Cambridge University Press
978-0-521-18430-4 - Econometric Theory and Practice: Frontiers of Analysis and Applied Research
Edited by Dean Corbae, Steven N. Durlauf and Bruce E. Hansen
Excerpt
More information

PART ONE

HIGHER-ORDER ASYMPTOTICS

ONE

Edgeworth Expansions for the Wald and GMM Statistics for Nonlinear Restrictions

Bruce E. Hansen*

1.1 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 the Wald test is generically 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) showed 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 generalized method of moments (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 of being 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 chapter attempts to fill this gap by providing an Edgeworth

* This research was supported by a grant from the National Science Foundation. I thank Yuichi Kitamura and Ken West for helpful comments and discussions.

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 formulas 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 the Edgeworth expansion approximation.

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, if the 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, whereas 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 chapter is organized as follows. Section 1.2 states the model and test statistics. Section 1.3 describes alternative methods to calculate the covariance matrix of the estimates and the weight matrix for GMM estimation. Section 1.4 presents our main results. Section 1.5 is a Monte Carlo simulation. A brief conclusion follows in Section 1.6. Appendix A is a restatement of the Park–Phillips (1988) Edgeworth expansion (for reference). Appendix B contains the proof of Theorem 1.1 (the Edgeworth expansion for the Wald statistic). Appendix C contains the proof of Theorem 1.2 (the Edgeworth expansion for the GMM statistic).

A Gauss program that calculates the GMM statistics described in this chapter can be downloaded from my Web page <www.ssc.wisc.edu/˜bhansen>.

1.2 LINEAR REGRESSION WITH NONLINEAR HYPOTHESES

The model is a linear regression

$$y_i = x_i'\beta + e_i$$
$$E(x_i e_i) = 0,$$

$i = 1, \ldots, n$, 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 ordinary least squares (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

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

The strengths of the Wald statistic are that it is easy to compute but asymptotically χ_1^2 under H_0 and 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 $\beta \in R^k$, that is,

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

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

The restricted GMM estimator minimizes $J(\beta)$ subject to 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 because the criterion $J(\beta)$ is

Cambridge University Press
978-0-521-18430-4 - Econometric Theory and Practice: Frontiers of Analysis and Applied Research
Edited by Dean Corbae, Steven N. Durlauf and Bruce E. Hansen
Excerpt
More information

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:

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

The statistic (4) has several 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. In the special case in which g is linear, however, 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).

1.3 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

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

For 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) may also be computed setting Ω_n to equal either $\hat{\Omega}_n$ or $\hat{\Omega}_n^0$, the latter being valid only under the assumption of homoskedasticity. These choices correspond to computing the weight matrix under the alternative hypothesis because 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

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

for the general case, or

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

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

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

1.4 EDGEWORTH EXPANSIONS

Park and Phillips (1988) used an Edgeworth expansion to show that the noninvariance 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 to use the same Edgeworth expansion argument to show that the GMM statistic has an Edgeworth approximation to the chi-square distribution superior to that of 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; thus, $\Omega_n = nI_n$. Although 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; thus, the divergence from the χ_1^2 is due only to the nonlinearity of g.

On the assumption that $g(\beta)$ is three-times continuously differentiable, define

$$\underset{k\times 1}{G(\beta)} = \frac{\partial}{\partial\beta} g(\beta),$$
$$\underset{k\times k}{D(\beta)} = \frac{\partial^2}{\partial\beta\partial\beta'} g(\beta), \text{ and}$$
$$\underset{k\times k^2}{C(\beta)} = \frac{\partial}{\partial\beta}\left((\operatorname{vec} D(\beta))'\right),$$

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

Define the projection matrices

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

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

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

where

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

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

and

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

Theorem 1.2 *The asymptotic expansion of DM as $n \to \infty$ is given by*

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

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 1.1 and 1.2.

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 $(1 - n^{-1}(G'G)^{-1}\alpha_1)^{-1}$ DM is $F(x) + o(n^{-1})$, and thus 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, because 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 noninvariance 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.

1.5 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 $E(e_i \mid x_i) = 0$. In our experiments, we generate x_{1i}, x_{2i}, and e_i as mutually independent, indepent and identically distributed (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. Although Gregory–Veall only examined the behavior of the Wald statistic constructed with a conventional covariance matrix

Table 1.1 *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

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).

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.[2] The results are presented in Tables 1.1 and 1.2. 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

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

Table 1.2 *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

matrix estimate is used, these tests have minimal size distortion (because 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.

1.6 CONCLUSION

We have extended the explicit matrix approach to Edgeworth expansions developed by Park and Phillips (1988), extended their Edgeworth expansion for the Wald statistic, and 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.