem 2 *Consider Arellano IVE estimator for the AR(1) model. Let* $\theta = (a, f, m)$ *where* (a, f, m) *are the pseudo-parameters defined in Section 5.1, and let* $\hat{\Delta}_j = \sum_{i=1}^{n} \delta_{ij}$ *be the influence function of* θ_j*. Then we have the following result:*

$$E(\hat{\Psi}(Z)\hat{g}(\beta_0)) \leq O_p(\sqrt{T})$$

Proof: See Appendix C.2.

We can compare this to our bias results in Part 1. There we showed that the GMM estimator for the AR(1) model has:

$$E\left(G_i'\Omega^{-1}g_i\right) = \sum_{t=1}^{T-2} B(\alpha, \sigma, T, t) * Term(t)$$

where $Term(t) = O_p(t)$ and $B(\alpha, \sigma, T, t) \simeq [1 + \alpha^{T-t}(T-t)]/(T-t)^2$. Combining the two terms and summing over t yields:

$$E\left(G_i'\Omega^{-1}g_i\right) = O_p(T)$$

Arellano's estimator is immune to asymptotic biases because it does not have first-stage coefficients increasing with T. By studying the influence function of Arellano's IVE estimator, we have shown $E\left(G_i'\Omega^{-1}g_i\right)$ is growing at a smaller rate than its GMM counterpart which simply sets the instrument $\hat{h}_{it}' = z_i^{t\prime}\hat{\Pi}_t$ where $\hat{\Pi}_t$ is an OLS estimate.

6 Conclusion

Bias calculations for standard GMM models, such as Hansen's (1982) two-step estimator, have already been covered in the literature. We have looked at the bias of a general GMM model where the weighting matrix $\hat{A}$ of the moment conditions $g(z, \beta)$ is left unspecified, except for some general conditions. Through a Taylor expansion, we have computed the bias of this GMM estimator and have compared it to the results of Newey and West (2003). An important case of GMM estimation with a general weighting matrix $\hat{A}$ is when $\hat{A}$ is a function of a vector of parameters of fixed dimension, $\hat{A} = A(\hat{\theta})$, in which case the order of the bias depends on the influence function for $\hat{\theta}$. One example of such an estimator is Arellano's IVE estimator. We considered this estimator's bias properties and found that it does better than the standard GMM estimator as T grows.

Technical Appendix

A Chapter 1 Appendix

A.1 Lemma 1

Consider our AR(1) model under assumptions 1-3. The finite sample GMM bias which grows with the number of moment restrictions is:

$$
\begin{aligned}
(\hat{\alpha}_{GMM} - \alpha) &= -(E(G_i(\alpha))'\Omega^{-1}E(G_i(\alpha)))^{-1}E(G_i'\Omega^{-1}g_i)/n, \quad \textit{where} \\
\Omega^{-1} &= E\left(g_i(\alpha)\, g_i(\alpha)'\right)^{-1} \\
G_i(\alpha) &= \tfrac{d}{d\alpha} g_i(\alpha)
\end{aligned}
$$

Proof: We make use of the bias calculation of Newey and Smith (2003):

$$
\begin{aligned}
Bias(\hat{\beta}_{GMM}) &= B_I + B_G + B_\Omega + B_W \\
B_I &= H(-a + E[G_i H g_i])/n \\
B_G &= -\Sigma E[G_i{}' P g_i]/n \\
B_\Omega &= HE[g_i g_i{}' P g_i]/n \\
B_W &= H\textstyle\sum_{j=1}^{K} \tilde{\Omega}_{\beta_j}(H_W - H)' e_j/n
\end{aligned}
$$

where $H = \Sigma G'\Omega^{-1}$, $H_W = (G'W^{-1}G)^{-1}G'W^{-1}$, $G = E(G_i)$, $P = \Omega^{-1} - \Omega^{-1}G\Sigma G'\Omega^{-1}$, $\tilde{\Omega}_{\beta_j} = E\left[\partial\{g_i(\beta_0)g_i(\beta_0)'\}/\partial\beta_j\right]$, $\Sigma = (G'\Omega^{-1}G)^{-1}$.

B_I is the asymptotic bias for GMM with the optimal linear combination $G'\Omega^{-1}g(z,\beta)$ - it will not grow with the number of moment restrictions. For our AR(1) model with Assumptions 1-3, we have symmetry, third moments equal to 0, and a weight matrix W that is a scalar multiple of Ω. From Newey and Smith (2003), this implies $B_\Omega = B_W = 0$. This leaves us with:

$$
B_G = -\Sigma E\left[G_i{}'\Omega^{-1}g_i\right]/n - \Sigma E\left[G_i{}'\Omega^{-1}G\Sigma G'\Omega^{-1}g_i\right]/n
$$

The second term on the RHS also fails to grow with the number of moment restrictions, yielding the desired result.

A.2 Lemma 2

We have the following two results:

- $f_i(a,b) \equiv f_i(a)\, f_i(b)$

- $f_i(x) \equiv v^*_{i,j+1} y_{i,x-[j(j-1)/2]}$, where $j \equiv j(x)$ where $j(x) \in I \;\; s.t. \;\; \frac{-1+\sqrt{1+8x}}{2} \leq j(x) < \frac{1+\sqrt{1+8x}}{2}$

Proof: We require a means of choosing the appropriate j for a given x such that $f_i(x) \equiv v^*_{i,j+1} y_{i,x-[j(j-1)/2]}$. Define the mapping Ψ from x to j by $\Psi : \mathbf{N} \rightarrow \mathbf{N}$:

$$\begin{array}{lllllllll} Psi(1) = 1 & Psi(2) = 2 & Psi(4) = 3 & Psi(7) = 4 & \cdots \\ & Psi(3) = 2 & Psi(5) = 3 & Psi(8) = 4 & \cdots \\ & & Psi(6) = 3 & Psi(9) = 4 & \cdots \\ & & & Psi(10) = 4 & \cdots \end{array}$$

For each j, define $X(j) = \{x : Psi(x) = j\}$. We have $max(X(j)) = j(j+1)/2$ and $min(X(j)) == j(j-1)/2 - 1$. For each $x \in X(j)$, $\frac{j(j-1)}{2} < x \leq \frac{j(j+1)}{2}$, we wish to solve for j in terms of x. Completing the squares for both expressions involving j gives: $\frac{(2j-1)^2-1}{8} < x \leq \frac{(2j+1)^2-1}{8}$, We can now solve algebraically for j: $\frac{-1+\sqrt{1+8x}}{2} \leq j(x) < \frac{1+\sqrt{1+8x}}{2}$

A.3 Theorem 1

- If $a = b$, then $E[f_i(a,a)] = \sigma^2(p+q)$
- If $a \neq b$ but a and b are in the same family, then $E[f_i(a,b)] = \sigma^2\left(\alpha^{|b-a|}p + q\right)$
- If $a \neq b$ and a and b are in different families, then $E[f_i(a,b)] = 0$

Proof: We make use of Lemma 2. We consider three cases in the next three sections: a, b in the same family with $a = b$ (case 1), a, b in the same family with $a \neq b$ (case 2), and a, b in different families (case 3).

A.3.1 Case 1: a=b

Let $a = b$. Define $d \equiv a - [j(j-1)/2]$.

$$E[f_i(a,a)] = E\left[v^{*2}_{i,j+1} y^2_{i,d}\right] = \sigma^2 E\left[y^2_{i,d}\right]$$

The first equality follows from $j+1 > d$ and independence of the $v^*_{i,j+1}$ terms. The second equality follows from

$$
\begin{aligned}
E\left(v_{i,j+1}^{*2}\right) &= c_{j+1}^{2}\left(E\left(v_{i,j+1}^{2}\right)+\frac{1}{(T-j-1)^{2}}E\left(\left(\sum_{k=j+2}^{T}v_{i,k}\right)^{2}\right)-\frac{2}{T-t}E\left(v_{i,j+1}\sum_{k=j+2}^{T}v_{i,k}\right)\right)\\
&= \frac{T-j-1}{T-j}\left[\sigma^{2}+\frac{1}{(T-j-1)^{2}}\left(T-j-1\right)\sigma^{2}-0\right]\\
&= \frac{T-j-1}{T-j}\left[\frac{T-j}{T-j-1}\sigma^{2}\right] = \sigma^{2}
\end{aligned} \tag{4}
$$

Expanding $y_{i,t}$, we can write:

$$
\begin{aligned}
y_{i,t} &= \alpha y_{i,t-1}+\eta_i+v_{i,t}\\
&= \quad \vdots\\
&= \left(1+\alpha+\alpha^{2}+\cdots\right)+\left(v_{i,t}+\alpha v_{i,t}+\cdots\right)\\
y_{i,t} &= \tfrac{\eta_i}{1-\alpha}+\textstyle\sum_{k=0}^{\infty}v_{i,t-k}\\
y_{i,d}^{2} &= \left(\tfrac{\eta_i}{1-\alpha}\right)^{2}+\textstyle\sum_{k=0}^{\infty}\alpha^{2k}v_{i,d-k}^{2}+cross-terms\\
E\left[y_{i,d}^{2}\right] &= \tfrac{\sigma_\eta^{2}}{(1-\alpha)^{2}}+\tfrac{\sigma^{2}}{1-\alpha^{2}}\\
&= q+p\\
E\left[f_i(a,a)\right] &= \sigma^{2}(p+q)
\end{aligned}
$$

A.3.2 Case 2: $a \neq b, j(a)=j(b)$

Let $a \neq b$. WLOG, take $a < b$. We have assumed that a, b are still in the same family. Define $d(a) \equiv a-\left[j(a)\left(j(a)-1\right)/2\right]$, $d(b) \equiv b-\left[j(b)\left(j(b)-1\right)/2\right]$, $d \equiv d(b)-d(a)=b-a$. Then we have:

$$
\begin{aligned}
E\left[f_i(a,b)\right] &= E\left[v_{i,j+1}^{*2}y_{i,d(a)}y_{i,d(b)}\right]\\
&= \sigma^{2}E\left[y_{i,d(a)}y_{i,d_2}\right]\\
&= \sigma^{2}E\left[y_{i,d(a)}\left(\alpha^{b-a}y_{i,d_1}+\left(1+\alpha+\cdots+\alpha^{(b-a)-1}\right)\eta_i+v_i\,terms\right)\right]\\
&= \sigma^{2}\left[\alpha^{b-a}E\left(y_{i,d(a)}^{2}\right)+\left(\frac{1-\alpha^{|b-a|}}{1-\alpha}\right)E\left(y_{i,d(a)}\eta_i\right)+E\left(y_{i,d(a)}\cdot v_i\,terms\right)\right]
\end{aligned}
$$

Solving out $E\left(y_{i,d(a)}\eta_i\right)$, we have:

$$
\begin{aligned}
E\left(y_{i,d(a)}\eta_i\right) &= E\left[\eta_i\left(\alpha y_{i,d(a)-1}+\eta_i+v_{i,d(a)}\right)\right] \\
&= E\left(\eta_i^2\right)+\alpha E\left(y_{i,d(a)-1}\eta_i\right)+0 \\
&= \vdots \\
&= \left(1+\alpha+\cdots\right)E\left(\eta_i^2\right) \\
&= \frac{\sigma_\eta^2}{1-\alpha}
\end{aligned}
$$

Plugging in for $E\left(y_{i,d_1}\eta_i\right)$, $E\left(y_{i,d_1}^2\right)$, and setting $E\left(y_{i,d_1}\cdot v_i\, terms\right)=0$, we have:

$$
\begin{aligned}
E\left[f_i(a,b)\right] &= \sigma^2\left[\alpha^{|b-a|}\left(\frac{\sigma_\eta^2}{(1-\alpha)^2}+\frac{\sigma^2}{1-\alpha^2}\right)+\left(\frac{1-\alpha^{|b-a|}}{1-\alpha}\right)\left(\frac{\sigma_\eta^2}{1-\alpha}\right)\right] \\
&= \sigma^2\left(\alpha^{|b-a|}p+q\right)
\end{aligned}
$$

A.3.3 Case 3: $a\neq b, j(a)\neq j(b)$

Let $a\neq b$. WLOG, take $a<b$. We now assume that a,b are in different families. Define $d(a)\equiv a-\left[j(a)\left(j(a)-1\right)/2\right]$, $d(b)\equiv b-\left[j(b)\left(j(b)-1\right)/2\right]$, $d\equiv d(b)-d(a)$. Then we have:

$$
\begin{aligned}
E\left[f_i(a,b)\right] &= E\left[v_{i,j(a)+1}^*v_{i,j(b)+1}^*y_{i,d(a)}y_{i,d(b)}\right] \\
&= E\left(v_{i,j(b)+1}^*\right)E\left[v_{i,j(a)+1}^*y_{i,d(a)}y_{i,d(b)}\right] \\
&= 0
\end{aligned}
$$

This follows from $j(b)>j(a)$ and $j(b)\geq d(b)$. This second inequality is easy to see:

$$
\begin{aligned}
j(b) &\geq d(b) \iff \\
j(b) &\geq b-\left[j(b)\left(j(b)-1\right)/2\right] \iff \\
j(b)^2+j(b)-2x &\geq 0 \iff \\
j(b) &\geq \frac{-1+\sqrt{1+8x}}{2}
\end{aligned}
$$

This last inequality holds by definition of $j(b)$.

