Edgeworth Expansions for the Wald and GMM Statistics for Nonlinear Restrictions

Bruce E. Hansen*
University of Wisconsin†

www.ssc.wisc.edu/~bhansen

May 2000

Abstract

An Edgeworth expansion is derived for the GMM distance statistic for a real-valued nonlinear restriction on a normal linear regression. The Edgeworth expansion takes the form $F\left(x - n^{-1}\alpha_1 x\right) + o(n^{-1})$, where F is the χ^2_1 distribution. We also provide a refinement of the Edgeworth expansion for the Wald statistic derived by Park and Phillips (1988), which takes the form $F\left(x - n^{-1}\left(\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3\right)\right) + o(n^{-1})$. Our calculations show that the leading coefficient α_1 is the same in these two expansions. This establishes that, to the order of approximation of the Edgeworth expansion, the GMM distance statistic has a superior approximation to the chi-square distribution than does the Wald statistic.

We also update the Monte Carlo simulation of Gregory and Veall (1985) to include both heteroskedasticity-robust covariance matrix estimation and the GMM distance statistic. We find that if the robust covariance matrix is calculated under the null, the GMM statistic has near perfect finite sample Type I error in our experiments, even in sample sizes as small as $n = 20$.

*This research was supported by a grant from the National Science Foundation. I thank Yuichi Kitamura and Ken West for helpful comments and discussions.
†Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706

1 Introduction

The Wald test is a popular test of statistical hypotheses largely because it is simple to compute. There are many reasons, however, to believe that generically the Wald test is a poor choice as a test of nonlinear hypothesis. One reason frequently mentioned is that the Wald statistic is not invariant to the algebraic formulation of the hypothesis. Gregory and Veall (1985) and Lafontaine and White (1986) show in Monte Carlo simulations the potentially large consequences of alternative algebraic formulations. Park and Phillips (1988) formalized this finding by showing that the coefficients of the Edgeworth expansion of the Wald statistic depend on the formulation.

Separately, Newey and West (1987) proposed a distance GMM statistic for nonlinear hypotheses. In the context of linear regression, their statistic is simply the GMM criterion function evaluated at the restricted estimates. When the hypothesis is a linear restriction on the parameters, their test corresponds to the Wald statistic. When the hypothesis is nonlinear the two statistics differ. A striking feature of the GMM distance statistic is that it is invariant to the algebraic formulation of the hypothesis. (The invariance follows directly from its definition in terms of the criterion function.) The GMM distance statistic also has the advantage that it is robust to heteroskedasticity (if a heteroskedasticity-consistent covariance matrix is used to define the GMM criterion). This is in contrast to the likelihood ratio statistic, which is invariant to formulation of the hypothesis, but is not robust to heteroskedasticity. For a pedagogical description of this statistic, see section 9.2 of Newey and McFadden (1994).

Little is known, however, about the finite sample behavior of the GMM statistic. This paper attempts to fill this gap by providing an Edgeworth expansion for the GMM statistic in the leading case considered by Park and Phillips (1988). We use the explicit matrix approach to Edgeworth expansions initiated by Park and Phillips (1988), and push their approach one step further, by using explicit matrix formulae for all our expressions. The advantage of this approach is that we are able to calculate greatly simplified expressions for our Edgeworth expansions, which enable us to make direct comparisons between statistics.

We rederive the Park-Phillips Edgeworth expansion for the Wald statistic, along with that for the GMM statistic. We find the striking result that the Edgeworth expansion for the GMM statistic is a strict simplification of that for the Wald statistic. Thus the chi-square approximation for the GMM statistic is as good as that for any algebraic formulation of the Wald statistic, at least up to the level of approximation of the Edgeworth expansion.

Gregory and Veall (1985) provided dramatic simulation evidence that two alternative formulations of the same hypothesis lead to very different finite sample behavior of the Wald statistic. We update their experiment, and contrast the performance of the Wald statistics with the GMM statistic. We also compare the performance of the tests when heteroskedasticity-robust covariance matrices and GMM weight matrices are used. The simulations show that the if GMM statistic is computed with a weight matrix calculated under the alternative hypothesis, its performance is nearly identical to the Gregory-Veall "good" form of the Wald statistic, while if the GMM statistic is computed with the weight matrix calculated under the null hypothesis, the size distortion virtually disappears. The results show that even in samples as small as $n = 20$, test statistics can be made robust to unknown heteroskedasticity without any loss of control over Type I error.

The paper is organized as follows. Section 2 states the model and test statistics. Section 3 describes alternative methods to calculate the covariance matrix of the estimates and the weight matrix for GMM estimation. Section 4 contains our main results. Section 5 is a Monte Carlo simulation. A brief conclusion follows in Section 6. Appendix A is a restatement of the Park-Phillips (1988) Edgeworth expansion (for reference). Appendix B contains the proof of Theorem 1 (the Edgeworth expansion for the Wald statistic). Appendix C contains the proof of Theorem 2 (the Edgeworth expansion for the GMM statistic).

A Gauss program which calculates the GMM statistics described in this paper can be downloaded from my webpage, www.ssc.wisc.edu/˜bhansen.

2 Linear Regression with NonLinear Hypotheses

The model is a linear regression

$$\begin{aligned} y_i &= x_i'\beta + e_i \\ E(x_i e_i) &= 0, \end{aligned}$$

$i = 1, ..., n$, where x_i and β are each $k \times 1$. Let β_0 denote the true value of β.

The goal is to test the nonlinear hypothesis

$$\begin{aligned} H_0 &: g(\beta) = 0 \\ H_1 &: g(\beta) \neq 0 \end{aligned} \tag{1}$$

where $g : R^k \to R$. We are interested in testing H_0 against H_1.

Let

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

be the OLS estimator of β, and let

$$V_n = (X'X)^{-1}\Omega_n(X'X)^{-1} \tag{2}$$

be an estimator of the covariance matrix of $\hat{\beta}$, where Ω_n is an estimate of $nE\left(x_i x_i' e_i^2\right)$. We discuss specific choices below.

A common test statistic for H_0 is the Wald statistic

$$\begin{aligned} W &= n\, g(\hat{\beta})' \left(\hat{G}' V_n \hat{G}\right)^{-1} g(\hat{\beta}) \\ \hat{G} &= \frac{\partial}{\partial \beta} g(\hat{\beta}). \end{aligned}$$

The strengths of the Wald statistic are that it is easy to compute, yet is asymptotically χ_1^2 under H_0 under very general conditions. A major weakness, however, is that the statistic is not invariant to the formulation of the hypothesis g.

A less commonly applied test of H_0 is the GMM distance statistic introduced by Newey and West (1987) and discussed in Newey and McFadden (1994, section 9.2). This statistic is defined as the difference in the GMM criterion evaluated at estimates calculated under the null and alternative, and constructed with the same efficient weight matrix. For the regression model, the GMM criterion function is

$$J(\beta) = (Y - X\beta)' X\, \Omega_n^{-1} X' (Y - X\beta),$$

where Ω_n again is an estimate of $nE\left(x_i x_i' e_i^2\right)$.

The unrestricted GMM estimator minimizes $J(\beta)$ over $\beta \in R^k$:

$$\begin{aligned} \hat{\beta} &= \underset{\beta \in R^k}{\operatorname{argmin}}\ J(\beta) \\ &= (X'X)^{-1}(X'Y) \end{aligned}$$

and is identical to the OLS estimator. Note that $J(\hat{\beta}) = 0$.

The restricted GMM estimator minimizes $J(\beta)$ subject to the constraint (1):

$$\tilde{\beta} = \underset{g(\beta)=0}{\operatorname{argmin}}\ J(\beta). \tag{3}$$

When $g(\beta)$ is nonlinear a closed-form expression for $\tilde{\beta}$ does not exist. However, in general $\tilde{\beta}$ is quite simple to calculate, as the criterion $J(\beta)$ is quadratic in β. Minimizing a quadratic function subject to a nonlinear constraint is a straightforward numerical optimization problem.

The Newey-West GMM distance test statistic is the difference in the criterion function evaluated at the two estimates:

$$\begin{aligned} DM &= J(\tilde{\beta}) - J(\hat{\beta}) \\ &= \min_{g(\beta)=0} \left(Y - X\beta\right)' X\Omega_n^{-1} X' \left(Y - X\beta\right). \end{aligned} \tag{4}$$

The statistic (4) has a number of wonderful advantages over the Wald statistic. Primarily, it is invariant to the formulation of the hypothesis (1). This is because the parameter space $\{\beta : g(\beta) = 0\}$ is invariant to its algebraic formulation. The lack of invariance is a major problem with implementation of the Wald statistic when g is nonlinear. However, in the special case when g is linear, then the two statistics are numerically identical (if the same Ω_n is used).

A by-product of the computation of the test statistic (4) is the restricted estimate $\tilde{\beta}$. For reference, an estimate of the covariance matrix for $\tilde{\beta}$ can be calculated as

$$\tilde{V}_n = V_n - V_n \hat{G} \left(\hat{G}' V_n \hat{G}\right)^{-1} \hat{G}' V_n,$$

where V_n is defined in (2). (For a derivation, see section 9.1 of Newey and McFadden, 1994).

3 Choice of Variance and Weight Matrix

The statistics depend on the choice of Ω_n. The Wald statistic is typically calculated from the unrestricted estimates $\hat{\beta}$. One choice for Ω_n is the Eicker-White estimator:

$$\begin{aligned} \hat{\Omega}_n &= \sum_{i=1}^{n} x_i x_i' \hat{e}_i^2 \\ \hat{e}_i &= y_i - x_i' \hat{\beta}, \end{aligned} \tag{5}$$

as this is asymptotically valid for the specified model without additional auxiliary assumptions. An alternative choice is the OLS estimator

$$\begin{aligned} \hat{\Omega}_n^0 &= X' X \hat{\sigma}^2 \\ \hat{\sigma}^2 &= \frac{1}{n-k} \sum_{i=1}^{n} \hat{e}_i^2, \end{aligned} \tag{6}$$

which is valid under the conditional homoskedasticity assumption $E\left(e_i^2 \mid x_i\right) = \sigma^2$.

