E C O N O M I C S   B U L L E T I N

A simple explanation for the non-invariance of a Wald statistic to a reformulation of a null hypothesis

Naorayex K Dastoor
Department of Economics, University of Alberta

Abstract

For a given null hypothesis and its reformulation, the associated Wald statistics are shown to be members of a wider family of statistics where all members are asymptotically equivalent under the null hypothesis. Therefore, the non-invariance of a Wald statistic (to a reformulation of a null hypothesis) is equivalent to using different members of the wider family and, in addition, this non-invariance implies that these members use different estimators of an appropriate variance-covariance matrix.

Citation: Dastoor, Naorayex K, (2008) "A simple explanation for the non-invariance of a Wald statistic to a reformulation of a null hypothesis." *Economics Bulletin*, Vol. 3, No. 62 pp. 1-10
Submitted: August 4, 2008. **Accepted:** September 30, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume3/EB-08C10010A.pdf

1. INTRODUCTION

It is well known that, in general, a Wald statistic is not invariant to a reformulation of a null hypothesis where a vector of restrictions $r(\theta) = 0$ is rewritten in an algebraically equivalent form $q(\theta) = 0$ with θ being a vector of unknown parameters. Initially, Gregory and Veall (1985) provided Monte Carlo evidence of the effect of a reformulation and, subsequently, Lafontaine and White (1986) and Breusch and Schmidt (1988) showed how this non-invariance could be exploited to obtain a desired numerical value for a Wald statistic, Phillips and Park (1988) examined the effect of a reformulation on the small sample distribution of a Wald statistic, and Kemp (2001) provided a justification for ruling out certain reformulations. In contrast to the explanations provided by Davidson (1990) and Critchley, Marriott, and Salmon (1996), which apply the methods of differential geometry, this note provides a simple explanation for the non-invariance of a Wald statistic.

Using the terminology in Dastoor (2003), the original family of Wald statistics for testing $H_0 : r(\theta) = 0$ is a family where all members are asymptotically equivalent under H_0, and each member (called an original Wald statistic) is a quadratic form in $\sqrt{n}r(\hat{\theta}_n)$ with all components of its weighting matrix evaluated at $\hat{\theta}_n$, the (unrestricted) maximum likelihood estimator of θ based on n observations. Then, the extended family of Wald statistics is a wider family where all members are asymptotically equivalent under H_0, and each member (called an extended Wald statistic) is a quadratic form in $\sqrt{n}r(\hat{\theta}_n)$ with all components of its weighting matrix not necessarily evaluated at $\hat{\theta}_n$. In both these families, the weighting matrix of any member is (under H_0) a consistent estimator of the inverse of the asymptotic variance-covariance matrix of $\sqrt{n}r(\hat{\theta}_n)$. Similarly, the original and extended families of Wald statistics for testing $H_0^* : q(\theta) = 0$ are families whose members are appropriate quadratic forms in $\sqrt{n}q(\hat{\theta}_n)$ and asymptotically equivalent under H_0^* or, equivalently, under H_0. In general, the two original families differ, but it can be shown that the two extended families are identical. Therefore, an original Wald statistic for testing H_0 and an original Wald statistic for testing H_0^* are members of the extended family for testing H_0. This provides a simple explanation for the non-invariance of a Wald statistic; i.e., when H_0 is replaced with H_0^*, the non-invariance of a Wald statistic is equivalent to replacing one extended statistic for testing H_0 with a different extended statistic for testing H_0, and it can be shown that this non-invariance implies that the two extended statistics use different estimators of the asymptotic variance-covariance matrix of $\sqrt{n}r(\hat{\theta}_n)$ under H_0.

The next section presents the original and extended families for testing each of H_0 and H_0^*. Section 3 derives the simple explanation, and some concluding remarks are stated in Section 4.

2. ORIGINAL AND EXTENDED FAMILIES

Let θ be a $p \times 1$ vector of unknown parameters, $\Omega \subseteq \mathbb{R}^p$ be the parameter space, $L_n(\theta)$ be a log-likelihood function for n observations, and $r(\theta) = 0$ be an $r \times 1$ vector of known restrictions with $r \leq p$. Then, $\hat{\theta}_n = \operatorname{argmax}_{\theta\in\Omega} L_n(\theta)$ is the (unrestricted) maximum likelihood estimator of θ, and the null and alternative hypotheses are $H_0 : \theta \in \Omega_0$ and $H_1 : \theta \in \Omega_1$, respectively, where $\Omega_0 = \{\theta \,|\, r(\theta) = 0,\, \theta \in \Omega\}$ and Ω_1 constitute a partition of Ω. Also, let $R(\theta) = \partial r(\theta)/\partial\theta^\top$ be the $r \times p$ matrix of derivatives with rank r for all $\theta \in \Omega$, $R_0 = R(\theta_0)$, $\hat{R} = R(\hat{\theta}_n)$, and $J_n(\theta)$ be a $p \times p$ symmetric nonsingular matrix such that $\hat{J}_n = J_n(\hat{\theta}_n) \xrightarrow{p} J_0$ where θ_0 is the true value of θ, $J_0 = -\operatorname{plim} n^{-1}\partial^2 L_n(\theta_0)/\partial\theta\partial\theta^\top$ is the (positive definite) limiting information matrix under H_0, and $\xrightarrow{p}$ denotes convergence in probability under H_0. Throughout, $\theta_0 \in \Omega_0$, all asymptotic results are obtained under H_0, the usual regularity conditions are assumed to hold, and standard results will be used. Rigorous statements of the appropriate conditions required and formal derivations of standard results can be found in, for example, Davidson and MacKinnon (1993) and Newey and McFadden (1994). Therefore, under H_0 and appropriate conditions, $\hat{\theta}_n \xrightarrow{p} \theta_0$, $\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{a}{\sim} N(0,\, J_0^{-1})$, and $\sqrt{n}r(\hat{\theta}_n) \overset{a}{\sim} N(0,\, R_0 J_0^{-1} R_0^\top)$. Then, an original Wald statistic for testing H_0 is