A.4 Theorem 2

Applying the partitioned matrix inverse formula gives us the following representation of Ω_i^{-1}:

$$\Omega_i^{-1} = \begin{pmatrix} A_1^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_{T-2}^{-1} \end{pmatrix}, \quad A_t^{-1} = \frac{1}{\sigma^4}\left(N_t - h_t R_t\right) \forall\, t$$

$$h_t = \left(\frac{\sigma_\eta^2}{\sigma^2 + \sigma_\eta^2\left(t + \frac{2\alpha}{1-\alpha}\right)}\right) \forall t$$

$$N_1 = (1-\alpha^2), \quad N_2 = \begin{pmatrix} 1 & -\alpha \\ -\alpha & 1 \end{pmatrix}, \quad R_1 = (1+\alpha)^2, \quad R_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \; For\ t \geq 3,$$

$$N_t = \begin{pmatrix} 1 & -\alpha & 0 & \cdots & 0 \\ -\alpha & 1+\alpha^2 & -\alpha & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\alpha & 1+\alpha^2 & -\alpha \\ 0 & \cdots & 0 & -\alpha & 1 \end{pmatrix}, \; R_t = \begin{pmatrix} 1 & 1-\alpha & \cdots & 1-\alpha & 1 \\ 1-\alpha & (1-\alpha)^2 & \cdots & (1-\alpha)^2 & 1-\alpha \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-\alpha & (1-\alpha)^2 & \cdots & (1-\alpha)^2 & 1-\alpha \\ 1 & 1-\alpha & \cdots & 1-\alpha & 1 \end{pmatrix}$$

Proof: We can rewrite A_t as $A_t = \sigma^2 \left(pM_t + qe_te_t'\right)$, where e_t is a $t \times 1$ column of 1's and M is defined by:

$$M_1 = 1 \; and \; M_t = \begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{i-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{i-2} \\ \alpha^2 & \alpha & 1 & \cdots & \alpha^{i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{i-1} & \alpha^{i-2} & \alpha^{i-3} & \cdots & 1 \end{pmatrix}, \; t \geq 2$$

Inverting M_t gives $M_1^{-1} = 1$, $M_2^{-1} = \frac{1}{1-\alpha^2}\begin{pmatrix} 1 & -\alpha \\ -\alpha & 1 \end{pmatrix}$, and for $t \geq 3$,

$$M_t^{-1} = \frac{1}{1-\alpha^2}\begin{pmatrix} 1 & -\alpha & 0 & \cdots & 0 \\ -\alpha & 1+\alpha^2 & -\alpha & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\alpha & 1+\alpha^2 & -\alpha \\ 0 & \cdots & 0 & -\alpha & 1 \end{pmatrix}$$

To invert A_t, we apply the partitioned matrix inverse formula (Linear Statistical Inference by Rao). The formula states that if B is a nonsingular matrix and U, V are column vectors, then

$$(B + UV')^{-1} = B^{-1} - \frac{(B^{-1}U)(V'B^{-1})}{1 + V'B^{-1}U}$$

For our problem, $A_t^{-1} = \frac{1}{\sigma^2}(pM_t + qe_te_t')^{-1}$. Applying the formula with $B = pM_t$, $U = qe_t$, and $V = e_t$ gives:

$$\begin{aligned}
A_t^{-1} &= \frac{1}{\sigma^2}\left(M_t^{-1}/p - \frac{((M_t^{-1}/p)qe_t)(e_t'M_t^{-1}/p)}{1 + e_t'(M_t^{-1}/p)qe_t}\right) \\
&= \frac{1}{\sigma^2}\left(M_t^{-1}/p - \frac{(q/p^2)(M_t^{-1}e_t)(e_t'M_t^{-1})}{(p + qe_t'M_t^{-1}e_t)/p}\right) \\
A_t^{-1} &= \frac{1-\alpha^2}{\sigma^4}\left(M_t^{-1} - \frac{q(M_t^{-1}e_t)(e_t'M_t^{-1})}{p + qe_t'M_t^{-1}e_t}\right)
\end{aligned}$$

To simplify this solution, let $N_t = (1-\alpha^2)M_t^{-1}$. This gives:

$$\begin{aligned}
A_t^{-1} &= \frac{1-\alpha^2}{\sigma^4}\left(N_t/(1-\alpha^2) - \frac{q\left(1/(1-\alpha^2)\right)^2(N_te_t)(e_t'N_t)}{p + q(1/(1-\alpha^2))e_t'N_te_t}\right) \\
A_t^{-1} &= \frac{1}{\sigma^4}\left(N_t - \frac{q(N_te_t)(e_t'N_t)}{p(1-\alpha^2) + qe_t'N_te_t}\right) \\
A_t^{-1} &= \frac{1}{\sigma^4}\left(N_t - \left(\frac{q}{p(1-\alpha^2) + qe_t'N_te_t}\right)(N_te_t)(e_t'N_t)\right)
\end{aligned}$$

Expanding $(N_te_t)(e_t'N_t)$, we have:

$$(N_te_t)(e_t'N_t) = (1-\alpha)^2 \underbrace{\begin{pmatrix} 1 & 1-\alpha & \cdots & 1-\alpha & 1 \\ 1-\alpha & (1-\alpha)^2 & \cdots & (1-\alpha)^2 & 1-\alpha \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-\alpha & (1-\alpha)^2 & \cdots & (1-\alpha)^2 & 1-\alpha \\ 1 & 1-\alpha & \cdots & 1-\alpha & 1 \end{pmatrix}}_{R_t} \quad t \geq 3$$

For $t = 1, 2$, we have $R_1 = (1+\alpha)^2$ and $R_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. We can now write:

$$A_t^{-1} = \frac{1}{\sigma^4}\left(N_t - (1-\alpha)^2\left(\frac{q}{p(1-\alpha^2)+qe_t'N_te_t}\right)R_t\right)$$

Solving for $e_t'N_te_t$ gives

$$\begin{array}{rcl} e_t'N_te_t & = & 2(1-\alpha)+(t-2)(1-\alpha)^2 \\ \downarrow & & \downarrow \\ A_t^{-1} & = & \frac{1}{\sigma^4}\left(N_t - \underbrace{\left(\frac{\sigma_\eta^2}{\sigma^2+\sigma_\eta^2\left(t+\frac{2\alpha}{1-\alpha}\right)}\right)}_{h_t} R_t\right) \end{array}$$

A.5 Solving for $E\left(G_i'\Omega^{-1}g_i\right)$: Part 1

$$E\left(G_i'\Omega^{-1}g_i\right) = \sum_{t=1}^{T-2} E\left(G_t'A_t^{-1}g_t\right)$$

$$E\left(G_t'A_t^{-1}g_t\right) = \begin{cases} \frac{c_2}{\sigma^4}\left(\frac{p+q}{c_2}\right)L_1\left[h_1(1+\alpha)^2-(1-\alpha)^2\right] & t=1 \\ \frac{c_3}{\sigma^4}\left\{B_{12}E(D_{12})+(2\alpha+B_{22})E(D_{22})\right\} & t=2 \\ \frac{c_{t+1}}{\sigma^4}\left\{B_{1t}E(D_{1t})+B_{2t}E(D_{2t})+B_{3t}E(D_{3t})+B_{4t}E(D_{4t})+B_{5t}E(D_{5t})\right\} & t\geq 3 \end{cases}$$

$$B_{1t}=h_t-1 \quad B_{2t}=2h_t \quad B_{3t}=h_t(1-\alpha)^2-(1+\alpha^2) \quad B_{4t}=2\alpha \quad B_{5t}=h_t(1-\alpha)^2$$

$$\begin{array}{rcl} D_{1t} & = & v_{t+1}^*\left(y_t-\frac{1}{T-t-1}\sum_{k=t+1}^{T-1}y_k\right)\left(y_1^2+y_t^2\right) \\ D_{2t} & = & v_{t+1}^*\left(y_t-\frac{1}{T-t-1}\sum_{k=t+1}^{T-1}y_k\right)\left(y_1y_t\right) \\ D_{3t} & = & v_{t+1}^*\left(y_t-\frac{1}{T-t-1}\sum_{k=t+1}^{T-1}y_k\right)\sum_{k=2}^{t-1}y_k^2 \\ D_{4t} & = & v_{t+1}^*\left(y_t-\frac{1}{T-t-1}\sum_{k=t+1}^{T-1}y_k\right)\sum_{k=1}^{t-1}y_ky_{k+1} \\ D_{5t} & = & v_{t+1}^*\left(y_t-\frac{1}{T-t-1}\sum_{k=t+1}^{T-1}y_k\right)\sum_{j\neq k}\sum_{k=2}^{t-1}y_jy_k \\ L_t & = & -\frac{\sigma^2}{(1-\alpha)(T-t)}\left[\left(1-\alpha^{T-t-1}\right)\left(1+\frac{1}{(T-t-1)(1-\alpha)}\right)-1\right] \end{array}$$

Proof: Our goal is to find an expression for $E\left(G_t'A_t^{-1}g_t\right)$ for $t=1,\ldots,T-2$. We will solve this in parts. Writing out $G_t'A_t^{-1}g_t$, we have:

$$G_t'A_t^{-1}g_t = -v_{t+1}^*c_{t+1}\left[y_t-\frac{1}{T-t-1}\sum_{k=t+1}^{T-1}y_k\right]\tilde{y}_t'A_t^{-1}\tilde{y}_t$$

Working out $\tilde{y}_t' A_t^{-1} \tilde{y}_t$ gives:

$$\tilde{y}_t' A_t^{-1} \tilde{y}_t \quad = \quad \frac{1}{\sigma^4}\left[\tilde{y}_t' N_t \tilde{y}_t - h_t \tilde{y}_t' R_t \tilde{y}_t\right]$$

Working out the two parts $\tilde{y}_t' N_t \tilde{y}_t$ and $\tilde{y}_t' R_t \tilde{y}_t$ gives:

$$\tilde{y}_t' N_t \tilde{y}_t = \begin{cases} y_1^2(1-\alpha)^2 & t=1 \\ y_1^2 + y_2^2 - 2\alpha y_1 y_2 & t=2 \\ y_1^2 + y_t^2 + (1+\alpha^2)\sum_{k=2}^{t-1} y_k^2 - 2\alpha \sum_{k=1}^{t-1} y_k y_{k+1} & t \geq 3 \end{cases}$$

$$\tilde{y}_t' R_t \tilde{y}_t = \begin{cases} y_1^2(1+\alpha)^2 & t=1 \\ y_1^2 + y_2^2 + 2 y_1 y_2 & t=2 \\ y_1^2 + y_t^2 + 2 y_1 y_t + 2(1-\alpha)(y_1 + y_t)\sum_{k=2}^{t-1} y_k + (1-\alpha)^2 (\sum_{k=2}^{t-1} y_k)^2 & t \geq 3 \end{cases}$$

$$Letting \quad (\sum_{k=2}^{t-1} y_k)^2 = \sum_{k=2}^{t-1} y_k^2 + \sum_{j \neq k} \sum_{k=2}^{t-1} y_j y_k, \; we\ have:$$

$$\tilde{y}_t'[N_t - h_t R_t]\tilde{y}_t = \begin{cases} y_1^2\left[(1-\alpha)^2 - h_1(1+\alpha)^2\right] & t=1 \\ (y_1^2 + y_2^2)(1-h_2) - 2 y_1 y_2 (\alpha + h_2) & t=2 \\ (y_1^2 + y_t^2)(1-h_t) - 2 h_t y_1 y_t - 2\alpha \sum_{k=1}^{t-1} y_k y_{k+1} & \\ \quad - h_t (1-\alpha)^2 \sum_{j \neq k} \sum_{k=2}^{t-1} y_j y_k & \\ \qquad + \left[(1+\alpha^2) - h_t (1-\alpha)^2\right] \sum_{k=2}^{t-1} y_k^2 & t \geq 3 \end{cases}$$

Plugging in gives us the following expression for $E\left(G_t' A_t^{-1} g_t\right)$:

$$E\left(G_t' A_t^{-1} g_t\right) = \begin{cases} \frac{c_2}{\sigma^4} E\left(v_2^* \left[y_1 - \frac{1}{T-2}\sum_{k=2}^{T-1} y_k\right] y_1^2\right)\left[h_1(1+\alpha)^2 - (1-\alpha)^2\right] & t=1 \\ \frac{c_3}{\sigma^4}\left\{B_{12} E(D_{12}) + (2\alpha + B_{22}) E(D_{22})\right\} & t=2 \\ \frac{c_{t+1}}{\sigma^4}\{B_{1t} E(D_{1t}) + B_{2t} E(D_{2t}) + B_{3t} E(D_{3t}) + B_{4t} E(D_{4t}) + B_{5t} E(D_{5t})\} & t \geq 3 \end{cases}$$

where B_{it}, D_{it} are:

$$B_{1t} = h_t - 1 \quad B_{2t} = 2h_t \quad B_{3t} = h_t(1-\alpha)^2 - (1+\alpha^2) \quad B_{4t} = 2\alpha \quad B_{5t} = h_t(1-\alpha)^2$$

$$
\begin{aligned}
D_{1t} &= v_{t+1}^* \left(y_t - \tfrac{1}{T-t-1} \textstyle\sum_{k=t+1}^{T-1} y_k \right) (y_1^2 + y_t^2) \\
D_{2t} &= v_{t+1}^* \left(y_t - \tfrac{1}{T-t-1} \textstyle\sum_{k=t+1}^{T-1} y_k \right) (y_1 y_t) \\
D_{3t} &= v_{t+1}^* \left(y_t - \tfrac{1}{T-t-1} \textstyle\sum_{k=t+1}^{T-1} y_k \right) \textstyle\sum_{k=2}^{t-1} y_k^2 \\
D_{4t} &= v_{t+1}^* \left(y_t - \tfrac{1}{T-t-1} \textstyle\sum_{k=t+1}^{T-1} y_k \right) \textstyle\sum_{k=1}^{t-1} y_k y_{k+1} \\
D_{5t} &= v_{t+1}^* \left(y_t - \tfrac{1}{T-t-1} \textstyle\sum_{k=t+1}^{T-1} y_k \right) \textstyle\sum_{j \neq k} \sum_{k=2}^{t-1} y_j y_k
\end{aligned}
$$

We can simplify our calculations by introducing a term L_t.

$$
\begin{aligned}
L_t &= -\tfrac{c_{t+1}^2 \sigma^2}{(1-\alpha)(T-t-1)} \left[\left(1 - \alpha^{T-t-1}\right) \left(1 + \tfrac{1}{(T-t-1)(1-\alpha)}\right) - 1 \right] \\
&= -\tfrac{\sigma^2}{(1-\alpha)(T-t)} \left[\left(1 - \alpha^{T-t-1}\right) \left(1 + \tfrac{1}{(T-t-1)(1-\alpha)}\right) - 1 \right]
\end{aligned}
$$

Other than the $E(D_{it})$ parts, we also require $E\left(v_2^* \left[y_1 - \frac{1}{T-2} \sum_{k=2}^{T-1} y_k\right] y_1^2\right)$

$$
E\left(v_2^* \left[y_1 - \tfrac{1}{T-2} \textstyle\sum_{k=2}^{T-1} y_k\right] y_1^2\right) = \underbrace{E(v_2^* y_1^3)}_{0} - \tfrac{1}{T-2} E(v_2^* y_1^2 \textstyle\sum_{k=2}^{T-1} y_k)
$$

$$
\begin{aligned}
E(v_2^* y_1^2 y_2) &= c_2 \sigma^2 E(y_1^2) \\
E(v_2^* y_1^2 y_3) &= c_2 \sigma^2 E(y_1^2)(\alpha - \tfrac{1}{T-2}) \\
\vdots \quad &\vdots \quad \vdots \\
E(v_2^* y_1^2 y_{T-1}) &= c_2 \sigma^2 E(y_1^2) \left(\alpha^{T-3} - \tfrac{1}{T-2}[1 + \alpha + \cdots + \alpha^{T-4}]\right)
\end{aligned}
$$

$$
\Downarrow \qquad \Downarrow
$$

$$
\begin{aligned}
E\left(v_2^* \left[y_1 - \tfrac{1}{T-2} \textstyle\sum_{k=2}^{T-1} y_k\right] y_1^2\right) &= -\tfrac{c_2 \sigma^2}{(1-\alpha)(T-2)} [\tfrac{\sigma_\eta^2}{(1-\alpha)^2} + \tfrac{\sigma^2}{1-\alpha^2}] \left[(1 - \alpha^{T-2})(1 + \tfrac{1}{(T-2)(1-\alpha)}) - 1\right] \\
&= (\tfrac{p+q}{c_2}) L_1
\end{aligned}
$$

Plugging back in for $t = 1$ gives the desired result.