The GMM statistic (4) also may be computed setting Ω_n to equal either $\hat{\Omega}_n$ or $\hat{\Omega}_n^0$, the latter valid only under the assumption of homoskedasticity. These choices correspond to computing the weight matrix under the alternative hypothesis, since they are computed from the unrestricted estimates. Another choice is to compute the weight matrix from estimates obtained under the null hypothesis. This requires iterated GMM. The first step sets Ω_n to equal (5) or (6) and calculates the first-step estimator $\tilde{\beta}$ as in (3). In the second step we calculate

$$\begin{aligned}\tilde{\Omega}_n &= \sum_{i=1}^{n} x_i x_i' \tilde{e}_i^2 \\ \tilde{e}_i &= y_i - x_i'\tilde{\beta},\end{aligned}$$

for the general case, or

$$\begin{aligned}\tilde{\Omega}_n^0 &= X'X\tilde{\sigma}^2, \\ \tilde{\sigma}^2 &= \frac{1}{n-k+1}\sum_{i=1}^{n} \tilde{e}_i^2\end{aligned}$$

under the homoskedasticity assumption. Then setting $\Omega_n = \tilde{\Omega}_n$ or $\Omega_n = \tilde{\Omega}_n^0$, (3) and (4) are re-computed as a second-step minimization.

Newey and West (1987) and Newey and McFadden (1994) do not provide any guidance to whether the weight matrix should be computed under the null ($\tilde{\Omega}_n$) or alternative ($\hat{\Omega}_n$). Since $\tilde{\Omega}_n$ is computed from the restricted estimates, we would expect it to be a more efficient estimator under the null hypothesis, and thus provide better finite sample Type I error approximations, at the cost of a somewhat greater computational burden and an uncertain effect upon the power of the test.

4 Edgeworth Expansions

Park and Phillips (1988) used an Edgeworth expansion to show that the non-invariance of the Wald statistic to the formulation of (1) is responsible for the poor size properties of the Wald statistic. Our goal in this section is use the same Edgeworth expansion argument to show that the GMM statistic has a superior Edgeworth approximation to the chi-square distribution than the Wald statistic, and thus should be expected to have better size properties.

Following Park and Phillips (1988), we derive our expansions under the assumptions that $e \mid X \sim N(0, I_n)$ and $X'X = nI_k$, and that this knowledge has been used to simplify the statistics, so that $\Omega_n = nI_n$. While this assumption is not relevant for applications, it places the focus on the nonlinearity. Under these conditions, if g were linear then both W and

DM would have exact χ_1^2 distributions, so the divergence from the χ_1^2 is due only to the nonlinearity of g.

Assuming that $g(\beta)$ is three-times continuously differentiable, define

$$\underset{k\times 1}{G(\beta)} = \frac{\partial}{\partial\beta} g(\beta),$$

$$\underset{k\times k}{D(\beta)} = \frac{\partial^2}{\partial\beta\partial\beta'} g(\beta),$$

$$\underset{k\times k^2}{C(\beta)} = \frac{\partial}{\partial\beta} \left((\operatorname{vec} D(\beta))' \right),$$

where $\operatorname{vec}(A)$ stacks the columns of the matrix A. Let $G = G(\beta_0)$, $D = D(\beta_0)$, $C = C(\beta_0)$.

Define the projection matrices

$$\begin{aligned} P &= G(G'G)^{-1}G' \\ \overline{P} &= I - P. \end{aligned}$$

Note that these are defined if $G'G > 0$, which holds when $\operatorname{rank}(G) = 1$, which is a standard condition for hypothesis testing.

Let F_W denote the cumulative distribution function (CDF) of W, let F_{DM} denote that of DM, and let F denote the CDF of the χ_1^2 distribution.

Theorem 1 *The asymptotic expansion of W as $n \to \infty$ is given by*

$$F_W(x) = F\left(x - n^{-1}\left(G'G\right)^{-1}\left(\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3\right)\right) + o(n^{-1}) \tag{7}$$

where

$$\alpha_1 = -\frac{1}{2}\operatorname{tr}\left(\overline{P}D\overline{P}D\right) + \frac{1}{4}\left(\operatorname{tr}\left(\overline{P}D\right)\right)^2,$$

$$\alpha_2 = \frac{3}{2}\left(\operatorname{tr}(PD)\right)^2 - \operatorname{tr}(PDD) - \frac{1}{2}\operatorname{tr}(D)\operatorname{tr}(PD) - \frac{2}{3}\operatorname{tr}(PC\otimes G),$$

and

$$\alpha_3 = \frac{1}{4}\left(\operatorname{tr}(PD)\right)^2.$$

Theorem 2 *The asymptotic expansion of DM as $n \to \infty$ is given by*

$$F_{DM}(x) = F\left(x - n^{-1}\left(G'G\right)^{-1}\alpha_1 x\right) + o(n^{-1}), \tag{8}$$

where α_1 is defined in Theorem 1.

The Edgeworth expansion (7) for W was derived by Park and Phillips (1988). The main difference is that our expression (7) provides a much more compact set of expressions for the coefficients α_1, α_2, α_3, which allows a direct comparison with the expansion for the GMM statistic. The Edgeworth expansion (8) for DM appears to be new.

There are several striking implications of Theorems 1 and 2.

First, the expansion for the GMM statistic is a strict simplification of that for the Wald statistic. The Wald statistic is approximately chi-square after a cubic transformation. The GMM statistic is approximately chi-square after a linear transformation, and the linear term is identical to that for the Wald statistic. Thus, up to order $o(n^{-1})$, the expansion for the GMM statistic is less distorted from the chi-square than is that for the Wald statistic.

Second, the expansion (8) shows that the CDF of $\left(1 - n^{-1}\left(G'G\right)^{-1}\alpha_1\right)^{-1} DM$ is $F(x) + o(n^{-1})$, so only a scale adjustment is necessary to achieve an $o(n^{-1})$ approximation to the chi-square distribution. This is a necessary condition for a statistic to be Bartlett correctable.

Third, since DM is invariant to the formulation of (1), so is its distribution F_{DM}, and hence so is its Edgeworth expansion. It follows that the coefficient α_1 is invariant to the formulation of (1). This is also the leading term in the Edgeworth expansion for W. It follows that the Wald statistic's non-invariance to the formulation (1) appears in the Edgeworth expansion (7) only through the higher-order coefficients α_2 and α_3. This generalizes the finding of Park and Phillips (1988) who found that α_1 was invariant to the formulation (1) in their examples. Indeed, the invariance of α_1 to the formulation of (1) is generally true.

5 Gregory-Veall Example

We illustrate the size performance of the GMM distance test in a replication of the Gregory-Veall (1985) experiment. The model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$$

with $\beta_1\beta_2 = 1$ and $E\left(e_i \mid x_i\right) = 0$. In our experiments, we generate x_{1i}, x_{2i} and e_i as mutually independent, iid, $N(0,1)$ variables. We consider two formulations of the Wald statistic, based on the hypotheses

$$H_0^A : \beta_1 - \frac{1}{\beta_2} = 0$$

and

$$H_0^B : \beta_1\beta_2 - 1 = 0.$$

Let W^A and W^B denote the Wald statistics corresponding to these two formulations of the null hypothesis. While Gregory-Veall only examined the behavior of the Wald statistic constructed with a conventional covariance matrix estimate, we also consider the performance of the Wald and GMM statistics constructed with Eicker-White covariance matrix estimates.

As shown by Park and Phillips (1988), the expansion of the W^A statistic has coefficients α_2 and α_3 which are very large, especially when β_2 is small, yet the expansion of the W^B statistic has coefficients α_2 and α_3 which are quite small, predicting that the W^A statistic will have larger size distortions than the W^B statistic.

We also consider the GMM statistic, which is invariant to the formulation H_0^A and H_0^B. Let DM^{alt} denote this statistic if the weight matrix is calculated using the unrestricted estimates (the alternative hypothesis), and let DM^{null} denote the statistic if the weight matrix is calculated using the restricted estimates (the null hypothesis).

Table 1
Percentage Rejections at the 5% Asymptotic Level
Tests Constructed Using Homoskedastic Covariance Matrix

Case	Test	$n=20$	$n=30$	$n=50$	$n=100$	$n=500$
$\beta_1 = 10,\ \beta_2 = 0.1$	W^A	.372	.317	.257	.189	.105
	W^B	.066	.059	.055	.052	.051
	DM^{alt}	.066	.059	.056	.052	.051
	DM^{null}	.039	.042	.046	.048	.050
$\beta_1 = 5,\ \beta_2 = 0.2$	W^A	.222	.183	.145	.115	.069
	W^B	.065	.061	.055	.053	.049
	DM^{alt}	.065	.061	.055	.053	.050
	DM^{null}	.038	.044	.046	.049	.049
$\beta_1 = 2,\ \beta_2 = 0.5$	W^A	.091	.082	.071	.059	.049
	W^B	.065	.058	.055	.052	.052
	DM^{alt}	.067	.059	.056	.053	.052
	DM^{null}	.040	.043	.046	.048	.051
$\beta_1 = 1,\ \beta_2 = 1$	W^A	.047	.043	.045	.046	.049
	W^B	.078	.069	.062	.055	.051
	DM^{alt}	.065	.060	.056	.052	.050
	DM^{null}	.039	.043	.046	.047	.049

Table 2
Percentage Rejections at the 5% Asymptotic Level
Tests Constructed Using Eicker-White Covariance Matrix

Case	Test	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 500$
$\beta_1 = 10,\ \beta_2 = 0.1$	W^A	.410	.342	.270	.198	.107
	W^B	.024	.097	.078	.064	.052
	DM^{alt}	.125	.097	.078	.064	.052
	DM^{null}	.051	.050	.051	.051	.050
$\beta_1 = 5,\ \beta_2 = 0.2$	W^A	.258	.204	.158	.121	.073
	W^B	.122	.095	.077	.064	.053
	DM^{alt}	.123	.096	.078	.064	.053
	DM^{null}	.049	.050	.050	.051	.051
$\beta_1 = 2,\ \beta_2 = 0.5$	W^A	.124	.104	.084	.064	.051
	W^B	.123	.096	.079	.062	.052
	DM^{alt}	.124	.098	.079	.063	.052
	DM^{null}	.051	.050	.051	.049	.050
$\beta_1 = 1,\ \beta_2 = 1$	W^A	.094	.077	.065	.058	.051
	W^B	.133	.104	.083	.067	.052
	DM^{alt}	.123	.096	.077	.064	.052
	DM^{null}	.049	.049	.049	.050	.049

We calculate the finite sample size (Type I error) of asymptotic 5% tests, using a selection of parameter values and sample sizes from $n = 20$ to $n = 500$, from 100,000 Monte Carlo replications[1]. The results are presented in Tables 1 and 2. As predicted by our theory, the W^A statistic has substantial size distortion when β_2 is small even if the sample size is quite large, regardless of the method to compute the covariance matrix. The size distortions of the W^B and DM^{alt} statistics are quite similar, and quite modest in comparison to the W^A statistic. In addition, the size distortions of W^B and DM^{alt} are insensitive to the true value of the parameters. If the homoskedastic covariance matrix estimate is used, these tests have minimal size distortion (as the true error is indeed homoskedastic) but have moderate size distortion if the heteroskedasticity-robust covariance matrix estimate is used.

The performance of the DM^{null} statistic is stunning. Regardless of the parameterization, sample size, or covariance matrix estimation method, the Type I error is excellent. If the heteroskedasticity-robust covariance matrix estimator is used, the estimated Type I error ranges from 4.9% to 5.1%, which is not statistically different from the nominal 5.0% level. Thus the robust DM^{null} statistic has dramatically better size performance than the robust W^B statistic or the robust DM^{alt} statistic.

[1] The standard error for the estimated rejection frequencies is about .0007.

6 Conclusion

We have extended the explicit matrix approach to Edgeworth expansions developed by Park and Phillips (1988), have extended their Edgeworth expansion for the Wald statistic, and have developed a new Edgeworth expansion for the GMM statistic. The major limitation of our results is that they are calculated for the restrictive setting of a normal regression with known error variance. Variance estimation would dramatically complicate the expansions. It would be quite desirable to relax this restriction in future work.

Our simulation reports near-perfect performance of the statistic DM^{null}. A theoretical explanation of this finding would be an important avenue for future research.

7 Appendix A: The Park-Phillips Expansion

For coherence, we repeat below a summary of the Edgeworth expansion of Park and Phillips (1988). Let K_{12} be the commutation matrix such that $K_{12}\operatorname{vec}A = \operatorname{vec}(A')$ if A is $k \times k^2$, K_{21} be the commutation matrix such that $K_{21}\operatorname{vec}A = \operatorname{vec}(A')$ if A is $k^2 \times k$, and $H = I + K_{12} + K_{21}$.

The following result is a restatement of Theorem 2.4 of Park and Phillips (1988) for the case $r = 1$ (in their notation). We use the result $(\operatorname{vec}P)(\operatorname{vec}P)' = P \otimes P$, and a few other minor algebraic simplifications.

Theorem 3 *For a statistic S which has the asymptotic expansion*

$$S = \frac{(G'm)^2 + n^{-1/2}u(m) + n^{-1}v(m)}{G'G} + O_p(n^{-3/2}),$$

where

$$\begin{aligned} u(m) &= J'(m \otimes m \otimes m), \\ v(m) &= \operatorname{tr}\left[L\left(mm' \otimes mm'\right)\right] \end{aligned}$$

for some $k^3 \times 1$ vector J and $k^2 \times k^2$ matrix L, then the asymptotic expansion of the distribution function $F_S(x)$ of S is given by

$$F_S(x) = F\left(x - n^{-1}(G'G)^{-1}\left(\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3\right)\right) + o(n^{-1})$$

where

$$\alpha_0 = \frac{1}{4}\left(4a_0 - b_1\right), \tag{9}$$
$$\alpha_1 = \frac{1}{4}\left(4a_1 + b_1 - b_2\right), \tag{10}$$
$$\alpha_2 = \frac{1}{12}\left(4a_2 + b_2 - b_3\right), \tag{11}$$
$$\alpha_3 = \frac{1}{60}b_3, \tag{12}$$

and

$$a_0 = \operatorname{tr}\left\{\left(\overline{P}\otimes\overline{P}\right)L\left(I+K\right)\right\} + \left(\operatorname{vec}\overline{P}\right)' L\left(\operatorname{vec}\overline{P}\right), \tag{13}$$

$$\begin{aligned} a_1 = {} & \operatorname{tr}\left\{\left(\left(P\otimes\overline{P}\right)+\left(\overline{P}\otimes P\right)\right)L\left(I+K\right)\right\} \\ & + \left(\operatorname{vec}P\right)' L\left(\operatorname{vec}\overline{P}\right) + \left(\operatorname{vec}\overline{P}\right)' L\left(\operatorname{vec}P\right) \end{aligned} \tag{14}$$

$$a_2 = 2\operatorname{tr}\left\{\left(P\otimes P\right)L\right\} + \operatorname{tr}\left\{\left(P\otimes P\right)LK\right\}, \tag{15}$$

$$\begin{aligned} b_1 = {} & \left(G'G\right)^{-1} J' \Big\{H\left(P\otimes\overline{P}\otimes\overline{P}\right)H + H\left(P\otimes\left(\operatorname{vec}\overline{P}\right)\left(\operatorname{vec}\overline{P}\right)'\right)H \\ & +2H\left(\overline{P}\otimes\left(\operatorname{vec}P\right)\left(\operatorname{vec}\overline{P}\right)'\right)H + \left(P\otimes K\left(\overline{P}\otimes\overline{P}\right)\right) \\ & +2\left(\overline{P}\otimes K\left(P\otimes\overline{P}\right)\right) + 2\left(K\left(P\otimes\overline{P}\right)\otimes\overline{P}\right) \\ & +\left(K\left(\overline{P}\otimes\overline{P}\right)\otimes P\right) + K_{12}\left(P\otimes K\left(\overline{P}\otimes\overline{P}\right)\right)K_{21} \\ & + 2K_{12}\left(\overline{P}\otimes K\left(P\otimes\overline{P}\right)\right)K_{21}\Big\} J, \end{aligned} \tag{16}$$

$$\begin{aligned} b_2 = {} & \left(G'G\right)^{-1} J' \left\{2H\left(\overline{P}\otimes P\otimes P\right)H\right. \\ & +2H\left(P\otimes\left(\operatorname{vec}\overline{P}\right)\left(\operatorname{vec}P\right)'\right)H + \left(\overline{P}\otimes K\left(P\otimes P\right)\right) \\ & +2\left(P\otimes K\left(\overline{P}\otimes P\right)\right) + 2\left(K\left(\overline{P}\otimes P\right)\otimes P\right) \\ & +\left(K\left(P\otimes P\right)\otimes\overline{P}\right) + K_{12}\left(\overline{P}\otimes K\left(P\otimes P\right)\right)K_{21} \\ & \left. + 2K_{12}\left(P\otimes K\left(\overline{P}\otimes P\right)\right)K_{21}\right\} J, \end{aligned} \tag{17}$$

$$\begin{aligned} b_3 = {} & \left(G'G\right)^{-1} J' \left\{H\left(P\otimes P\otimes P\right) + H\left(P\otimes P\otimes P\right)H + \left(P\otimes K\left(P\otimes P\right)\right)\right. \\ & \left. +\left(K\left(P\otimes P\right)\otimes P\right) + K_{12}\left(P\otimes K\left(P\otimes P\right)\right)K_{21}\right\} J. \end{aligned} \tag{18}$$

8 Appendix B: Proof of Theorem 1

8.1 A Vector-Valued 3rd-Order Taylor Expansion

Our Edgeworth expansions will involve third-order Taylor series expansions. To facilitate our explicit matrix formulation, the following algebraic development will be helpful.

Lemma 1 *If $g : R^k \to R$ is three times continuously differentiable, then*

$$g(\beta_0 + \delta) = g(\beta_0) + G'\delta + \frac{1}{2}\delta' D\delta + \frac{1}{6}\delta' C\left(\delta \otimes \delta\right) + O\left(|\delta|^4\right).$$

where G, D and C are defined in section 3.

Proof: Note that for any δ,

$$\delta' \frac{\partial}{\partial \beta_j} D(\beta_0)\delta = \frac{\partial}{\partial \beta_j}\left(\operatorname{vec} D(\beta_0)\right)'\left(\delta \otimes \delta\right),$$

so

$$\sum_{j=1}^{k} \delta_j \frac{\partial}{\partial \beta_j}\left(\operatorname{vec} D(\beta_0)\right)'\left(\delta \otimes \delta\right) = \delta' C\left(\delta \otimes \delta\right).$$

Thus a third-order Taylor expansion of $g(\beta_0 + \delta)$ about $\delta = 0$ yields

$$\begin{aligned} g(\beta_0 + \delta) &= g(\beta_0) + G'\delta + \frac{1}{2}\delta' D\delta + \frac{1}{6}\delta' \sum_{j=1}^{k} \frac{\partial}{\partial \beta_j} D(\beta_0)\delta\delta_j + O\left(|\delta|^4\right) \\ &= g(\beta_0) + G'\delta + \frac{1}{2}\delta' D\delta + \frac{1}{6}\delta' C\left(\delta \otimes \delta\right) + O\left(|\delta|^4\right), \end{aligned}$$

as stated. ■

Let K be the commutation matrix such that for any $k \times k$ matrix A, $K \operatorname{vec} A = \operatorname{vec}(A')$. Note that $K' = K$. The matrix D is symmetric, so that $\operatorname{vec} D = \operatorname{vec}(D')$ and

$$K \operatorname{vec} D = \operatorname{vec} D. \tag{19}$$

A similar set of properties hold for C. The following facts are useful.