A.6 Solving for $E\left(G_i' \Omega^{-1} g_i\right)$: Part 2

Solving for the various $E(D_{it})$ parts gives:

- $E(D_{1t}) = \left(\frac{p+q}{c_{t+1}}\right) L_t$

- $E(D_{2t}) = \alpha^{t-1}\left(\frac{p+q}{c_{t+1}}\right) L_t$
- $E(D_{3t}) = (t-2)\left(\frac{p+q}{c_{t+1}}\right) L_t$
- $E(D_{4t}) = (t-1)\left(\frac{\alpha p+q}{c_{t+1}}\right) L_t$
- $E(D_{5t}) = (2/c_{t+1})\left(\frac{\alpha}{1-\alpha}[(t-3) - \frac{\alpha}{1-\alpha}(1-\alpha^{t-1})]p + \frac{(t-3)(t-2)}{2}q\right) L_t$

Proof: We will calculate $E(D_{it})$ for $i = 1, 2, 3, 4$ in the next four sections.

A.6.1 $E(D_{1t})$

$$\begin{aligned} E(D_{1t}) &= E\left[v^*_{t+1}\left(y_t - \frac{1}{T-t-1}\sum_{k=t+1}^{T-1} y_k\right)(y_1^2 + y_t^2)\right] \\ &= E(D^1_{1t}) + E(D^2_{1t}) - (1/(T-t-1))\left[E(D^3_{1t}) + E(D^4_{1t})\right] \end{aligned}$$

where D^j_{1t} is given by:

$$\begin{array}{ll} D^1_{1t} = v^*_{t+1} y_t y_1^2 & D^3_{1t} = v^*_{t+1} y_1^2 \sum_{k=t}^{T-1} y_k \\ D^2_{1t} = v^*_{t+1} y_t^3 & D^4_{1t} = v^*_{t+1} y_t^2 \sum_{k=t}^{T-1} y_t^2 \end{array}$$

By the independence between v^*_{t+1} and y_t, we have:

- $E(D^1_{1t}) = 0$
- $E(D^2_{1t}) = 0$

$E(D^3_{1t})$ will require some calculation:

$$
\begin{aligned}
E(D_{1t}^3) &= E(v_{t+1}^* y_1^2 y_t + v_{t+1}^* y_1^2 y_{t+1} + \cdots + v_{t+1}^* y_1^2 y_{T-1}) \\
&= 0 + E(v_{t+1}^* y_1^2 y_{t+1}) + E(v_{t+1}^* y_1^2 y_{t+2}) + \cdots + E(v_{t+1}^* y_1^2 y_{T-1})
\end{aligned}
$$

$$
\begin{aligned}
E(v_{t+1}^* y_1^2 y_{t+1}) &= c_{t+1} E\left[(v_{t+1} - \tfrac{1}{T-t-1}\textstyle\sum_{j=t+2}^{T} v_k) y_1^2 y_{t+1}\right] \\
&= c_{t+1}\left\{ E(v_{t+1} y_1^2 y_{t+1}) - \tfrac{1}{T-t-1} \underbrace{E(v_{t+2} y_1^2 y_{t+1} + v_{t+3} y_1^2 y_{t+1} + \cdots)}_{0 \text{ by independence across time}} \right\} \\
&= c_{t+1} E\left(y_1^2 \left[\alpha^t y_1 + (1+\alpha+\cdots+\alpha^{t-1})\eta_i + \textstyle\sum_{k=0}^{t-1} \alpha^k v_{t+1-k}\right] v_{t+1}\right) \\
&= c_{t+1} E(y_1^2 v_{t+1}^2) \\
&= c_{t+1}\sigma^2 E(y_1^2) \\
&= c_{t+1}\sigma^2 [\tfrac{\sigma_\eta^2}{(1-\alpha)^2} + \tfrac{\sigma^2}{1-\alpha^2}]
\end{aligned}
$$

$$
\begin{aligned}
E(v_{t+1}^* y_1^2 y_{t+2}) &= c_{t+1} E\left[(v_{t+1} - \tfrac{1}{T-t-1}\textstyle\sum_{j=t+2}^{T} v_k) y_1^2 y_{t+2}\right] \\
&= c_{t+1}\left\{ E(v_{t+1} y_1^2 y_{t+2}) - \tfrac{1}{T-t-1} E(v_{t+2} y_1^2 y_{t+2}) \right\} \\
&= c_{t+1}\left\{ E(y_1^2[\alpha^2 y_t + (1-\alpha)\eta_i + v_{t+2} + \alpha v_{t+1}] v_{t+1}) - \tfrac{1}{T-t-1}\sigma^2 E(y_1^2) \right\} \\
&= c_{t+1}\left\{ \alpha\sigma^2 E(y_1^2) - \tfrac{1}{T-t-1}\sigma^2 E(y_1^2) \right\} \\
&= c_{t+1}\sigma^2 [\tfrac{\sigma_\eta^2}{(1-\alpha)^2} + \tfrac{\sigma^2}{1-\alpha^2}](\alpha - \tfrac{1}{T-t-1}) \\
&\vdots \\
E(v_{t+1}^* y_1^2 y_{T-1}) &= c_{t+1}\sigma^2 [\tfrac{\sigma_\eta^2}{(1-\alpha)^2} + \tfrac{\sigma^2}{1-\alpha^2}] \left(\alpha^{T-t-2} - \tfrac{1}{T-t-1}[1+\alpha+\cdots+\alpha^{T-t-3}]\right)
\end{aligned}
$$

$$
\Downarrow \qquad\qquad \Downarrow
$$

$$
E(D_{1t}^3) = \tfrac{c_{t+1}\sigma^2}{1-\alpha} [\tfrac{\sigma_\eta^2}{(1-\alpha)^2} + \tfrac{\sigma^2}{1-\alpha^2}] \left[(1-\alpha^{T-t-1})(1 + \tfrac{1}{(T-t-1)(1-\alpha)}) - 1\right]
$$

$E(D_{1t}^4)$ will require a similar calculation. We first make the following two assumptions: $E(v_{it}^3) = E(\eta_i^3) = 0 \ \forall\ i, t$. Now we solve for $E(D_{1t}^4)$:

$$
\begin{aligned}
E(D_{1t}^4) &= E(v_{t+1}^* y_t^2 \textstyle\sum_{k=t}^{T-1} y_t^2) \\
&= \underbrace{E(v_{t+1}^* y_t^4)}_{0} + E(v_{t+1}^* y_t^2 y_{t+1}^2) + \cdots + E(v_{t+1}^* y_t^2 y_{T-1}^2)
\end{aligned}
$$

$$
\begin{aligned}
E(v_{t+1}^* y_t^2 y_{t+1}^2) &= E\left(v_{t+1}^* y_t^2 [\alpha y_t + \eta_i + v_{t+1}]^2\right) \\
&= E\left(v_{t+1}^* y_t^2 [\alpha^2 y_t^2 + \eta_i^2 + v_{t+1}^2 + 2\alpha y_t \eta_i + 2\alpha y_t v_{t+1} + 2\eta_i v_{t+1}]\right) \\
&= 2\alpha E(v_{t+1}^* y_t^3 v_{t+1}) + E(v_{t+1}^* y_t^2 v_{t+1}^2) + 0 \\
&= 2\alpha c_{t+1} E(v_{t+1}^2 y_t^3) + c_{t+1} E(v_{t+1}^3 y_t^2) \\
&= 2\alpha c_{t+1} \sigma^2 E(Y_t^3) + c_{t+1} \underbrace{E(v_{t+1}^3)}_{0} E(y_t^2) \\
&= 2\alpha c_{t+1} \sigma^2 E\left([\tfrac{\eta_i}{1-\alpha} + \textstyle\sum_{k=0}^{\infty} \alpha^k v_{t-k}]^3 \right) \\
&= 2\alpha c_{t+1} \sigma^2 \left\{ E(\tfrac{\eta_i^3}{(1-\alpha)^3}) + \textstyle\sum_{k=0}^{\infty} \alpha^k \underbrace{E(v_{t-k}^3)}_{0} + \underbrace{E(crossterms)}_{0} \right\} \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
E(v_{t+1}^* y_t^2 y_{t+2}^2) &= E\left(v_{t+1}^* y_t^2 [\alpha^2 y_t + \eta_i + \alpha\eta_i + v_{t+2} + \alpha v_{t+1}]^2\right) \\
&= E(v_{t+1}^* y_t^2 [\alpha^4 y_t^2 + (1+\alpha)^2 \eta_i^2 + v_{t+2}^2 + \alpha^2 v_{t+1}^2 + 2\alpha^2 y_t \eta_i (1+\alpha) \\
&\quad + 2\alpha^2 y_t v_{t+2} + 2\alpha^3 y_t v_{t+1} + 2(1+\alpha)\eta_i v_{t+2} + 2(1+\alpha)\alpha\eta_i v_{t+1} + 2\alpha v_{t+1} v_{t+2}]) \\
&= \underbrace{E(v_{t+1}^* y_t^2 v_{t+2}^2) + \alpha^2 E(v_{t+1}^* y_t^2 v_{t+1}^2) + 2\alpha^2 E(v_{t+1}^* y_t^3 v_{t+2}) + 2\alpha^3 E(v_{t+1}^* y_t^3 v_{t+1})}_{All\, four\, terms\, are\, 0\, from\, E(v_t^3)=0, E(y_t^3)=0} \\
&\quad + 2\alpha \underbrace{E(v_{t+1}^* y_t^2 v_{t+1} v_{t+2})}_{0\, by\, independence\, of\, v_{t+1}, v_{t+2}} \\
&= 0 \\
\vdots \quad &\vdots \quad \vdots \\
E(v_{t+1}^* y_t^2 y_{T-1}^2) &= 0 \\
\Downarrow \quad &\quad \Downarrow \\
E(D_{1t}^4) &= 0
\end{aligned}
$$

Plugging in $E(D_{1t}^1) = E(D_{1t}^2) = E(D_{1t}^4) = 0$, and $E(D_{1t}^3)$ into $E(D_{1t})$, we have:

$$
\begin{aligned}
E(D_{1t}) &= -\frac{c_{t+1}\sigma^2}{(1-\alpha)(T-t-1)}\left[\frac{\sigma_\eta^2}{(1-\alpha)^2}+\frac{\sigma^2}{1-\alpha^2}\right]\left[(1-\alpha^{T-t-1})(1+\frac{1}{(T-t-1)(1-\alpha)})-1\right] \\
&= \left(\frac{p+q}{c_{t+1}}\right)L_t
\end{aligned}
$$

A.6.2 $E(D_{2t})$

$$
\begin{aligned}
E(D_{2t}) &= E\left(v_{t+1}^*\left[y_t-\frac{1}{T-t-1}\sum_{k=t+1}^{T-1}y_k\right](y_1y_t)\right) \\
&= (1-\frac{1}{T-t-1})\underbrace{E(v_{t+1}^*y_1y_t^2)}_{0}-\frac{1}{T-t-1}\sum_{t+1}^{T-1}E(v_{t+1}^*y_1y_ty_k)
\end{aligned}
$$

$$
\begin{aligned}
y_t &= \alpha^{t-1}y_1+(1+\alpha+\cdots+\alpha^{t-2})\eta_i+\sum_{k=0}^{t-2}\alpha^k v_{t-k} \\
\vdots &\quad \vdots \quad \vdots \\
y_{T-1} &= \alpha^{T-2}y_1+(1+\alpha+\cdots+\alpha^{T-3})\eta_i+\sum_{k=0}^{T-3}\alpha^k v_{T-1-k}
\end{aligned}
$$

The only parts of the term y_ty_{t+k} $\{k>0\}$ that will matter are the $v_{t+k}y_1$ $\{k>0\}$ terms.