Lemma 2

1. $CK = C$;

2. $\operatorname{vec} C = \operatorname{vec}(C')$.

3. For any $k \times k$ *matrix* B *and* $k \times 1$ *vector* a,

$$\operatorname{tr}[K(BC \otimes a)] = \operatorname{tr}[a \otimes BC] = \operatorname{tr}[BC \otimes a] = \operatorname{tr}[BC(I_k \otimes a)].$$

Proof: Let

$$C_j = \frac{\partial}{\partial \beta_j} D(\beta_0)$$

and

$$\begin{aligned} c_j &= \operatorname{vec} C_j \\ &= \frac{\partial}{\partial \beta_j} \operatorname{vec} D(\beta_0). \end{aligned}$$

Note that c_j is the $j'th$ column of C'. Since $K \operatorname{vec} D(\beta) = \operatorname{vec} D(\beta)$, it follows that

$$K c_j = \frac{\partial}{\partial \beta_j} K \operatorname{vec} D(\beta_0) = \frac{\partial}{\partial \beta_j} \operatorname{vec} D(\beta_0) = c_j.$$

Hence $KC' = C'$ and $CK = C$, establishing part 1.

Also, we see that

$$\begin{aligned} C_j &= \frac{\partial}{\partial \beta} \frac{\partial^2}{\partial \beta_j \partial \beta'} g(\beta_0) \\ &= \frac{\partial}{\partial \beta} d_j(\beta_0)', \end{aligned}$$

where $d_j(\beta)$ is the $j'th$ column of $D(\beta)$. Hence

$$C = \begin{bmatrix} C_1 & C_2 & \cdots & C_k \end{bmatrix}.$$

Thus

$$\operatorname{vec} C = \begin{pmatrix} \operatorname{vec} C_1 \\ \operatorname{vec} C_2 \\ \vdots \\ \operatorname{vec} C_k \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}.$$

But since c_j is the $j'th$ column of C',

$$\operatorname{vec}(C') = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}.$$

so we conclude that $\operatorname{vec} C = \operatorname{vec}(C')$, establishing part 2.

For part (3), the equality $\operatorname{tr}[K(BC \otimes a)] = \operatorname{tr}[a \otimes BC]$ is Magnus and Neudecker (1988, Chapter 3, Theorem 9). Then since $C = CK$,

$$\operatorname{tr}(a \otimes BC) = \operatorname{tr}(a \otimes BCK) = \operatorname{tr}((a \otimes BC)K) = \operatorname{tr}(K(a \otimes BC)) = \operatorname{tr}(BC \otimes a), \quad (20)$$

establishing part (3).

For part (4),

$$\operatorname{tr}[BC(I_k \otimes a)] = \operatorname{tr}[(I_k \otimes a)BC] = \operatorname{tr}[(I_k \otimes a)(BC \otimes 1)] = \operatorname{tr}[BC \otimes a].$$

■

8.2 Expansion for Wald Statistic

Let

$$\psi(\beta) = (G(\beta)'G(\beta))^{-1}$$

so that $\left(\hat{G}'\hat{G}\right)^{-1} = \psi(\hat{\beta})$.

First, note that

$$\frac{\partial}{\partial\beta}\psi(\beta) = -2\left(G(\beta)'G(\beta)\right)^{-2} D(\beta)G(\beta).$$

Second, as

$$\begin{aligned}\left(\frac{\partial}{\partial\beta'}D(\beta_0)G\right)' &= \frac{\partial}{\partial\beta}G'D(\beta_0) \\ &= \frac{\partial}{\partial\beta}\operatorname{vec}(G'D)' \\ &= \frac{\partial}{\partial\beta}\operatorname{vec}(D)'(I \otimes G) \\ &= C(I \otimes G),\end{aligned}$$

thus

$$\begin{aligned}\frac{\partial^2}{\partial\beta\partial\beta'}\psi(\beta_0) &= -2(G'G)^{-2}DD + 8(G'G)^{-3}DGG'D - 2(G'G)^{-2}C(I \otimes G) \\ &= 2(G'G)^{-2}\left[-DD + 4DPD - C(I \otimes G)\right].\end{aligned}$$

Hence

$$\begin{aligned}\psi(\hat{\beta}) &= \frac{1}{G'G}\big(1 - n^{-1/2}2(G'G)^{-1}G'Dm \\ &\quad + n^{-1}(G'G)^{-1}m'\left[-DD + 4DPD - C(I \otimes G)\right]m\big) + O_p(n^{-3/2}).\end{aligned} \quad (21)$$

Under the assumptions, we have

$$\begin{aligned} W &= n\, g(\hat{\beta})' \left(\hat{G}'\hat{G}\right)^{-1} g(\hat{\beta}) \\ &= \left(\sqrt{n} g(\hat{\beta})\right)^2 \psi(\hat{\beta}). \end{aligned}$$

Let

$$m = \sqrt{n}\left(\hat{\beta} - \beta\right) = n^{-1/2} X'e \sim N\left(0, I\right).$$

By Lemma 1,

$$\sqrt{n} g(\hat{\beta}) = G'm + n^{-1/2} m'Dm + n^{-1}\frac{1}{6} m'C\left(m \otimes m\right) + O_p(n^{-3/2}).$$

Hence

$$\begin{aligned} \left(\sqrt{n} g(\hat{\beta})\right)^2 &= \left(G'm\right)^2 + n^{-1/2} m'Gm'Dm \\ &\quad + n^{-1}\left[\frac{1}{4}\left(m'Dm\right)^2 + \frac{1}{3} m'Gm'C\left(m \otimes m\right)\right] + O_p(n^{-3/2}). \end{aligned} \tag{22}$$

Putting (22) and (21) together, we obtain the asymptotic expansion

$$W = \frac{\left(G'm\right)^2 + n^{-1/2} u + n^{-1} v}{G'G} + O_p(n^{-3/2}),$$

where

$$u = m'Gm'Dm - 2m'PmG'Dm$$

and

$$\begin{aligned} v = (&\tfrac{1}{4}\left(m'Dm\right)^2 + \tfrac{1}{3} m'Gm'C\left(m \otimes m\right) - 2m'Dmm'PDm \\ &+ m'Pmm'\left[-DD + 4DPD - C\left(I \otimes G\right)\right] m. \end{aligned}$$

Since

$$m'am'Dm = \left(a' \otimes \left(\mathrm{vec}\, A\right)'\right)\left(m \otimes m \otimes m\right),$$

we can write $u = J'\left(m \otimes m \otimes m\right)$, where

$$\begin{aligned} J &= \left(G \otimes \mathrm{vec}\, D\right) - 2\left(G \otimes G \otimes DG\right)\left(G'G\right)^{-1} \\ &= G \otimes \left(\mathrm{vec}\, D - 2\,\mathrm{vec}\left(DP\right)\right) \\ &= G \otimes \mathrm{vec}\left(D\left(\overline{P} - P\right)\right) \\ &= \left(G \otimes \left(\overline{P} - P\right) \otimes I\right) \mathrm{vec}\, D. \end{aligned}$$

Similarly, since

$$\begin{aligned} m'Am\; m'Bm &= (m\otimes m)'(\mathrm{vec}\,A)'(\mathrm{vec}\,B)'(m\otimes m) \\ &= \mathrm{tr}\left[\left((\mathrm{vec}\,A)(\mathrm{vec}\,B)'\right)(m\otimes m)(m\otimes m)'\right] \\ &= \mathrm{tr}\left[\left((\mathrm{vec}\,A)(\mathrm{vec}\,B)'\right)(mm'\otimes mm')\right], \end{aligned}$$

and by Magnus and Neudecker (1988, Chapter 2), Theorem 3,

$$\begin{aligned} m'am'A(m\otimes m) &= (m\otimes m)'(A'\otimes a)\,\mathrm{vec}\,(mm') \\ &= \mathrm{tr}\left[(A'\otimes a)(mm'\otimes mm')\right] \end{aligned}$$

then $v = \mathrm{tr}\left[L\,(mm'\otimes mm')\right]$, where

$$\begin{aligned} L &= \frac{1}{4}(\mathrm{vec}\,D)(\mathrm{vec}\,D)' - 2(\mathrm{vec}\,PD)(\mathrm{vec}\,D)' - (\mathrm{vec}\,P)(\mathrm{vec}\,DD)' \\ &\quad +4(\mathrm{vec}\,P)(\mathrm{vec}\,DPD)' - (\mathrm{vec}\,P)(\mathrm{vec}\,(C\,(I\otimes G)))' + \frac{1}{3}(C\otimes G). \end{aligned}$$

This expansion is equivalent to equation (7) of Park and Phillips (1988), but is in a different algebraic form. The above expression turns out to be more convenient for evaluation of the coefficients of the Edgeworth expansion. Using Theorem 3 in Appendix A, the coefficients of the expansion (7) are found by explicit calculation of the coefficients $a_0, a_1, a_2, b_1, b_2, b_3$ from the expressions for J and L.

8.3 Calculation of a_0

First, since $\left(\overline{P}\otimes\overline{P}\right)\mathrm{vec}\,P = \mathrm{vec}\,\overline{P}P\overline{P} = 0$ and $\left(\overline{P}\otimes\overline{P}\right)\mathrm{vec}\,PD = \mathrm{vec}\,\overline{P}PD\overline{P} = 0$, then

$$\left(\overline{P}\otimes\overline{P}\right)L = \frac{1}{4}\left(\mathrm{vec}\left(\overline{P}D\overline{P}\right)\right)(\mathrm{vec}\,D)'.$$

Using (19) and the fact that $\mathrm{vec}\,A'\,\mathrm{vec}\,B = \mathrm{tr}\,(A'B)$,

$$\begin{aligned} \mathrm{tr}\left\{\left(\overline{P}\otimes\overline{P}\right)L\,(I+K)\right\} &= \frac{1}{4}\,\mathrm{tr}\left\{\left(\mathrm{vec}\left(\overline{P}D\overline{P}\right)\right)(\mathrm{vec}\,D)'(I+K)\right\} \\ &= \frac{1}{2}\,\mathrm{tr}\left\{\left(\mathrm{vec}\left(\overline{P}D\overline{P}\right)\right)(\mathrm{vec}\,D)'\right\} \\ &= \frac{1}{2}\,\mathrm{tr}\left(D\overline{P}D\overline{P}\right). \end{aligned}$$

Second, as $\left(\mathrm{vec}\,\overline{P}\right)'(\mathrm{vec}\,P)' = \mathrm{tr}\left(\overline{P}P\right) = 0$, and $\left(\mathrm{vec}\,\overline{P}\right)'(\mathrm{vec}\,PD) = \mathrm{tr}\left(\overline{P}PD\right) = 0$, and

$$0 = \mathrm{vec}\left(G'\overline{P}C\right) = (C'\otimes G')\,\mathrm{vec}\left(\overline{P}\right),$$

then

$$\begin{aligned}\left(\mathrm{vec}\,\overline{P}\right)' L \left(\mathrm{vec}\,\overline{P}\right) &= \frac{1}{4}\left(\mathrm{vec}\,\overline{P}\right)' (\mathrm{vec}\,D)\,(\mathrm{vec}\,D)' \left(\mathrm{vec}\,\overline{P}\right)' + \frac{1}{3}\left(\mathrm{vec}\,\overline{P}\right)' C\,(I \otimes G) \left(\mathrm{vec}\,\overline{P}\right) \\ &= \frac{1}{4}\left(\mathrm{tr}\left(D\overline{P}\right)\right)^2.\end{aligned}$$

Summing these terms, we see that (13) equals

$$a_0 = \frac{1}{2}\,\mathrm{tr}\left(D\overline{P}D\overline{P}\right) + \frac{1}{4}\left(\mathrm{tr}\left(D\overline{P}\right)\right)^2. \tag{23}$$

8.4 Calculation of a_1

First,

$$\mathrm{tr}\left[\left(P \otimes \overline{P}\right) L\,(I+K)\right] = \frac{1}{4}\,\mathrm{tr}\left[\left(\mathrm{vec}\,\overline{P}DP\right)(\mathrm{vec}\,D)'\,(I+K)\right] = \frac{1}{2}\,\mathrm{tr}\left(D\overline{P}DP\right).$$

Second, using Lemma 2 part 3,

$$\begin{aligned}\mathrm{tr}\left[\left(\overline{P} \otimes P\right) L\,(I+K)\right] &= \mathrm{tr}\left\{\left[-\frac{7}{4}\left(\mathrm{vec}\,PD\overline{P}\right)(\mathrm{vec}\,D)' + \frac{1}{3}\left(\overline{P}C \otimes G\right)\right](I+K)\right\} \\ &= -\frac{7}{2}\,\mathrm{tr}\left(D\overline{P}DP\right) + \frac{2}{3}\,\mathrm{tr}\left(\overline{P}C \otimes G\right).\end{aligned}$$

Third, using the fact that $\mathrm{tr}\,(P) = 1$, and Lemma 2 part 4,

$$\begin{aligned}(\mathrm{vec}\,P)' L \left(\mathrm{vec}\,\overline{P}\right) &= \frac{1}{4}\,\mathrm{tr}\,(DP)\,\mathrm{tr}\left(D\overline{P}\right) - 2\,\mathrm{tr}\,(DP)\,\mathrm{tr}\left(D\overline{P}\right) - \mathrm{tr}\,(P)\,\mathrm{tr}\left(DD\overline{P}\right) \\ &\quad +4\,\mathrm{tr}\,(P)\,\mathrm{tr}\left(DPD\overline{P}\right) - \mathrm{tr}\,(P)\,\mathrm{tr}\left(\overline{P}C\,(I \otimes G)\right) \\ &\quad +\frac{1}{3}\,(\mathrm{vec}\,P)'\,(C \otimes G)\left(\mathrm{vec}\,\overline{P}\right) \\ &= -\frac{7}{4}\,\mathrm{tr}\,(DP)\,\mathrm{tr}\left(D\overline{P}\right) - \mathrm{tr}\left(DD\overline{P}\right) + 4\,\mathrm{tr}\left(DPD\overline{P}\right) - \frac{2}{3}\,\mathrm{tr}\left(\overline{P}C \otimes G\right).\end{aligned}$$

The final equality uses Lemma 2 part 3 and the fact that $K\,\mathrm{vec}\,P = \mathrm{vec}\,P$, $\mathrm{vec}\,(C') = \mathrm{vec}\,C$, Magnus and Neudecker (1988, Theorem 3, Chapter 2), and Lemma 2 part 3 imply that

$$\begin{aligned}(\mathrm{vec}\,P)'\,(C \otimes G)\left(\mathrm{vec}\,\overline{P}\right) &= (\mathrm{vec}\,P)'\,K\,(C \otimes G)\left(\mathrm{vec}\,\overline{P}\right) \\ &= (\mathrm{vec}\,P)'\,(G \otimes C)\left(\mathrm{vec}\,\overline{P}\right) \\ &= (\mathrm{vec}\,(C'PG))'\left(\mathrm{vec}\,\overline{P}\right) \\ &= (\mathrm{vec}\,(C'))'\,(G \otimes I \otimes I)\left(\mathrm{vec}\,\overline{P}\right) \\ &= (\mathrm{vec}\,C)'\,((G \otimes I) \otimes I)\left(\mathrm{vec}\,\overline{P}\right)\end{aligned}$$

$$
\begin{aligned}
&= \operatorname{tr}\left[\overline{P}C\,(G \otimes I)\right] \\
&= \operatorname{tr}\left[\left(1 \otimes \overline{P}C\right)(G \otimes I)\right] \\
&= \operatorname{tr}\left(G \otimes \overline{P}C\right) \\
&= \operatorname{tr}\left(\overline{P}C \otimes G\right).
\end{aligned}
$$

Fourth,

$$
\begin{aligned}
\left(\operatorname{vec}\overline{P}\right)' L\,(\operatorname{vec}P) &= \frac{1}{4}\operatorname{tr}\left(D\overline{P}\right)\operatorname{tr}(DP) + \frac{1}{3}\left(\operatorname{vec}\overline{P}\right)'(C \otimes G)(\operatorname{vec}P) \\
&= \frac{1}{4}\operatorname{tr}\left(D\overline{P}\right)\operatorname{tr}(DP).
\end{aligned}
$$

Summing these three terms, we find that (14) equals

$$
\begin{aligned}
a_1 &= -\frac{3}{2}\operatorname{tr}(DP)\operatorname{tr}\left(D\overline{P}\right) + \operatorname{tr}\left(DPD\overline{P}\right) - \operatorname{tr}\left(DD\overline{P}\right) \\
&= -\frac{3}{2}\operatorname{tr}(DP)\operatorname{tr}\left(D\overline{P}\right) - \operatorname{tr}\left(D\overline{P}D\overline{P}\right). \qquad (24)
\end{aligned}
$$

8.5 Calculation of a_2

Observe that

$$
\begin{aligned}
(P \otimes P)\,L &= \frac{1}{4}(\operatorname{vec}PDP)(\operatorname{vec}D)' - 2(\operatorname{vec}PDP)(\operatorname{vec}D)' - (\operatorname{vec}P)(\operatorname{vec}DD)' \\
&\quad + 4(\operatorname{vec}P)(\operatorname{vec}DPD)' - (\operatorname{vec}P)(\operatorname{vec}(C(I \otimes G)))' + \frac{1}{3}(PC \otimes G).
\end{aligned}
$$

Hence using Lemma 2 part 3, (15) equals

$$
\begin{aligned}
a_2 &= 3\left\{-\frac{7}{4}\operatorname{tr}(DPDP) - \operatorname{tr}(DDP) + 4\operatorname{tr}(DPDP) - \operatorname{tr}(PC \otimes G)\right\} \\
&\quad + \frac{2}{3}\operatorname{tr}(PC \otimes G) + \frac{1}{3}\operatorname{tr}(G \otimes PC) \\
&= \frac{27}{4}(\operatorname{tr}(DP))^2 - 3\operatorname{tr}(DDP) - 2\operatorname{tr}(PC \otimes G), \qquad (25)
\end{aligned}
$$

where the final equality uses the fact that $\operatorname{tr}(DPDP) = (\operatorname{tr}(DP))^2$ since P has rank one.