$$
\begin{aligned}
y_ty_{t+1} &= \alpha^{t-1}y_1v_{t+1}+(other\ stuff) \\
E(v_{t+1}^*y_1y_ty_{t+2}) &= \alpha^{t-1}c_{t+1}E(y_1^2v_{t+1}^2) \\
&= \alpha^{t-1}c_{t+1}\sigma^2\left[\frac{\sigma_\eta^2}{(1-\alpha)^2}+\frac{\sigma^2}{1-\alpha^2}\right]
\end{aligned}
$$

$$
\begin{aligned}
y_ty_{t+2} &= \alpha^{t-1}y_1(\alpha v_{t+1}+v_{t+2})+(other\ stuff) \\
E(v_{t+1}^*y_1y_ty_{t+2}) &= \alpha^{t-1}c_{t+1}\{\alpha E(y_1^2v_{t+1}^2)-\frac{1}{T-t-1}E(y_1^2v_{t+2}^2)\} \\
&= \alpha^{t-1}c_{t+1}\sigma^2\left[\frac{\sigma_\eta^2}{(1-\alpha)^2}+\frac{\sigma^2}{1-\alpha^2}\right](\alpha-\frac{1}{T-t-1}) \\
\vdots &\quad \vdots \quad \vdots \\
E(v_{t+1}^*y_1y_ty_{T-1}) &= \alpha^{t-1}c_{t+1}\sigma^2\left[\frac{\sigma_\eta^2}{(1-\alpha)^2}+\frac{\sigma^2}{1-\alpha^2}\right][\alpha^{T-t-2}-\frac{1}{T-t-1}(1+\alpha+\cdots+\alpha^{T-t-3})]
\end{aligned}
$$

$$
\Downarrow \qquad \Downarrow
$$

$$
\begin{aligned}
E(D_{2t}) &= \alpha^{t-1}E(D_{1t}) \\
&= \alpha^{t-1}\left(\frac{p+q}{c_{t+1}}\right)L_t
\end{aligned}
$$

A.6.3 $E(D_{3t})$

$$\begin{aligned}
E(D_{3t}) &= E\left(v^*_{t+1}\left[y_t - \tfrac{1}{T-t-1}\textstyle\sum_{k=t}^{T-1} y_k\right]\textstyle\sum_{k=2}^{t-1} y_k^2\right) \\
&= \underbrace{E(v^*_{t+1}y_t\sum_{k=2}^{t-1} y_k^2)}_{0} - \tfrac{1}{T-t-1}E\left(v^*_{t+1}\left[\textstyle\sum_{k=t}^{T-1} y_k\right]\textstyle\sum_{k=2}^{t-1} y_k^2\right) \\
&= -\tfrac{1}{T-t-1}\underbrace{E\left(v^*_{t+1}y_t\sum_{k=2}^{t-1} y_k^2\right)}_{0} - \tfrac{1}{T-t-1}E\left(v^*_{t+1}\left[\textstyle\sum_{k=t+1}^{T-1} y_k\right]\textstyle\sum_{k=2}^{t-1} y_k^2\right) \\
&= -\tfrac{1}{T-t-1}E\left(v^*_{t+1}\left[\textstyle\sum_{k=t+1}^{T-1} y_k\right]\textstyle\sum_{k=2}^{t-1} y_k^2\right)
\end{aligned}$$

$$\begin{aligned}
E(v^*_{t+1}y_{t+1}y_2^2) &= E(v^*_{t+1}[\alpha y_t + \eta_i + v_{t+1}]y_2^2) \\
&= E(v^*_{t+1}v_{t+1}y_2^2) \\
E(v^*_{t+1}y_{t+1}y_2^2) &= c_{t+1}\sigma^2 E(y_2^2) \\
\vdots \quad & \vdots \quad \vdots \\
E(v^*_{t+1}y_{t+1}y_{t-1}^2) &= c_{t+1}\sigma^2 E(y_2^2)
\end{aligned}$$

$$\begin{aligned}
E(v^*_{t+1}y_{t+1}\textstyle\sum_{k=2}^{t-1} y_k^2) &= (t-2)c_{t+1}\sigma^2 E(y_2^2) \\
E(v^*_{t+1}y_{t+2}\textstyle\sum_{k=2}^{t-1} y_k^2) &= (t-2)c_{t+1}\sigma^2 E(y_2^2)[\alpha - \tfrac{1}{T-t-1}] \\
\vdots \quad & \vdots \quad \vdots \\
E(v^*_{t+1}y_{T-1}\textstyle\sum_{k=2}^{t-1} y_k^2) &= (t-2)c_{t+1}\sigma^2 E(y_2^2)\left[\alpha^{T-t-2} - \tfrac{1}{T-t-1}(1+\alpha+\cdots+\alpha^{T-t-3})\right]
\end{aligned}$$

$$\Downarrow \qquad \Downarrow$$

$$\begin{aligned}
E(D_{3t}) &= (t-2)E(D_{1t}) \\
&= (t-2)\left(\tfrac{p+q}{c_{t+1}}\right)L_t
\end{aligned}$$

A.6.4 $E(D_{4t})$

$$
\begin{aligned}
E(D_{4t}) &= E\left(v^*_{t+1}\left(y_t - \tfrac{1}{T-t-1}\textstyle\sum_{k=t}^{T-1} y_k\right)\sum_{k=1}^{t-1} y_k y_{k+1}\right) \\
&= \underbrace{E(v^*_{t+1} y_t \sum_{k=1}^{t-1} y_k y_{k+1})}_{0} - \tfrac{1}{T-t-1} E\left(v^*_{t+1}\left[\textstyle\sum_{k=t}^{T-1} y_k\right]\sum_{k=1}^{t-1} y_k y_{k+1}\right) \\
&= -\tfrac{1}{T-t-1} E\left(v^*_{t+1}\left[\textstyle\sum_{k=t+1}^{T-1} y_k\right]\sum_{k=1}^{t-1} y_k y_{k+1}\right) \\
&\quad - \tfrac{1}{T-t-1}\underbrace{E\left(v^*_{t+1} y_t \sum_{k=1}^{t-1} y_k y_{k+1}\right)}_{0}
\end{aligned}
$$

$$
\begin{aligned}
E\left(v^*_{t+1} y_{t+1} \textstyle\sum_{k=1}^{t-1} y_k y_{k+1}\right) &= c_{t+1} E(v^2_{t+1} \textstyle\sum_{k=1}^{t-1} y_k y_{k+1}) \\
&= c_{t+1}\sigma^2 E(\textstyle\sum_{k=1}^{t-1} y_k y_{k+1})
\end{aligned}
$$

$$
\begin{aligned}
E(y_1 y_2) &= E(\alpha y_1^2 + y_1\eta_i + y_1 v_2) \\
&= \alpha E(y_1^2) + E(y_1\eta_i) \\
\vdots \;&\vdots\; \vdots \\
E(y_{t-1} y_t) &= \alpha E(y_1^2) + E(y_1\eta_i)
\end{aligned}
$$

$$
\begin{aligned}
E\left(v^*_{t+1} y_{t+1} \textstyle\sum_{k=1}^{t-1} y_k y_{k+1}\right) &= (t-1)c_{t+1}\sigma^2[\alpha E(y_1^2) + E(y_1\eta_i)] \\
E\left(v^*_{t+1} y_{t+2} \textstyle\sum_{k=1}^{t-1} y_k y_{k+1}\right) &= (t-1)c_{t+1}\sigma^2[\alpha E(y_1^2) + E(y_1\eta_i)](\alpha - \tfrac{1}{T-t-1}) \\
\vdots \;&\vdots\; \vdots \\
E\left(v^*_{t+1} y_{T-1} \textstyle\sum_{k=1}^{t-1} y_k y_{k+1}\right) &= (t-1)c_{t+1}\sigma^2[\alpha^{T-t-2} - \tfrac{1}{T-t-1}(1+\alpha+\cdots+\alpha^{T-t-3})] \\
&\quad \times(\alpha E(y_1^2) + E(y_1\eta_i))
\end{aligned}
$$

$$
E(y_t\eta_i) = \frac{\sigma_\eta^2}{1-\alpha}\ \forall i, t
$$

$$
\Downarrow \qquad \Downarrow
$$

$$
\begin{aligned}
E(D_{4t}) &= \alpha(t-1)E(D_{1t}) \\
&\quad - \tfrac{c_{t+1}\sigma^2(t-1)}{(1-\alpha)(T-t-1)}\left[\tfrac{\sigma_\eta^2}{1-\alpha}\right]\left[(1-\alpha^{T-t-1})(1+\tfrac{1}{(T-t-1)(1-\alpha)}) - 1\right] \\
&= (t-1)\left(\tfrac{\alpha p + q}{c_{t+1}}\right) L_t
\end{aligned}
$$

A.6.5 $E(D_{5t})$

$$\begin{aligned}
E(D_{5t}) &= E(v^*_{t+1}\left(y_t - \tfrac{1}{T-t-1}\textstyle\sum_{k=t+1}^{T-1} y_k\right)\textstyle\sum_{j\neq k}\sum_{k=2}^{t-1} y_j y_k) \\
&= \underbrace{E(v^*_{t+1} y_t \sum_{j\neq k}\sum_{k=2}^{t-1} y_j y_k)}_{0} - \tfrac{1}{T-t-1}E(v^*_{t+1}\left[\textstyle\sum_{k=t}^{T-1} y_k\right]\textstyle\sum_{j\neq k}\sum_{k=2}^{t-1} y_j y_k) \\
&= -\tfrac{1}{T-t-1}E(v^*_{t+1}\left[\textstyle\sum_{k=t+1}^{T-1} y_k\right]\textstyle\sum_{j\neq k}\sum_{k=2}^{t-1} y_j y_k) - \tfrac{1}{T-t-1}\underbrace{E(v^*_{t+1} y_t \sum_{j\neq k}\sum_{k=2}^{t-1} y_j y_k)}_{0}
\end{aligned}$$

$$\begin{aligned}
E(v^*_{t+1} y_{t+1}\textstyle\sum_{j\neq k}\sum_{k=2}^{t-1} y_j y_k) &= c_{t+1}E(v^2_{t+1}\textstyle\sum_{j\neq k}\sum_{k=2}^{t-1} y_j y_k) \\
&= c_{t+1}\sigma^2 E(\textstyle\sum_{j\neq k}\sum_{k=2}^{t-1} y_j y_k) \\
&= 2c_{t+1}\sigma^2 E(\textstyle\sum_{2\leq j<k}^{t-1} y_j y_k) \\
&= 2c_{t+1}\sigma^2\left(E(y_2\textstyle\sum_{k=3}^{t-1} y_k) + E(y_3\sum_{k=4}^{t-1} y_k) + \cdots + E(y_{t-2}y_{t-1})\right)
\end{aligned}$$

$$\begin{aligned}
E(y_{t-2}y_{t-1}) &= E(y_{t-2}(\alpha y_{t-2} + \eta_i)) \\
&= \alpha E(y_{t-2}) + E(y_{t-2}\eta_i) \\
&= \alpha(p+q) + (1-\alpha)q \\
&= \alpha p + q
\end{aligned}$$

$$\begin{aligned}
E(y_{t-3}(y_{t-2} + y_{T-1}) &= E(y_{t-3}(\alpha y_{t-3} + \eta_i)) + E(y_{t-3}(\alpha^2 y_{t-3} + (1+\alpha)\eta_i)) \\
&= \alpha(1+\alpha)E(y_{t-3}) + (2+\alpha)E(y_{t-3}\eta_i) \\
&= \alpha(1+\alpha)(p+q) + (2+\alpha)(1-\alpha)q \\
&= \alpha(1+\alpha)p + 2q \\
\Downarrow \quad & \quad \Downarrow \\
E(y_s\textstyle\sum_{k=s+1}^{t-1} y_k) &= \tfrac{\alpha}{1-\alpha}(1-\alpha^{t-1-s})p + (t-1-s)q
\end{aligned}$$

Summing up, we have:

$$E(v^*_{t+1} y_{t+1}\textstyle\sum_{j\neq k}\sum_{k=2}^{t-1} y_j y_k) = 2c_{t+1}\sigma^2\left(\tfrac{\alpha}{1-\alpha}[(t-3) - \tfrac{\alpha}{1-\alpha}(1-\alpha^{t-1})]p + \tfrac{(t-3)(t-2)}{2}q\right)$$

We can write in the other terms now:

$$E(v^*_{t+1}y_{t+2}\sum_{j\neq k}\sum_{k=2}^{t-1}y_jy_k) = 2c_{t+1}\sigma^2\left(\tfrac{\alpha}{1-\alpha}[(t-3)-\tfrac{\alpha}{1-\alpha}(1-\alpha^{t-1})]p+\tfrac{(t-3)(t-2)}{2}q\right) \times(\alpha-\tfrac{1}{T-t-1})$$

$$\vdots \quad \vdots \quad \vdots$$

$$E(v^*_{t+1}y_{T-1}\sum_{j\neq k}\sum_{k=2}^{t-1}y_jy_k) = 2c_{t+1}\sigma^2\left(\tfrac{\alpha}{1-\alpha}[(t-3)-\tfrac{\alpha}{1-\alpha}(1-\alpha^{t-1})]p+\tfrac{(t-3)(t-2)}{2}q\right) \times(\alpha^{T-t-2}-\tfrac{1}{T-t-1}(1+\alpha+\cdots+\alpha^{T-t-3})$$

Putting the pieces together, we have:

$$E(D_{5t}) = (2/c_{t+1})\left(\tfrac{\alpha}{1-\alpha}[(t-3)-\tfrac{\alpha}{1-\alpha}(1-\alpha^{t-1})]p+\tfrac{(t-3)(t-2)}{2}q\right)L_t$$

A.7 Theorem 3

$$E\left(G_i'\Omega^{-1}g_i\right)=\sum_{t=1}^{T-2}B(\alpha,\sigma,T,t)*Term(t)$$

$$B(\alpha,\sigma,T,t)=-\left[1+\alpha^{T-t-1}[\alpha(T-t-1)-(T-t)]\right]/\left[\sigma^2(1-\alpha)^4(1+\alpha)(T-t)(T-t-1)\right]$$

$$Term(t) = \begin{cases} 2\sigma_\eta^2\alpha(1-\alpha^2)-\sigma^2(1-\alpha)^3 & t=1 \\ \left[\sigma_\eta^4(1+\alpha)^2+2\sigma^2\sigma_\eta^2\alpha(1-\alpha^2)-\sigma^4(1-\alpha)^2(1-2\alpha)\right]/\left[2\sigma_\eta^2+\sigma^2(1-\alpha)\right] & t=2 \\ \left[\sigma^4C(\alpha,t)+\sigma^2\sigma_\eta^2D(\alpha,t)+\sigma_\eta^4E(\alpha,t)\right]/\left[\sigma^2(1-\alpha)+\sigma_\eta^2(t(1-\alpha)+2\alpha)\right] & t\geq 3 \end{cases}$$

$$\begin{aligned} C(\alpha,t) &= (1-\alpha)^2(1-t(1-\alpha^2)) \\ D(\alpha,t) &= (1-\alpha)\left[(1-\alpha)[2\alpha(\alpha^t+\alpha^{t-2}-2)-1-t(t(1-\alpha^2)+2\alpha(1+\alpha)-1)]+1+3\alpha\right] \\ E(\alpha,t) &= (1+\alpha)\left[(1-\alpha)[5-3t+2\alpha^{t-1}+4\alpha(t-2)]+2\alpha\right] \end{aligned}$$