8.6 Calculation of b_1

First observe that

$$
\begin{aligned}
K_{12}J &= \left(I \otimes G \otimes \left(\overline{P} - P\right)\right)\operatorname{vec}D, \\
K_{21}J &= \left(\left(\overline{P} - P\right) \otimes I \otimes G\right)\operatorname{vec}D.
\end{aligned}
$$

Second, observe that

$$\left(\left(\overline{P}-P\right)\otimes I\right)\operatorname{vec}\overline{P}=\operatorname{vec}\overline{P}$$

and

$$\left(\left(\overline{P}-P\right)\otimes I\right)\operatorname{vec}P=-\operatorname{vec}P$$

Using the facts that J lies in the span of $(G\otimes I\otimes I)$, $K_{12}J$ lies in the span of $(I\otimes G\otimes I)$, and $K_{21}J$ lies in the span of $(I\otimes I\otimes G)$, we see that (16) equals

$$\begin{aligned}
b_1 &= (G'G)^{-1}\left\{J'\left(P\otimes\overline{P}\otimes\overline{P}\right)J+J'\left(P\otimes\left(\operatorname{vec}\overline{P}\right)\left(\operatorname{vec}\overline{P}\right)'\right)J\right. \\
&\quad \left.+J'\left(P\otimes K\left(\overline{P}\otimes\overline{P}\right)\right)J\right\} \\
&= 2\operatorname{vec}D'\left(\overline{P}\otimes\overline{P}\right)\operatorname{vec}D'+\operatorname{vec}D'\left(\operatorname{vec}\overline{P}\right)\left(\operatorname{vec}\overline{P}\right)'\operatorname{vec}D \\
&= 2\operatorname{tr}\left(\overline{P}D\overline{P}D\right)+\left(\operatorname{tr}\left(\overline{P}D\right)\right)^2. \qquad (26)
\end{aligned}$$

8.7 Calculation of b_2

Using similar reasoning, we find that (17) equals

$$\begin{aligned}
b_2 &= (G'G)^{-1}\left\{2\left(K_{12}J+K_{21}J\right)'\left(\overline{P}\otimes P\otimes P\right)\left(K_{12}J+K_{21}J\right)\right. \\
&\quad +2J'\left(P\otimes\left(\operatorname{vec}\overline{P}\right)\left(\operatorname{vec}P\right)'\right)HJ+2J'\left(P\otimes K\left(\overline{P}\otimes P\right)\right)J \\
&\quad \left.+J'\left(K\left(P\otimes P\right)\otimes\overline{P}\right)J+J'K_{12}\left(\overline{P}\otimes K\left(P\otimes P\right)\right)K_{21}J\right\}.
\end{aligned}$$

We take the terms on the RHS in turn. First, since

$$P\otimes G=(G'G)^{-1}\left(GG'\otimes G\right)=(G'G)^{-1}\left(G\otimes GG'\right)=G\otimes P,$$

then

$$\left(\overline{P}\otimes P\otimes P\right)\left(K_{12}J+K_{21}J\right)=\left[-\left(\overline{P}\otimes G\otimes P\right)+\left(\overline{P}\otimes P\otimes G\right)\right]\operatorname{vec}D=0.$$

Second,

$$\begin{aligned}
J'\left(P\otimes\left(\operatorname{vec}\overline{P}\right)\left(\operatorname{vec}P\right)'\right)HJ &= \operatorname{vec}D'\left(G'\otimes\left(\operatorname{vec}\overline{P}\right)\left(\operatorname{vec}P\right)'\right)HJ \\
&= -3(G'G)\left(\operatorname{vec}D\right)'\operatorname{vec}\overline{P}\left(\operatorname{vec}P\right)'\operatorname{vec}D \\
&= -3(G'G)\operatorname{tr}\left(\overline{P}D\right)\operatorname{tr}\left(PD\right).
\end{aligned}$$

Third

$$\begin{aligned} J'\left(P\otimes K\left(\overline{P}\otimes P\right)\right)J &= -G'G\,\mathrm{vec}\,D'\left(P\otimes\overline{P}\right)\mathrm{vec}\,D \\ &= -G'G\,\mathrm{tr}\left(\overline{P}DPD\right). \end{aligned}$$

Fourth,

$$J'\left(K\left(P\otimes P\right)\otimes\overline{P}\right)J = G'G\,\mathrm{vec}\,D'\left(P\otimes\overline{P}\right)\mathrm{vec}\,D = G'G\,\mathrm{tr}\left(\overline{P}DPD\right).$$

Fifth,

$$J'K_{12}\left(\overline{P}\otimes K\left(P\otimes P\right)\right)K_{21}J = G'G\,\mathrm{tr}\left(\overline{P}DPD\right).$$

Hence

$$b_2 = -6\,\mathrm{tr}\left(\overline{P}D\right)\mathrm{tr}\left(PD\right). \tag{27}$$

8.8 Calculation of b_3

First, (18) equals

$$\begin{aligned} b_3 &= \left(G'G\right)^{-1}\{J'H\left(P\otimes P\otimes P\right)J + J'H\left(P\otimes P\otimes P\right)HJ \\ &\quad +J'\left(P\otimes K\left(P\otimes P\right)\right)J + J'\left(K\left(P\otimes P\right)\otimes P\right)J \\ &\quad +J'K_{12}\left(P\otimes K\left(P\otimes P\right)\right)K_{21}J\}. \end{aligned} \tag{28}$$

Next, note that

$$\left(P\otimes P\otimes P\right)J = -\left(G\otimes P\otimes P\right)\mathrm{vec}\,D$$

and

$$\begin{aligned} \left(P\otimes P\otimes P\right)HJ &= -\left[\left(G\otimes P\otimes P\right)+\left(P\otimes G\otimes P\right)+\left(P\otimes P\otimes G\right)\right]\mathrm{vec}\,D \\ &= -3\left(G\otimes P\otimes P\right)\mathrm{vec}\,D, \end{aligned}$$

since $G\otimes P = P\otimes G$. Thus the first term in (28) is

$$\begin{aligned} J'H\left(P\otimes P\otimes P\right)J &= 3(G'G)\,\mathrm{vec}\,D'\left(P\otimes P\right)\mathrm{vec}\,D \\ &= 3(G'G)\,\mathrm{tr}\left(PDPD\right) \\ &= 3(G'G)\left(\mathrm{tr}\left(PD\right)\right)^2, \end{aligned}$$

the last equality since the rank of PD is one. Similarly, the second term is

$$J'H\left(P\otimes P\otimes P\right)HJ=9(G'G)\left(\mathrm{tr}\left(PD\right)\right)^2.$$

The third fourth, and fifth terms are similar:

$$J'\left(P\otimes K\left(P\otimes P\right)\right)J+J'\left(K\left(P\otimes P\right)\otimes P\right)J+J'K_{12}\left(P\otimes K\left(P\otimes P\right)\right)K_{21}J=3(G'G)\left(\mathrm{tr}\left(PD\right)\right)^2.$$

Hence

$$b_3=15\left(\mathrm{tr}\left(PD\right)\right)^2. \tag{29}$$

8.9 Calculation of Final Coefficients

We now can calculate the coefficients α_0 through α_3 for the expansion of the distribution of the Wald statistic. From (9), (23) and (26), we have

$$\alpha_0=\frac{1}{4}\left(4a_0-b_1\right)=\frac{1}{2}\mathrm{tr}\left(D\overline{P}D\overline{P}\right)+\frac{1}{4}\left(\mathrm{tr}\left(D\overline{P}\right)\right)^2-\frac{1}{2}\mathrm{tr}\left(\overline{P}D\overline{P}D\right)-\frac{1}{4}\left(\mathrm{tr}\left(\overline{P}D\right)\right)^2=0.$$

From (10), (24), (26) and (27), we have

$$\begin{aligned}\alpha_1 &= \frac{1}{4}\left(4a_1+b_1-b_2\right)\\ &= -\frac{3}{2}\mathrm{tr}\left(DP\right)\mathrm{tr}\left(D\overline{P}\right)-\mathrm{tr}\left(D\overline{P}D\overline{P}\right)+\frac{1}{2}\mathrm{tr}\left(\overline{P}D\overline{P}D\right)\\ &\quad +\frac{1}{4}\left(\mathrm{tr}\left(\overline{P}D\right)\right)^2+\frac{3}{2}\mathrm{tr}\left(\overline{P}D\right)\mathrm{tr}\left(PD\right)\\ &= -\frac{1}{2}\mathrm{tr}\left(\overline{P}D\overline{P}D\right)+\frac{1}{4}\left(\mathrm{tr}\left(\overline{P}D\right)\right)^2.\end{aligned}$$

From (11), (25), (27) and (29)

$$\begin{aligned}\alpha_2 &= \frac{1}{12}\left(4a_2+b_2-b_3\right)\\ &= \frac{9}{4}\left(\mathrm{tr}\left(DP\right)\right)^2-\mathrm{tr}\left(DDP\right)-\frac{2}{3}\mathrm{tr}\left(PC\otimes G\right)-\frac{1}{2}\mathrm{tr}\left(\overline{P}D\right)\mathrm{tr}\left(PD\right)-\frac{5}{4}\left(\mathrm{tr}\left(PD\right)\right)^2\\ &= \left(\mathrm{tr}\left(DP\right)\right)^2-\mathrm{tr}\left(PDD\right)-\frac{2}{3}\mathrm{tr}\left(PC\otimes G\right)-\frac{1}{2}\mathrm{tr}\left(\left(I-P\right)D\right)\mathrm{tr}\left(PD\right)\\ &= \frac{3}{2}\left(\mathrm{tr}\left(DP\right)\right)^2-\mathrm{tr}\left(PDD\right)-\frac{2}{3}\mathrm{tr}\left(PC\otimes G\right)-\frac{1}{2}\mathrm{tr}\left(D\right)\mathrm{tr}\left(PD\right).\end{aligned}$$

From (12) and (29)

$$\alpha_3=\frac{1}{60}b_3=\frac{1}{4}\left(\mathrm{tr}\left(PD\right)\right)^2,$$

completing the proof of Theorem 1.