Proof: Combining Sections A.5 and A.6 gives:

$$E\left(G_i'\Omega^{-1}g_i\right)=\sum_{t=1}^{T-2}E\left(G_t'A_t^{-1}g_t\right)$$

$$E\left(G_t'A_t^{-1}g_t\right) = \begin{cases} \tfrac{p+q}{\sigma^4}L_1\left[h_1(1+\alpha)^2-(1-\alpha)^2\right] & t=1 \\ \tfrac{p+q}{\sigma^4}L_2\left[B_{12}+(2\alpha+B_{22})\right] & t=2 \\ \tfrac{1}{\sigma^4}L_t\left\{\left[B_{1t}+B_{2t}\alpha^{t-1}+B_{3t}(t-2)\right](p+q)+B_{4t}(t-1)(\alpha p+q)+B_{5t}B_{6t}\right\} & for\ t\geq 3 \end{cases}$$

$$B_{1t} = h_t - 1 \quad B_{2t} = 2h_t \quad B_{3t} = h_t(1-\alpha)^2 - (1+\alpha^2) \quad B_{4t} = 2\alpha \quad B_{5t} = h_t(1-\alpha)^2$$

$$\begin{aligned} B_{6t} &= 2\left(\tfrac{\alpha}{1-\alpha}[(t-3) - \tfrac{\alpha}{1-\alpha}(1-\alpha^{t-1})]p + \tfrac{(t-3)(t-2)}{2}q\right) \\ h_t &= \sigma_\eta^2 / \left[\sigma^2 + \sigma_\eta^2\left((t-2) + \tfrac{2}{1-\alpha}\right)\right] \\ L_t &= -\tfrac{\sigma^2}{(1-\alpha)(T-t)}\left[\left(1-\alpha^{T-t-1}\right)\left(1 + \tfrac{1}{(T-t-1)(1-\alpha)}\right) - 1\right] \end{aligned}$$

Plugging in everywhere for the $B_{j,t}$, $j = 1, \ldots, 6$ gives the following (somewhat simplified) representation of $E\left(G_t' A_t^{-1} g_t\right)$:

$$E\left(G_t' A_t^{-1} g_t\right) = \begin{cases} \frac{(p+q)L_1}{\sigma^4}\left[h_1(1+\alpha)^2 - (1-\alpha)^2\right] & t = 1 \\ \frac{(p+q)L_2}{\sigma^4}(3h_2 + 2\alpha - 1) & t = 2 \\ \frac{L_t}{\sigma^4}\left(p[h_t\hat{C}(\alpha,t) + \hat{E}(\alpha,t)] + q[h_t\hat{D}(\alpha,t) + \hat{F}(\alpha,t)]\right) & t \geq 3 \end{cases}$$

$$\begin{aligned} h_t &= \sigma_\eta^2 / \left[\sigma^2 + \sigma_\eta^2\left((t-2) + \tfrac{2}{1-\alpha}\right)\right] \\ L_t &= -\tfrac{\sigma^2}{(1-\alpha)(T-t)}\left[\left(1-\alpha^{T-t-1}\right)\left(1 + \tfrac{1}{(T-t-1)(1-\alpha)}\right) - 1\right] \end{aligned}$$

$$\begin{aligned} \hat{C}(\alpha,t) &= t(1-\alpha^2) - 1 + 2\alpha(\alpha^t + \alpha^{t-2} + \alpha - 1) & \hat{E}(\alpha,t) &= 1 - t(1-\alpha^2) \\ \hat{D}(\alpha,t) &= 1 + 2\alpha^{t-1} + (1-\alpha)^2(t-2)^2 & \hat{F}(\alpha,t) &= 1 - t(1-\alpha)^2 - 2\alpha(1-\alpha) \end{aligned}$$

We can simplify the expression further by plugging in for L_t, h_t, $\hat{C}(\alpha,t)$, $\hat{D}(\alpha,t)$,$\hat{E}(\alpha,t)$,$\hat{F}(\alpha,t)$ and solving. Some algebra gives us the final form.

A.8 Lemma 3

$$E(G_i(\alpha)) = (r_1\tilde{\alpha}_1 \cdots r_{T-2}\tilde{\alpha}_{T-2}), \quad where$$

$$\begin{aligned} \tilde{\alpha}_t &= (\alpha^{t-1} \cdots 1)' \\ r_t &= -\tfrac{c_{t+1}\sigma^2}{(1-\alpha)^2(1+\alpha)(T-t-1)}[(T-t)(1-\alpha) - (1-\alpha^{T-t})] \end{aligned}$$

Proof:

$$
\begin{aligned}
E(G_i) &= E(G_{i1} \cdots G_{i,T-2})' \\
E(G_{it}) &= -c_{t+1} E\left(\left[y_{it} - \tfrac{1}{T-t-1}\textstyle\sum_{k=t+1}^{T-1} y_{ik}\right] \tilde{y}_{it}\right)
\end{aligned}
$$

$$
\begin{aligned}
E(y_{it}\tilde{y}_{it}) &= E\left(y_{i1}y_{it} \;\; y_{i2}y_{it} \cdots y_{it}^2\right)' \\
&= E\left(\alpha^{t-1}p + q \;\cdots\; p+q\right)'
\end{aligned}
$$

$$
\begin{aligned}
E(\textstyle\sum_{k=t+1}^{T-1} y_{ik}\tilde{y}_{it}) &= E\left(y_{i1}\textstyle\sum_{k=t+1}^{T-1} y_{ik} \cdots y_{it}\sum_{k=t+1}^{T-1} y_{ik}\right)' \\
&= E\left(\tfrac{\alpha^t}{1-\alpha}(1-\alpha^{T-t-1})p + (T-t-1)q \;\cdots\; \tfrac{\alpha}{1-\alpha}(1-\alpha^{T-t-1})p + (T-t-1)q\right)'
\end{aligned}
$$

$$
\begin{aligned}
E(G_{it}) &= -c_{t+1}\begin{pmatrix} \begin{pmatrix} \alpha^{t-1}p+q \\ \alpha^{t-2}p+q \\ \vdots \\ \alpha^1 p+q \\ p+q \end{pmatrix} - \frac{\alpha(1-\alpha^{T-t-1})}{(T-t-1)(1-\alpha)}\begin{pmatrix} \alpha^{t-1}p \\ \alpha^{t-2}p \\ \vdots \\ \alpha p \\ p \end{pmatrix} - \begin{pmatrix} q \\ q \\ \vdots \\ q \\ q \end{pmatrix} \end{pmatrix} \\
&= -c_{t+1}p\left(1 - \frac{\alpha(1-\alpha^{T-t-1})}{(T-t-1)(1-\alpha)}\right)\begin{pmatrix} \alpha^{t-1} \\ \alpha^{t-2} \\ \vdots \\ \alpha \\ 1 \end{pmatrix} \\
&= r_t\tilde{\alpha}_t
\end{aligned}
$$

where $\tilde{\alpha}_t,\ r_t$ *are* :

$$
\begin{aligned}
\tilde{\alpha}_t &= (\alpha^{t-1} \cdots 1)' \\
r_t &= -\frac{c_{t+1}\sigma^2}{(1-\alpha)^2(1+\alpha)(T-t-1)}[(T-t)(1-\alpha) - (1-\alpha^{T-t})]
\end{aligned}
$$

$$
E(G_i(\alpha)) = (r_1\tilde{\alpha}_1 \cdots r_{T-2}\tilde{\alpha}_{T-2})
$$

A.9 Theorem 4

Plugging in for Ω_i^{-1} and $E(G_i(\alpha))$ (Lemma 3) gives:

$$
\Sigma = \frac{(1+\alpha)}{\sigma^4}\left(\sum_{t=1}^{T-2}\frac{T-t}{T-t-1}\left[1 - \frac{1-\alpha^{T-t}}{(1-\alpha)(T-t)}\right]^2\left(\frac{\sigma^2 + \sigma_\eta(t-1)}{\sigma^2(1-\alpha) + \sigma_\eta[t(1-\alpha)+2\alpha]}\right)\right)^{-1}
$$

Proof:

$$\Sigma = \left(\sum_{t=1}^{T-2} r_t^2 \tilde{\alpha}_t' A_t^{-1} \tilde{\alpha}_t\right)^{-1}$$

$$r_t^2 = \frac{\sigma^4(T-t)}{(1-\alpha^2)^2(T-t-1)}\left[1 - \frac{1-\alpha^{T-t}}{(1-\alpha)(T-t)}\right]^2$$

$$\tilde{\alpha}_t' A_t^{-1} \tilde{\alpha}_t = \tilde{\alpha}_t' N_t \tilde{\alpha}_t - h_t \tilde{\alpha}_t' R_t \tilde{\alpha}_t$$

$$\begin{aligned}\tilde{\alpha}_t' N_t \tilde{\alpha}_t &= (1-\alpha^2)\ \forall t \\ \tilde{\alpha}_t' R_t \tilde{\alpha}_t &= (1+\alpha)^2\ \forall t\end{aligned}$$

$$\begin{aligned}\tilde{\alpha}_t' A_t^{-1} \tilde{\alpha}_t &= (1-\alpha^2) - h_t(1+\alpha)^2\ \forall t \\ &= (1+\alpha)(1-\alpha)^2\left(\frac{\sigma^2+\sigma_\eta(t-1)}{\sigma^2(1-\alpha)+\sigma_\eta[t(1-\alpha)+2\alpha]}\right)\end{aligned}$$

$\Downarrow$ $\Downarrow$

$$\Sigma = \frac{(1+\alpha)}{\sigma^4}\left(\sum_{t=1}^{T-2}\frac{T-t}{T-t-1}\left[1 - \frac{1-\alpha^{T-t}}{(1-\alpha)(T-t)}\right]^2\left(\frac{\sigma^2+\sigma_\eta(t-1)}{\sigma^2(1-\alpha)+\sigma_\eta[t(1-\alpha)+2\alpha]}\right)\right)^{-1}$$

B Chapter 2 Appendix

Preliminary Definitions For a measurable map $(x, y, \theta) \mapsto f(y, x, \theta)$, define maps

$$\theta \mapsto E_n f(Y, X, \theta) \equiv \frac{1}{n} \sum_{t=1}^{n} f(Y_t, X_t, \theta),$$
$$\theta \mapsto G_n f(Y, X, \theta) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (f(Y_t, X_t, \theta) - E_{P_n} f(Y_t, X_t, \theta)).$$

B.1 Proof of Theorem 1

Define the process $Z_n \equiv \left\{ \sqrt{n}(\hat{\beta}_n(\tau) - \beta_n(\tau)), \tau \in \mathcal{N} \right\}$

1. Show *consistency* $Z_n(\cdot)/\sqrt{n} \xrightarrow{P_n^*} 0$. For some $\tilde{\beta}$, $\hat{\beta}_n(\tau)$ maximizes

$$\begin{aligned} Q_n(\tau, \beta) \equiv E_n \left[\rho_\tau(Y - X'\tilde{\beta}) - \rho_\tau(Y - X'\beta)\right] &\xrightarrow{P_n^*} \\ Q_\infty(\tau, \beta) \equiv E_P \left[\rho_\tau(Y - X'\tilde{\beta}) - \rho_\tau(Y - X'\beta)\right], & \end{aligned} \tag{5}$$

uniformly in $(\tau, \beta) \in \mathcal{T} \times \Theta$. Pointwise convergence is by **L**, since $(\tau, \beta) \mapsto \rho_\tau(Y - X'\tilde{\beta}) - \rho_\tau(Y - X'\beta)$ is bounded, and uniform convergence follows, since this map is linear in τ and Lipshitz in β (uniformly in Y, in $X \in \mathbf{X}$, and in τ).

The following extends Amemiya's (1973) proof for nonlinear estimators to the process case. For any $\epsilon > 0$, wp $\to 1$, uniformly in $\tau \in \mathcal{T}$:

- $Q_n(\tau, \hat{\beta}_n(\tau)) \geq Q_n(\tau, \beta(\tau))$, by definition,
- $Q_\infty(\tau, \hat{\beta}_n(\tau)) > Q_n(\tau, \hat{\beta}_n(\tau)) - \epsilon/2$, by (5),
- $Q_n(\tau, \beta(\tau)) > Q_\infty(\tau, \beta(\tau)) - \epsilon/2$, by (5).

Therefore, wp $\to 1$,

$$Q_\infty(\tau, \hat{\beta}_n(\tau)) > Q_n(\tau, \hat{\beta}_n(\tau)) - \epsilon/2 \geq Q_n(\tau, \beta(\tau)) - \epsilon/2 > Q_\infty(\tau, \beta(\tau)) - \epsilon.$$

Let $\{B(\tau), \tau \in (T)\}$ be a collection of balls radius $\delta/2$, each centered at $\beta(\tau)$. Then

$$\epsilon \equiv \inf_{\tau \in \mathcal{T}} \left[Q_\infty(\beta(\tau)) - \sup_{\beta_n \in \Theta \setminus B(\tau)} Q_\infty(\beta_n(\tau)) \right] > 0,$$

by assumption M (i) (put $T = \infty$). It now follows wp $\to 1$, uniformly in τ

$$Q_\infty(\hat{\beta}(\tau)) > Q_\infty(\beta(\tau)) - Q_\infty(\beta(\tau)) + \sup_{\beta_n \in \Theta \setminus B(\tau)} Q_\infty(\beta_n(\tau)) = \sup_{\beta_n \in \Theta \setminus B(\tau)} Q_\infty(\beta_n(\tau)).$$

Thus wp $\to 1$, $\sup_{\tau \in \mathcal{T}} \|\hat{\beta}_n(\tau) - \beta(\tau)\| \leq \delta$, for any $\delta > 0$.