9 Appendix C: Proof of Theorem 2

9.1 Expansion for GMM Statistic

In the simplified setting for the theorem, $\tilde{\beta} = \text{argmin}_{g(\beta)=0} \ J(\beta)$, where

$$J(\beta) = \ n^{-1} \left(Y - X\beta\right)' XX' \left(Y - X\beta\right).$$

Newey and McFadden (1994) established that $\tilde{\beta} \to_p \beta_0$ (Theorem 9.1) and that $\tilde{q} = \sqrt{n}\left(\tilde{\beta} - \beta_0\right) = O_p(1)$ (p. 2219). Thus with probability that tends to one as $n \to \infty$, $\tilde{\beta}$ lies in the interior of the parameter space and there exists a Lagrange multiplier $\tilde{\lambda}$ such that

$$0 = -\tilde{m} + \tilde{G}\tilde{\lambda}. \tag{30}$$

where $\tilde{m} = n^{-1/2}X'\left(Y - X\tilde{\beta}\right)$ and $\tilde{G} = G(\tilde{\beta})$. Hence we can write $\tilde{\lambda} = \left(G'\tilde{G}\right)^{-1} G'\tilde{m}$ and

$$\tilde{m} = \tilde{G}\left(G'\tilde{G}\right)^{-1} G'\tilde{m}. \tag{31}$$

Expanding $\tilde{G} = G(\tilde{\beta})$ about β_0, and using the fact that $\tilde{q} = O_p(1)$, we see

$$\tilde{G} = G + n^{-1/2}D\tilde{q} = G + O_p(n^{-1/2}). \tag{32}$$

(31) and (32) combine to yield $\tilde{m} = P\tilde{m} + O_p(n^{-1/2})$, or

$$\overline{P}\tilde{m} = O_p(n^{-1/2}). \tag{33}$$

From Lemma 1 evaluated at $\tilde{\beta}$ and noting that $g(\tilde{\beta}) = g(\beta_0) = 0$, we have

$$0 = G'\tilde{q} + n^{-1/2}\frac{1}{2}\tilde{q}'D\tilde{q} + n^{-1}\frac{1}{6}\tilde{q}'C\left(\tilde{q}\otimes\tilde{q}\right) + O_p\left(n^{-3/2}\right). \tag{34}$$

One implication of (34) is that $G'\tilde{q} = O_p(n^{-1/2})$ and hence $P\tilde{q} = O_p(n^{-1/2})$. Thus since $\tilde{q} = m - \tilde{m}$,

$$\begin{aligned} \tilde{q} &= \overline{P}\tilde{q} + O_p(n^{-1/2}) \\ &= \overline{P}m - \overline{P}\tilde{m} + O_p(n^{-1/2}) \\ &= \overline{P}m + O_p(n^{-1/2}), \end{aligned}$$

where the last equality is (33). We have established that

$$\tilde{q} = \overline{P}m + O_p(n^{-1/2}). \tag{35}$$

Our next task is to obtain an expansion of the form $\tilde{q} = \overline{P}m + n^{-1/2}q_1 + O_p(n^{-1})$. Applying (35) to (32), we find

$$\tilde{G} = G + n^{-1/2}D\overline{P}m + O_p(n^{-1}).$$

Thus

$$\begin{aligned} \tilde{G}\left(G'\tilde{G}\right)^{-1} &= G\left(G'G\right)^{-1} + n^{-1/2}\left[\left(G'G\right)^{-1}D\overline{P}m - G\left(G'G\right)^{-2}G'D\overline{P}m\right] + O_p(n^{-1}) \\ &= G\left(G'G\right)^{-1} + n^{-1/2}\left(G'G\right)^{-1}\overline{P}D\overline{P}m + O_p(n^{-1}). \end{aligned} \tag{36}$$

Applying $\tilde{m} = m - \tilde{q}$ to (34), we find that

$$G'\tilde{m} = G'm + n^{-1/2}\frac{1}{2}\tilde{q}'D\tilde{q} + n^{-1}\frac{1}{6}\tilde{q}'C\left(\tilde{q}\otimes\tilde{q}\right) + O_p\left(n^{-3/2}\right). \tag{37}$$

Combined with (35), this implies

$$G'\tilde{m} = G'm + n^{-1/2}\frac{1}{2}m'\overline{P}D\overline{P}m + O_p(n^{-1}). \tag{38}$$

Thus (31), (36), and (38) combine as

$$\tilde{m} = G\left(G'G\right)^{-1}G'\tilde{m} + n^{-1/2}\left(G'G\right)^{-1}\left[G'm\overline{P}D\overline{P}m + \frac{1}{2}Gm'\overline{P}D\overline{P}m\right] + O_p(n^{-1}).$$

This implies

$$\tilde{q} = \overline{P}m - n^{-1/2}\left(G'G\right)^{-1}\left[G'm\overline{P}D\overline{P}m + \frac{1}{2}Gm'\overline{P}D\overline{P}m\right] + O_p(n^{-1}), \tag{39}$$

as desired.

From (31), we have the representation for the test statistic

$$\begin{aligned} DM &= \tilde{m}'\tilde{m} \\ &= \tilde{m}'G\left(\tilde{G}'G\right)^{-1}\tilde{G}'\tilde{G}\left(G'\tilde{G}\right)^{-1}G'\tilde{m} \\ &= \left(\tilde{m}'G\right)^2\ \Psi(\tilde{\beta}) \end{aligned} \tag{40}$$

where

$$\Psi(\beta) = \frac{G(\beta)'G(\beta)}{\left(G'G(\beta)\right)^2}.$$

We proceed by developing expansions for $\left(G'\tilde{m}\right)^2$ and $\Psi(\tilde{\beta})$, each to the order $O_p(n^{-3/2})$.

We first take $\left(G'\tilde{m}\right)^2$. (37) combined with (39) yields

$$G'\tilde{m} = G'm + n^{-1/2}\frac{1}{2}m'\overline{P}D\overline{P}m + n^{-1}\Upsilon + O_p\left(n^{-3/2}\right),$$

where

$$\Upsilon = -\left(G'G\right)^{-1} m'\overline{P}D\left[G'm\overline{P}D\overline{P}m + \frac{1}{2}Gm'\overline{P}D\overline{P}m\right] + \frac{1}{6}m'\overline{P}C\left(\overline{P}m \otimes \overline{P}m\right).$$

Thus

$$\left(G'\tilde{m}\right)^2 = \left(G'm\right)^2 + n^{-1/2}G'mm'\overline{P}D\overline{P}m + n^{-1}\left[\frac{1}{4}\left(m'\overline{P}D\overline{P}m\right)^2 + 2G'm\Upsilon\right] + O_p\left(n^{-3/2}\right). \tag{41}$$

Second, consider $\Psi(\tilde{\beta})$. Note that

$$\frac{\partial}{\partial\beta}\Psi(\beta) = \frac{2}{\left(G'G(\beta)\right)^2}D(\beta)G(\beta) - \frac{2G(\beta)'G(\beta)}{\left(G'G(\beta)\right)^3}D(\beta)G(\beta),$$

so

$$\begin{aligned}\frac{\partial}{\partial\beta}\Psi(\beta_0) &= \frac{2}{\left(G'G\right)^2}DG - \frac{2G'G}{\left(G'G\right)^3}DG \\ &= 0.\end{aligned}$$

We also calculate that

$$\begin{aligned}\frac{\partial^2}{\partial\beta\partial\beta'}\Psi(\beta_0) &= \frac{2}{\left(G'G\right)^2}DD - \frac{2}{\left(G'G\right)^3}DGG'H \\ &= \frac{2}{\left(G'G\right)}D\overline{P}D.\end{aligned}$$

Thus a second-order Taylor expansion yields

$$\begin{aligned}\Psi(\tilde{\beta}) &= \Psi(\beta_0) + n^{-1/2}\frac{\partial}{\partial\beta}\Psi(\beta_0)\tilde{q} + n^{-1}\frac{1}{2}\tilde{q}'\frac{\partial^2}{\partial\beta\partial\beta'}\Psi(\beta_0)\tilde{q} + O_p(n^{-3/2}) \\ &= (G'G)^{-1}\left\{1 + n^{-1}(G'G)^{-1}m'\overline{P}D\overline{P}D\overline{P}m\right\} + O_p(n^{-3/2}). \end{aligned} \tag{42}$$

Combining (40), (41), and (42),

$$DM = \frac{\left(G'm\right)^2 + n^{-1/2}u + n^{-1}v}{G'G} + O_p(n^{-3/2}),$$

where

$$u = G'mm'\overline{P}D\overline{P}m$$

and

$$\begin{aligned} v &= \frac{1}{4}\left(m'\overline{P}D\overline{P}m\right)^2 + 2G'm\Upsilon + (G'm)^2(G'G)^{-1}m'\overline{P}D\overline{P}D\overline{P}m \\ &= \frac{1}{4}\left(m'\overline{P}D\overline{P}m\right)^2 - m'Pmm'\overline{P}D\overline{P}D\overline{P}m \\ &\quad -m'PD\overline{P}mm'\overline{P}D\overline{P}m + G'm\frac{1}{3}m'\overline{P}C\left(\overline{P}m \otimes \overline{P}m\right)\end{aligned}$$

We can write $u = J' (m \otimes m \otimes m)$, where

$$J = \left(G \otimes \overline{P} \otimes \overline{P}\right) \operatorname{vec} D,$$

and $v = \operatorname{tr}[L (mm' \otimes mm')]$, where

$$\begin{aligned} L &= \frac{1}{4} \left(\overline{P} \otimes \overline{P}\right) (\operatorname{vec} D) (\operatorname{vec} D)' \left(\overline{P} \otimes \overline{P}\right) - (\operatorname{vec} P) \left(\operatorname{vec} D\overline{P}D\right)' \left(\overline{P} \otimes \overline{P}\right) \\ &\quad - \left(\overline{P} \otimes P\right) (\operatorname{vec} D) (\operatorname{vec} D)' \left(\overline{P} \otimes \overline{P}\right) + \frac{1}{3} \left(\overline{P}C \left(\overline{P} \otimes \overline{P}\right)\right) \otimes G. \end{aligned}$$

Using Theorem 3 in Appendix A, the coefficients of the expansion (8) are found by calculation of the coefficients $a_0, a_1, a_2, b_1, b_2, b_3$ from the above expressions for J and L. We calculate each explicitly.