Using similar steps, we can show $\sup_{\tau \in \mathcal{T}} \|\beta_n(\tau) - \beta(\tau)\| \leq \epsilon$, as $T \to \infty$, for any $\epsilon > 0$, exploiting that $\beta(\tau)$ maximizes $\tilde{Q}_n(\beta) \equiv E_{P_T} Q_n(\beta)$.

So $\sup_{\tau \in \mathcal{T}} \|\hat{\beta}_n(\tau) - \beta(\tau)\| \xrightarrow{p_n} 0$, $\sup_{\tau \in \mathcal{T}} \|\hat{\beta}_n(\tau) - \beta_n(\tau)\| \xrightarrow{p_n} 0$. ■

2. Next show asymptotic gaussianity of Z_n. First, by the computational properties of $\hat{\beta}_n(\tau)$, for all $\tau \in \mathcal{T}$, cf. [?]

$$E_n \varphi_\tau [Y - X'\hat{\beta}_n(\tau)] X = O\left(n^{-1}\right). \tag{6}$$

Second,

$$(\tau, \beta) \mapsto G_n \varphi_\tau \left[Y - X'\beta_n(\tau)\right] X \text{ is s.e. over } \Theta \times \mathcal{T} \tag{7}$$

by assumption E and linearity in τ, which implies that

$$\tau \mapsto G_n \varphi_\tau \left[Y - X'\beta_n(\tau)\right] X \text{ is s.e. over } \mathcal{T} \tag{8}$$

by assumption and continuity of $\tau \mapsto \beta_n(\tau)$. (7) and uniform consistency give

$$G_n \varphi_\tau [Y - X'\hat{\beta}_n(\tau)] X = G_n \varphi_\tau [Y - X'\beta_n(\tau)] X + o_{p*}(1), \text{ uniformly in } \tau. \tag{9}$$

Third, by M and uniform consistency, uniformly in τ

$$E_{P_n} \varphi_\tau [Y - X'\beta] X|_{\beta = \hat{\beta}_n(\tau)} = J_x(\tau)(\hat{\beta}_n(\tau) - \beta_n(\tau)) + o_{p*}\big(\sup_\tau \|\hat{\beta}_n(\tau) - \beta_n(\tau)\|\big) \tag{10}$$

Since the LHS of (6) $= n^{-1/2}$(lhs of(9) + lhs of(11)), we have uniformly in τ

$$\begin{aligned} O\left(T^{-1}\right) &= J_x(\tau)(\hat{\beta}_n(\tau) - \beta_n(\tau)) + o_{p*}\big(\sup_{\tau \in \mathcal{T}} \|\hat{\beta}_n(\tau) - \beta_n(\tau)\|\big) \\ &\quad + n^{-1/2} G_n[\varphi_\tau(Y - X'\beta_n(\tau) X] + o_{p^*}(n^{-1/2}). \end{aligned} \tag{11}$$

Since mineig $(J_x(\tau)) > \lambda$, uniformly in τ and $n > n_o$ by **M**, for some $\lambda > 0$,

$$\begin{aligned} &\sup_\tau \left|\left| n^{-1/2} G_n[\varphi_\tau [Y - X'\beta_n(\tau)] X + o_{p^*}(n^{-1/2}) \right|\right| \\ &\quad = \sup_\tau \left|\left| J_x(\tau)(\hat{\beta}_n(\tau) - \beta_n(\tau)) + o_{p^*}\big(\sup_\tau \|\hat{\beta}_n(\tau) - \beta_n(\tau)\|\big) \right|\right| \\ &\quad \geq (\lambda - o_{p^*}(1)) \cdot \sup_\tau \|\hat{\beta}_n(\tau) - \beta_n(\tau)\|, \end{aligned} \tag{12}$$

where $\lambda \geq 0$. By (8) and condition **L**

$$G_n \left[\varphi_\tau(Y - X'\beta_n(\tau)X\right] \Rightarrow v(\tau) \text{ in } \ell^\infty(\mathcal{T}), \tag{13}$$

where v is a P-Brownian Bridge with covariance function $w_v(\tau, \tau') \equiv V_M(k(\tau), k(\tau'))$, where $k(\tau) = \varphi_\tau(Y - X'\beta(\tau))X$. Therefore, the left-hand side of (12) is $O_{p^*}(n^{-1/2})$, hence

$$\begin{aligned}\sqrt{n}(\hat{\beta}_n(\tau) - \beta_n(\tau)) &= J_x^{-1}(\tau) G_n \left[\varphi_\tau(Y - X'\beta_n(\tau)\right] + o_{p^*}(1) \\ &\Rightarrow J_x^{-1}(\tau) \cdot v(\tau) \text{ in } \ell^\infty(\mathcal{T}).\end{aligned} \tag{14}$$

by (11) and (13). ■

B.2 Proof of Theorem 2

A special case of Theorem 3. ■

B.3 Proof of Theorem 3

Write

$$\begin{aligned}\hat{\mu}(\tau, g) &\equiv \sqrt{n} E_n \left[\varphi_\tau(Y - X'\hat{\beta}_n(\tau)) g(X)\right] \\ &\equiv \sqrt{n} E_{P_n} [\varphi_\tau(Y - X'\beta) g(X)]_{\beta = \hat{\beta}_n(\tau)} \\ &+ \sqrt{n} G_n \ \left[\varphi_\tau(Y - X'\beta) g(X)\right]_{\beta = \hat{\beta}_n(\tau)}.\end{aligned} \tag{15}$$

We have uniformly in $\{(\beta, \tau) : \|\beta - \beta(\tau)\| \leq \delta_n, \tau \in \mathcal{T}\}$ for $\delta_n \downarrow 0$

$$E_{P_n} \left[\varphi_\tau(Y - X'\beta) g(X)\right] = E_P \left[f_{Y|X}(X'\beta(\tau)) g(X) X'\right] (\beta - \beta_n(\tau)) + o\left(\delta_n\right) \tag{16}$$

By assumption **E** and linearity in τ, $\sqrt{n} G_n \left[\varphi_\tau(Y - X'\beta) g(X)\right]$ is s.e. in $\ell^\infty(\mathcal{T} \times \Theta \times \mathcal{G})$. Therefore, uniformly in τ

$$\sqrt{n} G_n \ \left[\varphi_\tau(Y - X'\beta) g(X)\right]_{\beta = \hat{\beta}_n(\tau)} \equiv \sqrt{n} G_n \ \left[\varphi_\tau(Y - X'\beta_n(\tau)) g(X)\right] + o_{p^*}(1). \tag{17}$$

uniformly in τ. Combining terms,

$$\begin{aligned}\hat{\mu}(\tau, g) &\equiv \sqrt{n} E_n \left[\varphi_\tau(Y - X'\hat{\beta}_n(\tau)) g(X)\right] \\ &= E_P \left[f_{Y|X}(X'\beta(\tau)) g(X) X'\right] \sqrt{n} \left(\hat{\beta}_n(\tau) - \beta_n(\tau)\right) + o_{p^*}(1) \\ &+ \sqrt{n} G_n \left[\varphi_\tau(Y - X'\beta_n(\tau)) g(X)\right] + o_{p^*}(1)\end{aligned} \tag{18}$$

Substituting the right-hand side of (14) into (18), conclude. ■

B.4 Proof of Theorem 4

The second part of the proof follows the arguments of PRW (1999), Theorem 2.6.1, 3.5.1. The first has to deal with the presence of estimated centering.

I. To prove (i), (ii), define $\dot{G}_{n,b}(x)$ as $\hat{G}_{n,b}(x)$ when $\hat{\mu}(\tau, g) = \mu(\tau, g)$, $\hat{W}(\tau, g) = W(\tau, g)$ (i.e. case with the known centering). $\dot{G}_{n,b}(x)$ is a U-statistic of degree b, with $E_P\hat{G}_{n,b}(x) = G_b(x, P) = P(\hat{\theta}_b \leq x)$. PRW (1999), cf. proofs of Theorem 2.2.1, Theorem 3.2.1, for dependent data, show that $\dot{G}_{n,b}(x) \xrightarrow{p_n} G(x)$. Next note that

$$\hat{G}_{n,b}(x) = B_n^{-1} \sum_{i \leq B_n} 1\left[\mathbf{f}\left(\left\|\hat{W}^{1/2}(r_b(\hat{\mu}_{n,b,i} - \mu) + r_b(\mu - \hat{\mu}))\right\|_2^2\right) < x\right],$$

$$\dot{G}_{n,b}(x) = B_n^{-1} \sum_{i \leq B_n} 1\left[\mathbf{f}\left(\left\|W^{1/2}(r_b(\hat{\mu}_{n,b,i} - \mu))\right\|_2^2\right) < x\right]$$

Collect three facts: fact 1, uniformly in i

$$\frac{1}{\dot{\lambda}_n} \leq \frac{\left\|\hat{W}^{1/2}(r_b(\hat{\mu}_{n,b,i} - \mu) + r_b(\mu - \hat{\mu}))\right\|_2^2}{\left\|W^{1/2}(r_b(\hat{\mu}_{n,b,i} - \mu) + r_b(\mu - \hat{\mu}))\right\|_2^2} \leq \bar{\lambda}_n,$$

where

$$\bar{\lambda}_n = \sup_{\tau, g \in \mathcal{L} \times \mathcal{G}} \text{maxeig}\left(W^{-1/2}(\tau, g)\hat{W}(\tau, g)W^{-1/2}\right),$$

$$\dot{\lambda}_n = \sup_{\tau, g \in \mathcal{L} \times \mathcal{G}} \text{maxeig}\left(\hat{W}^{-1/2}(\tau, g)W(\tau, g)\hat{W}^{-1/2}\right).$$

by inequality 10. on p.460 in Amemiya (1985).

Fact 2 is trivial:

$$\left\|\hat{W}^{1/2}(r_b(\hat{\mu}_{n,b,i}-\mu)+r_b(\mu-\hat{\mu}))\right\|_2^2 \leq \left\|\hat{W}^{1/2}(r_b(\hat{\mu}_{n,b,i}-\mu)\right\|_2^2+\left\|\hat{W}^{1/2}r_b(\mu-\hat{\mu}))\right\|_2^2$$

Fact 3 is, by the assumed properties of $\mathbf{f}$ need to check

$$1\left[\mathbf{f}\left(\left\|W^{1/2}(r_b(\hat{\mu}_{n,b,i} - \mu)\right\|_2^2\right) < \frac{(x - w_n)}{\dot{u}_n}\right] \leq 1\left[\mathbf{f}\left(\left\|\hat{W}^{1/2}(r_b(\hat{\mu}_{n,b,i} - \mu) + r_b(\mu - \hat{\mu}))\right\|_2^2\right) < x\right]$$
$$\leq 1\left[\mathbf{f}\left(\left\|W^{1/2}(r_b(\hat{\mu}_{n,b,i} - \mu)\right\|_2^2\right) < \frac{(x + w_n)}{l_n}\right]$$

where $l_n = \max(\bar{\lambda}_n, 1/\dot{\lambda}_n)$ and $u_n = min(\bar{\lambda}_n, 1/\dot{\lambda}_n)$, and v_n and w_n are defined below.

Noting that the error of estimation vanishes by Theorem 2 and 3

$$w_n \equiv \sup_{\tau,g} b \left\| W(\tau, g)(\hat{\mu}(\tau, g) - \mu(\tau, g)) \right\|_2 = bO_p(1)O_p(T^{-1}) \xrightarrow{P_n} 0$$

$$v_n \equiv \max\left[|u_n - 1|, |l_n - 1|\right] \xrightarrow{P_n} 0$$

so that with probability tending to one $1(E_n) = 1$, where $E_n\{v_n, w_n \leq \delta\}$ for any $\delta > 0$.

II. Thus, for small enough $\epsilon > 0$, there is $\delta > 0$, so that by Fact 3:

$$\dot{G}_{n,b}(x - \epsilon)1(E_n) \leq \hat{G}_{n,b}(x)1(E_n) \leq \dot{G}_{n,b}(x + \epsilon)$$

So that with probability tending to one: $\dot{G}_{n,b}(x - \epsilon) \leq \hat{G}_{n,b}(x) \leq \dot{G}_{n,b}(x + \epsilon)$. for any $\epsilon > 0$. Hence if $x + \epsilon$ and $x - \epsilon$ are continuity points of $G(x)$, then $\dot{G}_{n,b}(x + c) \xrightarrow{P_n} G(x - c)$, for $c = \epsilon, -\epsilon$, implies $G(x - \epsilon) - \epsilon \leq \hat{G}_{n,b}(x) \leq G(x + \epsilon) + \epsilon$ w.p. $\to 1$. Let $\epsilon \to 0$ and $\delta(\epsilon) \to 0$, so that $x + c$ are continuity points of $G(x)$. Hence $G_{n,b}(x) \xrightarrow{P_n} G(x)$. Finally, note that $x = G^{-1}(x)$ is assumed a continuity point.

Moreover, the convergence of quantiles implies the convergence of distribution functions at all continuity points.

III. Now $G(x) = H_0(x)$ under H_0, so that (i) and (ii) follow. To prove (iii), note that by (i): $c_{n,b}(1 - \alpha) \xrightarrow{P} H_0^{-1}(1 - \alpha)$. Contiguity of $\{P_n^{[n]}\}$ to $\{P^{[n]}\}$ forces that $c_{n,b}(1 - \alpha) \xrightarrow{P_n} H_0^{-1}(1 - \alpha)$

(iv) follows from Davydov's Theorem. ■

C Chapter 3 Appendix

C.1 Taylor Expansion of $\hat{A}\hat{g}$

We make the following assumptions:

- $g(\beta_0) = 0$. $\hat{\beta}$ is chosen as to set $\hat{A}\hat{g}(Z, \hat{\beta}) = 0$, where $\hat{A}$ is KXJ, $\hat{g}(Z, \beta)$ is $JX1$, $\hat{\beta}$ is $KX1$, Z is our data $Z_1, \ldots, Z_n$.
- $\hat{A} = A + \frac{1}{n}\sum_{i=1}^{n} \psi(z_i) + O_p(n^{-1})$, $E(\psi(z_i)) = 0$, $\hat{\Psi}(Z) = \frac{1}{n}\sum_{i=1}^{n} \psi(z_i)$
- $\hat{g}(\beta) = \frac{1}{n}\sum_{i=1}^{n} g(z_i, \beta)$
- $\hat{\beta}_j = \beta_{j0} + \hat{\Lambda}_j + O_p(n^{-1})$, where $\hat{\Lambda}_j = \frac{1}{n}\sum_{k=1}^{n} \lambda_{kj}$.

Then our GMM estimator has the following bias:

$$E(\hat{\beta}_{GMM} - \beta) = -\Sigma\left[E(\hat{\Psi}(Z)\hat{g}(\beta_0)) - E(\hat{\Psi}(Z)G(\beta_0)Q) - AE(\hat{G}(\beta_0)Q) - \frac{1}{2}\sum_{i}^{k} H_i)\right],$$

$$\Sigma = [AG(\beta_0)]^{-1}, \qquad H_i = E(\hat{\Lambda}_i A G_i'(\beta_0)Q) \qquad Q = \Sigma A\hat{g}(\beta_0)$$

Proof:

We will make use of the following notation:

- $\hat{\Psi} = \frac{1}{n}\sum_{i=1}^{n} \psi(z_i)$
- $\hat{h}(\beta) = \hat{A}\hat{g}(\beta)$
- $M = -[AG(\beta_0)]^{-1}$
- β_{i0} is the i^{th} component of β_0
- $\hat{G}(\beta_0) = \hat{g}'(\beta_0)$, where $\hat{g}'(\beta_0) = \frac{\partial \hat{g}(\beta)}{\partial \beta}|_{\beta=\beta_0}$
- $\hat{G}_i'(\beta_0) = \frac{\partial \hat{G}(\beta_0)}{\partial \beta_i}$, where β_i is the i^{th} component of β

We now carry out a Taylor Expansion of $\hat{A}\hat{g}(\hat{\beta})$ around β_0:

$$
\begin{aligned}
\hat{h}(\hat{\beta}) &= \hat{A}\hat{g}(\hat{\beta}) \\
&= 0 \\
&= \hat{A}\hat{g}(\beta_0) + \hat{A}\hat{G}(\beta_0)(\hat{\beta}-\beta_0) + \tfrac{1}{2}\textstyle\sum_{i=1}^{k}(\hat{\beta}_i-\beta_{i0})\hat{A}\hat{G}_i'(\beta_0)(\hat{\beta}-\beta_0) + o_p(n^{-1}) \\
&= \hat{A}\hat{g}(\beta_0) + [\hat{A}\hat{G}(\beta_0) - AG(\beta_0)](\hat{\beta}-\beta_0) + AG(\beta_0)(\hat{\beta}-\beta_0) \\
&\quad + \tfrac{1}{2}\textstyle\sum_{i}^{k}(\hat{\beta}_i-\beta_{i0})\hat{A}\hat{G}_i'(\beta_0)(\hat{\beta}-\beta_0) + o_p(n^{-1}) \\
[-AG(\beta_0)](\hat{\beta}-\beta_0) &= \hat{A}\hat{g}(\beta_0) + [\hat{A}\hat{G}(\beta_0) - AG(\beta_0)](\hat{\beta}-\beta_0) \\
&\quad + \tfrac{1}{2}\textstyle\sum_{i}^{k}(\hat{\beta}_i-\beta_{i0})\hat{A}\hat{G}_i'(\beta_0)(\hat{\beta}-\beta_0) + o_p(n^{-1})
\end{aligned}
$$

Let $M = [-AG(\beta_0)]^{-1}$. *Then,*

$$
\begin{aligned}
(\hat{\beta}-\beta_0) &= M\overbrace{\hat{A}\hat{g}(\beta_0)}^{t1} + M\overbrace{[\hat{A}\hat{G}(\beta_0) - AG(\beta_0)](\hat{\beta}-\beta_0)}^{t2} \\
&\quad + \underbrace{M\frac{1}{2}\sum_{i}^{k}(\hat{\beta}_i-\beta_{i0})\hat{A}\hat{G}_i'(\beta_0)(\hat{\beta}-\beta_0)}_{t3} + o_p(n^{-1})
\end{aligned}
$$

Let us replace the bias terms on the RHS :

$$
\begin{aligned}
(\hat{\beta}-\beta_0) &= M\hat{A}\hat{g}(\beta_0) + M[\hat{A}\hat{G}(\beta_0) - AG(\beta_0)]M[\hat{A}\hat{g}(\beta_0) + t2 + t3 + o_p(n^{-1})] \\
&\quad + \tfrac{M}{2}\textstyle\sum_{i}^{k}(\hat{\beta}_i-\beta_{i0})\hat{A}\hat{G}_i'(\beta_0)M[\hat{A}\hat{g}(\beta_0) + t2 + t3 + o_p(n^{-1})] + o_p(n^{-1})
\end{aligned}
$$

- $t2$ *is* $O_p(n^{-1}) \rightarrow [\hat{A}\hat{G}(\beta_0) - AG(\beta_0)]M * t2$ *is* $o_p(n^{-1})$
- $t2$ *is* $O_p(n^{-1}) \rightarrow (\hat{\beta}_i-\beta_{i0})\hat{A}\hat{G}_i'(\beta_0)M * t2$ *is* $o_p(n^{-1})$
- $t3$ *is* $O_p(n^{-1}) \rightarrow [\hat{A}\hat{G}(\beta_0) - AG(\beta_0)]M * t3$ *is* $o_p(n^{-1})$
- $t3$ *is* $O_p(n^{-1}) \rightarrow (\hat{\beta}_i-\beta_{i0})\hat{A}\hat{G}_i'(\beta_0)M * t3$ *is* $o_p(n^{-1})$

$$
\begin{aligned}
(\hat{\beta}-\beta_0) &= M\hat{A}\hat{g}(\beta_0) + M[\hat{A}\hat{G}(\beta_0) - AG(\beta_0)]M\hat{A}\hat{g}(\beta_0) \\
&\quad + \tfrac{M}{2}\textstyle\sum_{i}^{k}(\hat{\beta}_i-\beta_{i0})\hat{A}\hat{G}_i'(\beta_0)M\hat{A}\hat{g}(\beta_0) + o_p(n^{-1})
\end{aligned}
$$

- $\hat{A} = A + \hat{\Psi}(Z) + o_p(n^{-1}),\ \hat{\Psi}(Z)$ *is* $O_p(n^{-1/2})$
- $\hat{g}(\beta_0)$ *is* $O_p(n^{-1/2})$
- $\hat{A}\hat{G}(\beta_0) - AG(\beta_0)$ *is* $O_p(n^{-1/2})$
- $(\hat{\beta}_i-\beta_{i0})$ *is* $O_p(n^{-1/2})$

$$
\begin{aligned}
(\hat{\beta}-\beta_0) &= M\hat{A}\hat{g}(\beta_0) + M[\hat{A}\hat{G}(\beta_0) - AG(\beta_0)]M(A+O_p(n^{-1/2}))\hat{g}(\beta_0) \\
&\quad + \tfrac{M}{2}\textstyle\sum_i^k (\hat{\beta}_i - \beta_{i0})(A+O_p(n^{-1/2}))\hat{G}_i'(\beta_0)M(A+O_p(n^{-1/2}))\hat{g}(\beta_0) + o_p(n^{-1}) \\
&= M\hat{A}\hat{g}(\beta_0) + M[\hat{A}\hat{G}(\beta_0) - AG(\beta_0)]MA\hat{g}(\beta_0) \\
&\quad + \tfrac{M}{2}\textstyle\sum_i^k (\hat{\beta}_i - \beta_{i0})A\hat{G}_i'(\beta_0)MA\hat{g}(\beta_0) + o_p(n^{-1})
\end{aligned}
$$

Expand the 1^{st} *and* 2^{nd} *terms :*

$$
\begin{aligned}
(\hat{\beta}-\beta_0) &= MA\hat{g}(\beta_0) + M\hat{\Psi}(Z)\hat{g}(\beta_0) + M\hat{A}\hat{G}(\beta_0)MA\hat{g}(\beta_0) - MAG(\beta_0)MA\hat{g}(\beta_0) \\
&\quad + \tfrac{M}{2}\textstyle\sum_i^k (\hat{\beta}_i - \beta_{i0})A\hat{G}_i'(\beta_0)MA\hat{g}(\beta_0) + o_p(n^{-1})
\end{aligned}
$$

Plug in $\hat{A} = A + \hat{\Psi}(Z) + o_p(n^{-1})$,
Recall $M = [-AG(\beta_0)]^{-1} \rightarrow MAG(\beta_0) = -1$:

$$
\begin{aligned}
(\hat{\beta}-\beta_0) &= MA\hat{g}(\beta_0) + M\hat{\Psi}(Z)\hat{g}(\beta_0) + M(A+\hat{\Psi}(Z))\hat{G}(\beta_0)MA\hat{g}(\beta_0) + MA\hat{g}(\beta_0) \\
&\quad + \tfrac{M}{2}\textstyle\sum_i^k (\hat{\beta}_i - \beta_{i0})A\hat{G}_i'(\beta_0)MA\hat{g}(\beta_0) + o_p(n^{-1}) \\
&= 2MA\hat{g}(\beta_0) + M\hat{\Psi}(Z)\hat{g}(\beta_0) + MA\hat{G}(\beta_0)MA\hat{g}(\beta_0) + M\hat{\Psi}(Z)\hat{G}(\beta_0)MA\hat{g}(\beta_0) \\
&\quad + \tfrac{M}{2}\textstyle\sum_i^k (\hat{\beta}_i - \beta_{i0})A\hat{G}_i'(\beta_0)MA\hat{g}(\beta_0) + o_p(n^{-1})
\end{aligned}
$$

•$\hat{\Psi}(Z)$ *and* $\hat{g}(\beta_0)$ *are* $O_p(n^{-1/2})$, $\hat{G}(\beta_0) = G(\beta_0) + O_p(n^{-1/2})$
$\rightarrow \hat{\Psi}(Z)\hat{G}(\beta_0)MA\hat{g}(\beta_0) = \hat{\Psi}(Z)G(\beta_0)MA\hat{g}(\beta_0)$

•$(\hat{\beta}_i - \beta_{i0}) = \hat{\Lambda}_i + O_p(n^{-1})$
•$\hat{\Lambda}_i$ *and* $\hat{g}(\beta_0)$ *are* $O_p(n^{-1/2})$, $\hat{G}_i'(\beta_0) = G_i'(\beta_0) + O_p(n^{-1/2})$
$\rightarrow (\hat{\beta}_i - \beta_{i0})A\hat{G}_i'(\beta_0)MA\hat{g}(\beta_0) = \hat{\Lambda}_i AG_i'(\beta_0)MA\hat{g}(\beta_0)$

$$
\begin{aligned}
(\hat{\beta}-\beta_0) &= 2MA\hat{g}(\beta_0) + M\hat{\Psi}(Z)\hat{g}(\beta_0) + MA\hat{G}(\beta_0)MA\hat{g}(\beta_0) + M\hat{\Psi}(Z)G(\beta_0)MA\hat{g}(\beta_0) \\
&\quad + \tfrac{M}{2}\textstyle\sum_i^k \hat{\Lambda}_i AG_i'(\beta_0)MA\hat{g}(\beta_0) + o_p(n^{-1})
\end{aligned}
$$

Taking expectaions, we have :

$$\begin{aligned} E(\hat{\beta}-\beta_0) &= 2MAE(\hat{g}(\beta_0)) + ME(\hat{\Psi}(Z)\hat{g}(\beta_0)) + MAE(\hat{G}(\beta_0)MA\hat{g}(\beta_0)) \\ &\quad + ME(\hat{\Psi}(Z)G(\beta_0)MA\hat{g}(\beta_0)) + \tfrac{M}{2}\textstyle\sum_i^k E(\hat{\Lambda}_i AG_i'(\beta_0)MA\hat{g}(\beta_0)) \end{aligned}$$

$\bullet E(\hat{g}(\beta_0)) = 0$
$\bullet MA\hat{g}(\beta_0)) = (\hat{\beta}-\beta_0) + O_p(n^{-1})$. *Let* $Q = MA\hat{g}(\beta_0))$

$$\begin{aligned} E(\hat{\beta}-\beta_0) &= ME(\hat{\Psi}(Z)\hat{g}(\beta_0)) + MAE(\hat{G}(\beta_0)Q) \\ &\quad + ME(\hat{\Psi}(Z)G(\beta_0)Q) + \tfrac{M}{2}\textstyle\sum_i^k E(\hat{\Lambda}_i AG_i'(\beta_0)Q) \end{aligned}$$

And our formula for the bias of GMM with general weighting matrix $\hat{A}$ is given by:

$$\begin{aligned} E(\hat{\beta}-\beta_0) &= -[AG(\beta_0)]^{-1}\left[E(\hat{\Psi}(Z)\hat{g}(\beta_0)) + AE(\hat{G}(\beta_0)Q) + E(\hat{\Psi}(Z)G(\beta_0)Q) + \tfrac{1}{2}\textstyle\sum_i^k H_i\right] \\ H_i &= E(\hat{\Lambda}_i AG_i'(\beta_0)Q) \quad , \quad Q = -[AG(\beta_0)]^{-1}A\hat{g}(\beta_0) \end{aligned}$$