9.2 Calculation of a_0

First,

$$\left(\overline{P} \otimes \overline{P}\right) L = \frac{1}{4} \left(\overline{P} \otimes \overline{P}\right) (\operatorname{vec} D) (\operatorname{vec} D)' \left(\overline{P} \otimes \overline{P}\right),$$

so

$$\operatorname{tr}\left\{\left(\overline{P} \otimes \overline{P}\right) L (I + K)\right\} = \frac{1}{2} (\operatorname{vec} D)' \left(\overline{P} \otimes \overline{P}\right) (\operatorname{vec} D) = \frac{1}{2} \operatorname{tr}\left(\overline{P}D\overline{P}D\right).$$

Second,

$$\left(\operatorname{vec} \overline{P}\right)' L \left(\operatorname{vec} \overline{P}\right) = \frac{1}{4} \left(\operatorname{vec} \overline{P}\right)' (\operatorname{vec} D) (\operatorname{vec} D)' \left(\operatorname{vec} \overline{P}\right) = \frac{1}{4} \left(\operatorname{tr}\left(\overline{P}D\right)\right)^2.$$

Summing these terms, we see that (13) equals

$$a_0 = \frac{1}{2} \operatorname{tr}\left(\overline{P}D\overline{P}D\right) + \frac{1}{4} \left(\operatorname{tr}\left(\overline{P}D\right)\right)^2. \tag{43}$$

9.3 Calculation of a_1

First

$$\left(P \otimes \overline{P}\right) L + \left(\overline{P} \otimes P\right) L = -\left(\overline{P} \otimes P\right) (\operatorname{vec} D) (\operatorname{vec} D)' \left(\overline{P} \otimes \overline{P}\right) + \frac{1}{3} \left(\overline{P}C \left(\overline{P} \otimes \overline{P}\right)\right) \otimes G,$$

so

$$
\begin{aligned}
& \operatorname{tr}\left\{\left(\left(P \otimes \overline{P}\right)+\left(\overline{P} \otimes P\right)\right) L\,(I+K)\right\} \\
= & -2\left(\operatorname{vec} P D \overline{P}\right)^{\prime}\left(\operatorname{vec} \overline{P} D \overline{P}\right)+\frac{1}{3} \operatorname{tr}\left\{\left(\overline{P} C\left(\overline{P} \otimes \overline{P}\right)\right) \otimes G\right\}+\frac{1}{3} \operatorname{tr}\left\{G \otimes\left(\overline{P} C\left(\overline{P} \otimes \overline{P}\right)\right)\right\} \\
= & \frac{1}{3} \operatorname{tr}\left\{\left(\overline{P} C \otimes G\right)\left(\overline{P} \otimes \overline{P}\right)\right\}+\frac{1}{3} \operatorname{tr}\left\{\left(G \otimes \overline{P} C\right)\left(\overline{P} \otimes \overline{P}\right)\right\} \\
= & \frac{1}{3} \operatorname{tr}\left\{\left(\overline{P} \otimes \overline{P}\right)\left(\overline{P} C \otimes G\right)\right\}+\frac{1}{3} \operatorname{tr}\left\{\left(\overline{P} \otimes \overline{P}\right)\left(G \otimes \overline{P} C\right)\right\} \\
= & 0
\end{aligned}
$$

Second,

$$
\begin{aligned}
\left(\operatorname{vec} P\right)^{\prime} L\left(\operatorname{vec} \overline{P}\right) &= -\left(\operatorname{vec} D \overline{P} D\right)^{\prime}\left(\operatorname{vec} \overline{P}\right)+\frac{1}{3}\left(\operatorname{vec} P\right)^{\prime}\left(\left(\overline{P} C\left(\overline{P} \otimes \overline{P}\right)\right) \otimes G\right)\left(\operatorname{vec} \overline{P}\right) \\
&= -\operatorname{tr}\left(\overline{P} D \overline{P} D\right)+\frac{1}{3}\left(\operatorname{vec} P\right)^{\prime}\left(\overline{P} C \otimes G\right)\left(\operatorname{vec} \overline{P}\right) \\
&= -\operatorname{tr}\left(\overline{P} D \overline{P} D\right)+\frac{1}{3}\left(\operatorname{vec} G^{\prime} P \overline{P} C\right)^{\prime}\left(\operatorname{vec} \overline{P}\right) \\
&= -\operatorname{tr}\left(\overline{P} D \overline{P} D\right)
\end{aligned}
$$

Third,

$$
\begin{aligned}
\left(\operatorname{vec} \overline{P}\right)^{\prime} L\left(\operatorname{vec} P\right) &= \frac{1}{3}\left(\operatorname{vec} \overline{P}\right)^{\prime}\left(\left(\overline{P} C\left(\overline{P} \otimes \overline{P}\right)\right) \otimes G\right)\left(\operatorname{vec} P\right) \\
&= \frac{1}{3}\left(\operatorname{vec} \overline{P}\right)^{\prime}\left(\overline{P} C \otimes G\right)\left(\overline{P} \otimes \overline{P}\right)\left(\operatorname{vec} P\right) \\
&= 0.
\end{aligned}
$$

Summing these terms, we see that (14) equals

$$
a_1=-\operatorname{tr}\left(\overline{P} D \overline{P} D\right). \tag{44}
$$

9.4 Calculation of a_2

Note that

$$
\left(P \otimes P\right) L=-\left(\operatorname{vec} P\right)\left(\operatorname{vec} D \overline{P} D\right)^{\prime}\left(\overline{P} \otimes \overline{P}\right),
$$

so

$$
a_2=2 \operatorname{tr}\left\{\left(P \otimes P\right) L\right\}+\operatorname{tr}\left\{\left(P \otimes P\right) L K\right\}=-3 \operatorname{tr}\left\{P \overline{P} D \overline{P} D \overline{P}\right\}=0. \tag{45}
$$

9.5 Calculation of $b_1,\ b_2,\ b_3$.

First, observe that

$$\begin{aligned}
\left(P \otimes \overline{P} \otimes \overline{P}\right) HJ &= \left(G \otimes \overline{P} \otimes \overline{P}\right) \operatorname{vec} D, \\
\left(P \otimes K\left(\overline{P} \otimes \overline{P}\right)\right) J &= \left(G \otimes \overline{P} \otimes \overline{P}\right) \operatorname{vec} D, \\
K_{21} J &= \left(\overline{P} \otimes \overline{P} \otimes G\right) \operatorname{vec} D
\end{aligned}$$

Thus (16) equals

$$\begin{aligned}
b_1 &= \left(G'G\right)^{-1} (\operatorname{vec} D)' \left\{2G'G\left(\overline{P} \otimes \overline{P}\right) + G'G\left(\operatorname{vec} \overline{P}\right)\left(\operatorname{vec} \overline{P}\right)'\right\} \operatorname{vec} D \\
&= 2 \operatorname{tr}\left(\overline{P}D\overline{P}D\right) + \left(\operatorname{tr}\left(\overline{P}D\right)\right)^2. \qquad (46)
\end{aligned}$$

By simple projection calculations, it is simple to calculate that $b_2 = 0$ and $b_3 = 0$.

9.6 Calculation of Final Coefficients

We now can calculate the coefficients α_0 through α_3 for the expansion of the distribution of the GMM statistic. From (9), (43) and (46), we have

$$\begin{aligned}
\alpha_0 &= \frac{1}{4}\left(4a_0 - b_1\right) \\
&= \frac{1}{2} \operatorname{tr}\left(\overline{P}D\overline{P}D\right) + \frac{1}{4}\left(\operatorname{tr}\left(\overline{P}D\right)\right)^2 - \left(\frac{1}{2} \operatorname{tr}\left(\overline{P}D\overline{P}D\right) + \frac{1}{4}\left(\operatorname{tr}\left(\overline{P}D\right)\right)^2\right) \\
&= 0.
\end{aligned}$$

From (10), (44), (46) and $b_2 = 0$, we have

$$\begin{aligned}
\alpha_1 &= \frac{1}{4}\left(4a_1 + b_1 - b_2\right) \\
&= -\operatorname{tr}\left(\overline{P}D\overline{P}D\right) + \frac{1}{2} \operatorname{tr}\left(\overline{P}D\overline{P}D\right) + \frac{1}{4}\left(\operatorname{tr}\left(\overline{P}D\right)\right)^2 \\
&= -\frac{1}{2} \operatorname{tr}\left(\overline{P}D\overline{P}D\right) + \frac{1}{4}\left(\operatorname{tr}\left(\overline{P}D\right)\right)^2.
\end{aligned}$$

From (11), $a_2 = 0$, $b_2 = 0$, and $b_3 = 0$

$$\alpha_2 = \frac{1}{12}\left(4a_2 + b_2 - b_3\right) = 0$$

From (12) and $b_3 = 0$

$$\alpha_3 = \frac{1}{60} b_3 = 0,$$

completing the proof of Theorem 2.

References

[1] Gregory, A. and M. Veall (1985): "On formulating Wald tests of nonlinear restrictions," *Econometrica*, 53, 1465-1468.

[2] Lafontaine, F. and K.J. White (1986): "Obtaining any Wald statistic you want," *Economics Letters*.

[3] Magnus, J. R., and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: John Wiley and Sons.

[4] Newey, W.K. and D.L. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of Econometrics, Vol IV*, R.F. Engle and D.L McFadden, eds., 2113-2245. New York: Elsevier.

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

[6] Park, J. Y. and P.C.B. Phillips (1988): "On the formulation of Wald tests of nonlinear restrictions," *Econometrica*, 56, 1065-1083.