C.2 Theorem 2

Consider Arellano IVE estimator for the AR(1) model. Let $\theta = (a, f, m)$ where (a, f, m) are the pseudo-parameters defined in Section 5.1, and let $\hat{\Delta}_j = \sum_{i=1}^n \delta_{ij}$ be the influence function of θ_j. Then we have the following result:

$$E(\hat{\Psi}(Z)\hat{g}(\beta_0)) \leq O_p(\sqrt{T})$$

Proof: Recall from Example 2 that when $\hat{A}$ depends on a parameter $\hat{\theta}$ of fixed dimension, we have:

$$E(\hat{\Psi}(Z)\hat{g}(\beta_0)) = (1/n)^2 \sum_{i=1}^{n}\sum_{j=1}^{J} E(\delta_{ij}\frac{\partial A(\theta_0)}{\partial\theta_j} g(z_i, \beta_0))$$

where $\hat{\Delta}_j = \sum_{i=1}^n \delta_{ij}$ is the influence function of θ_j. For the Arellano IVE estimator, $j = 3, \theta_1 = a, \theta_2 = f, \theta_3 = m$. Consider calculating $\frac{\partial A(\theta_0)}{\partial\theta_j} g(z_i, \beta_0)$. It would seem from Example 1 that such a calculation would require an expression for A in terms of the π_t^s coefficients. In fact, we can simplify the work involved by noting that $\frac{\partial A}{\partial\theta_j} g_i = \frac{\partial A g_i}{\partial\theta_j}$ and

$$Ag_i = \sum_{t=1}^{T} h_t(y_i^{t-1}, \theta)v_{it}^*, where\ v_{it}^* = y_{it}^* - \alpha x_{it}^*\ and\ x_{it} = y_{it-1}$$

This implies:

$$\frac{\partial A}{\partial \theta_j} g_i = \sum_{t=1}^{T} \frac{\partial h_t}{\partial \theta_j} v_{it}^*$$

Recall from Lemma 1 that:

$$\begin{aligned} h_t(y_i^{t-1}, \theta) &= c_t \left[1 - \tfrac{a}{1-a}\left(\tfrac{1-a^{T-t}}{T-t}\right)\right]\left[y_{i,t-1} - m_{t-1}(y_i^{t-1}, c)\right] & t \geq 1 \\ m_{t-1}(y_i^{t-1},], theta) &= \tfrac{m + f\left[(1-a)\sum_{s=1}^{t-1} u_{is} + (1-a^2)y_{i0}\right]}{1 + f\left[(t-1)(1-a)^2 + 1 - a^2\right]} & t \geq 2 \end{aligned}$$

where $u_{is} = y_{is} - \alpha y_{i,s-1}$. Differentiating our instruments with respect to the parameters, we obtain:

$$\begin{aligned} \partial h_t/\partial a &= c_t(\partial S_t/\partial a)[y_{i,t-1} - m_{t-1}] - c_t S_t(\partial m_{t-1}/\partial a) \\ \partial h_t/\partial f &= -c_t S_t(\partial m_{t-1}/\partial f) \\ \partial h_t/\partial m &= -c_t S_t(\partial m_{t-1}/\partial m) \end{aligned}$$

As $t \to \infty$, we have:

- $m_{t-1} \to lim_{t\to\infty}(1/t)\sum_{s=0}^{t-1} y_{is} = E_t(y_{it})$.
  To see this, first note that:

  $$\sum_{s=1}^{t-1} u_{is} = y_{i,t-1} + (1-a)\sum_{s=1}^{t-2} y_{is} - ay_{i0}$$

  Plugging in and taking the limit yields the desired result.

- Since m_{t-1} does not depend on the parameters (a, f, m) as $t \to \infty$, we will have $\partial m_{t-1}/\partial \theta_j \to 0$ as $t \to \infty$.

These results imply $\partial h_t/\partial f \to 0$ and $\partial h_t/\partial m \to 0$ as $t \to \infty$. Also, for large T_0 and $t > T_0$, we have

$$\sum_{t=T_0}^{T} \frac{\partial h_t}{\partial a} v_{it}^* \simeq \sum_{t=T_0}^{T} R_t \left[y_{i,t-1} - (1/t)\sum_{s=0}^{t-1} y_{is}\right] v_{it}^*$$

where $R_t = c_t \left[\frac{a^{T-t}}{1-a} - \frac{1-a^{T-t}}{(1-a)^2}\frac{1}{T-t}\right]$. Indeed, this result is justified as $T \to \infty$ even if we did not consider $t > T_0$, since $R_t \simeq 0$ for small t. We now have:

$$\begin{aligned}
(\partial A/\partial a)g_i &\simeq \sum_{t=1}^{T} R_t \left[y_{i,t-1} - (1/t)\sum_{s=0}^{t-1} y_{is}\right] v_{it}^* \\
E\left[([\partial A/\partial a]g_i)^2\right] &\simeq \sum_{t=T_0}^{T} R_t^2 E\left[\left(y_{i,t-1} - (1/t)\sum_{s=0}^{t-1} y_{is}\right)^2 v_{it}^{*2}\right] \\
&\simeq O_p(T)
\end{aligned}$$

Since $\partial h_t/\partial f \to 0$ and $\partial h_t/\partial m \to 0$ as $t \to \infty$, we have

$$E\left[([\partial A/\partial \theta_j]g_i)^2\right] \leq O_p(T), \quad j = 1,2,3$$

This implies:

$$|E[\delta_{ij}(\partial A/\partial \theta_j)g_i]| \leq \sqrt{E(\delta_{ij}^2)E[([\partial A/\partial \theta_j]g_i)^2]} \leq O_p(\sqrt{T})$$

which in turn gives:

$$E(\hat{\Psi}(Z)\hat{g}(\beta_0)) \leq O_p(\sqrt{T})$$

References

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

[2] Alonso-Borrego C., and M. Arellano. (1999). "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data." *Journal of Business and Economic Statistics* 17: 36-49

[3] Arellano, M., and O. Bover. (1995). "Another Look at the Instrumental Variable Estimation of Error-Component Models." *Journal of Econometrics* 68: 1-268

[4] Arellano, M. (2003). "Modelling Optimal Instrumental Variables for Dynamic Panel Data Models." Econometrics Invited Lecture, European Meeting of the Econometric Society, Venice, August 2002. CEMFI Working Paper no. 0310.

[5] Bierens, H.J. (1990). "A Consistent Conditional Moment Test of Functional Form." *Econometrica* 58: 1443-1458

[6] Bierens, H.J., and W. Ploberger. (1997). "Asymptotic Theory of Integrated Conditional Moment Tests." *Econometrica* 65: 1129-1152

[7] Hahn, J. (1999). "How Informative is the Initial Condition in the Dynamic Panel Data Model with Fixed Effects?" *Journal of Econometrics* 93: 309-326

[8] Hahn, J., J.A. Hausman, and G. Kuersteiner. (2001). "Bias Corrected Instrumental Variables Estimators for Panel Data." Working Paper, Department of Economics, M.I.T..

[9] Hansen, L.P. (2001). "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50, 1269-1286

[10] Koenker, R., and G. Basset. (1978). "Regression Quantiles." *Econometrica* 46: 33-51

[11] Koenker, R., and Z. Xiao. (2001). "Inference on the Quantile Regression Process." *Econometrica* 70: 1583-1604

[12] Newey, W., and Smith, R. (2003). "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators." Working Paper, Presented at 2002 Winter Econometric Society Meetings

[13] Sakov, A., and P. Bickel. (2000). "An Edgeworth Expansion for the m out of n Bootstrapped Median." *Statistics and Probability Letters* 49: 217-223

[14] Stinchcombe, M., and H. White. (1998). "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative." *Econometric Theory* 14: 295-324

[15] Van Der Vaart, A.W. (1999). *Asymptotic Statistics.* Cambridge University Press.

[16] Van Der Vaart, A.W., and J.A. Wellner. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Tables

Table 1: Bias Calculations

T	n	α	Actual Bias	Actual % Bias	2nd Order Bias	2nd Order % Bias	Our Bias	Our % Bias
5	100	0.1	-0.016	-16.00	-0.018	-17.71	-0.015	-15.38
10	100	0.1	-0.014	-14.26	-0.016	-15.78	-0.018	-18.00
5	500	0.1	-0.004	-3.72	-0.004	-3.54	-0.003	-3.08
10	500	0.1	-0.003	-3.20	-0.003	-3.16	-0.004	-3.60
5	100	0.3	-0.028	-9.23	-0.032	-10.60	-0.016	-5.17
10	100	0.3	-0.021	-7.11	-0.024	-8.13	-0.026	-8.71
5	500	0.3	-0.006	-2.08	-0.006	-2.12	-0.003	-1.03
10	500	0.3	-0.005	-1.58	-0.005	-1.63	-0.005	-1.74
5	100	0.5	-0.052	-10.32	-0.060	-12.09	0.024	4.90
10	100	0.5	-0.034	-6.78	-0.040	-8.00	-0.036	-7.22
5	500	0.5	-0.011	-2.29	-0.012	-2.42	0.005	0.98
10	500	0.5	-0.008	-1.51	-0.008	-1.60	-0.007	-1.44
5	100	0.8	-0.224	-28.06	-0.302	-37.81	3.500	437.48
10	100	0.8	-0.108	-13.53	-0.152	-18.98	0.402	50.26
5	500	0.8	-0.056	-7.02	-0.060	-7.56	0.700	87.50
10	500	0.8	-0.027	-3.44	-0.030	-3.80	0.080	10.05
5	100	0.9	-0.455	-50.56	-1.068	-118.64	57.675	6408.32
10	100	0.9	-0.220	-24.47	-0.474	-52.66	8.888	987.54
5	500	0.9	-0.184	-20.48	-0.214	-23.73	11.535	1281.66
10	500	0.9	-0.078	-8.64	-0.095	-10.53	1.778	197.51

The GMM Bias is based on 10000 Monte Carlo Runs from HHK. The second Order Bias calculations are also from HHK, Table 1. We see that our calculations of the bias are poor for values of α close to 1.

Table 2: % Bias of $\hat{a}_{GMM}$, $\hat{a}_{BC2}$, and $\hat{a}_Z$

			HHK Monte Carlo 5000 Runs		Our Monte Carlo 5000 Runs			σ_η,σ known	
T	n	α	$\hat{a}_{GMM}$	$\hat{a}_{BC2}$	$\hat{a}_{GMM}$	$\hat{a}_{Z,trunc}$	$\hat{a}_{Z,med}$	$\tilde{a}_{Z,trunc}$	$\tilde{a}_{Z,med}$
5	100	0.1	-14.96	0.25	-20.48	11.13	10.16	-5.93	-7.70
10	100	0.1	-14.06	-0.77	-15.88	8.17	7.84	1.55	1.17
5	500	0.1	-3.68	-0.38	-2.72	3.46	4.10	0.33	0.94
10	500	0.1	-3.15	-0.16	-3.15	1.69	1.62	0.46	0.40
5	100	0.3	-8.86	-0.47	-10.97	-0.51	0.04	-6.82	-5.45
10	100	0.3	-7.06	-0.66	-7.78	4.03	4.28	0.61	0.89
5	500	0.3	-2.03	-0.16	-1.99	-0.17	-0.54	-1.03	-1.39
10	500	0.3	-1.58	-0.10	-1.78	0.59	0.51	-0.04	-0.12
5	100	0.5	-10.05	-1.14	-10.25	-23.54	-24.15	-23.76	-16.60
10	100	0.5	-6.76	-0.93	-6.96	3.19	3.53	-0.08	0.26
5	500	0.5	-2.25	-0.15	-2.06	-5.45	-5.61	-3.26	-3.04
10	500	0.5	-1.53	-0.11	-1.39	0.56	0.59	0.05	0.09
5	100	0.8	-27.56	-11.33	-16.11	-218.33	-157.11	-1157.99	-45.66
10	100	0.8	-13.45	-4.55	-8.53	-6.73	-6.22	-7.91	-5.83
5	500	0.8	-6.98	-0.72	-3.00	-50.91	-48.68	-34.01	-18.93
10	500	0.8	-3.48	-0.37	-1.68	-2.20	-2.18	-1.80	-1.63
5	100	0.9	-50.22	-42.10	-49.53	-3425.12	-1007.76	-3836.30	-72.01
10	100	0.9	-24.27	-15.82	-19.39	-372.45	-230.11	-155.13	-31.22
5	500	0.9	-20.50	-6.23	-13.09	-7686.80	-1298.08	-94924.19	-62.53
10	500	0.9	-8.74	-2.02	-4.70	-229.46	-188.10	-207.22	-50.82

- $\hat{a}_{Z,trunc}$ is the mean of 4800 $\hat{a}_Z$ estimates (the lowest 100 and highest 100 are thrown out).
- $\hat{a}_{Z,med}$ is computed by simply looking at the median of the 5000 $\hat{a}_Z$ values.
- $\tilde{a}_{Z,trunc}$ is computed in the same way as $\hat{a}_{Z,trunc}$, except we assume σ_η,σ are known and set to 1.
- $\tilde{a}_{Z,med}$ is computed in the same way as $\hat{a}_{Z,med}$, except we assume σ_η,σ are known and set to 1.

Table 3: Rejection Levels under Models 1-3

			Model 1 Sig. Level				Model 2 Sig. Level				Model 3 Sig. Level			
k	n	d	0.20	0.10	0.05	0.01	0.20	0.10	0.05	0.01	0.20	0.10	0.05	0.01
3	100	3	0.33	0.16	0.08	0.03	0.76	0.57	0.44	0.24	0.59	0.32	0.16	0.06
3	200	3	0.36	0.17	0.09	0.03	0.87	0.77	0.65	0.42	0.79	0.58	0.35	0.08
4	100	3	0.46	0.23	0.12	0.04	0.92	0.82	0.67	0.48	0.73	0.48	0.29	0.11
4	200	3	0.56	0.30	0.20	0.06	0.98	0.96	0.91	0.80	0.86	0.66	0.47	0.18
3	100	5	0.30	0.14	0.07	0.01	0.77	0.56	0.41	0.21	0.61	0.34	0.18	0.06
3	200	5	0.34	0.14	0.05	0.02	0.93	0.82	0.71	0.46	0.84	0.58	0.32	0.09
4	100	5	0.48	0.25	0.14	0.05	0.95	0.87	0.74	0.51	0.71	0.44	0.28	0.10
4	200	5	0.52	0.27	0.14	0.05	0.99	0.99	0.96	0.89	0.89	0.72	0.49	0.17

The rejection levels are based on 500 Monte Carlo runs. The table shows how often our test rejects the null of a linear location-scale model, under the three models. We see higher rejection levels under model 2 and model 3, as we would expect. We also see the rejection levels for all three models increasing with n, k.