$$W_n(\hat{J}_n) = nr(\hat{\theta}_n)^\top \{\hat{R}\hat{J}_n^{-1}\hat{R}^\top\}^{-1} r(\hat{\theta}_n), \quad \textbf{(1)}$$

which is asymptotically distributed as a $\chi^2(r)$ variate under H_0.

Henceforth, as all asymptotic results are obtained under H_0, the terms 'consistent estimator' and 'family of asymptotically-equivalent Wald statistics' will mean 'consistent estimator under H_0' and 'a family of Wald statistics where all members are asymptotically equivalent under H_0', respectively. Let $\hat{\mathscr{J}} = \{\hat{J}_n \,|\, \hat{J}_n = J_n(\hat{\theta}_n) = J_n(\hat{\theta}_n)^\top \xrightarrow{p} J_0,\ \hat{J}_n^{-1} \text{ exists}\}$, a set of consistent estimators of J_0 that are evaluated at $\hat{\theta}_n$. Then,

$$\mathscr{W} = \left\{ W_n(\hat{J}_n) \,|\; W_n(\hat{J}_n) \text{ is given by (1)}, \quad \hat{J}_n \in \hat{\mathscr{J}} \right\}$$

is the original family of asymptotically-equivalent Wald statistics for testing H_0. In this family, one member is distinguished from another only by the choice of $J_n(\theta)$ and all members evaluate their chosen $J_n(\theta)$ at $\hat{\theta}_n$. Therefore, an original Wald statistic is a quadratic form in $\sqrt{n}r(\hat{\theta}_n)$ where all components of its weighting matrix are evaluated at $\hat{\theta}_n$. Now, let $\mathscr{J} = \{J_n \,|\, J_n = J_n^\top \xrightarrow{p} J_0,\ J_n^{-1} \text{ exists}\}$ and $\mathscr{R} = \{R_n \,|\, R_n \xrightarrow{p} R_0,\ R_n \text{ has rank } r\}$ be sets of consistent estimators of J_0 and R_0, respectively. Then, replacing $\hat{J}_n$ and $\hat{R}$ in $W_n(\hat{J}_n)$ with the more general estimators J_n and R_n, respectively, gives a statistic that is asymptotically equivalent to $W_n(\hat{J}_n)$ under H_0. Therefore, an extended Wald statistic is

$$W_n(J_n, R_n) = nr(\hat{\theta}_n)^\top \{R_n J_n^{-1} R_n^\top\}^{-1} r(\hat{\theta}_n), \tag{2}$$

which corresponds to W_{1n} in Newey and McFadden (1994, Table 2, p. 2222) and which is a special case of ξ_n^w in Gourieroux and Monfort (1989, equation (37), p. 75) where the restrictions are written in a more general form than $r(\theta) = 0$. Then,

$$\mathscr{E}\mathscr{W} = \left\{ W_n(J_n, R_n) \,|\; W_n(J_n, R_n) \text{ is given by (2)}, \; J_n \in \mathscr{J}, \; R_n \in \mathscr{R} \right\}$$

is the extended family of asymptotically-equivalent Wald statistics for testing H_0; the sets $\hat{\mathscr{J}}$, $\mathscr{J}$, $\mathscr{R}$, $\mathscr{W}$, and $\mathscr{E}\mathscr{W}$ correspond to $\hat{\mathscr{A}}$, $\mathscr{A}$, $\mathscr{R}$, $\bar{\mathscr{W}}$, and $\bar{\mathscr{E}}_1$, respectively, in Dastoor (2003). In this extended family, one member is distinguished from another by the choice of J_n and R_n and, for each member, the chosen J_n and R_n need not necessarily be matrices evaluated at $\hat{\theta}_n$. Since $\hat{\mathscr{J}} \subset \mathscr{J}$, $\hat{R} \in \mathscr{R}$, and $W_n(\hat{J}_n) = W_n(\hat{J}_n, \hat{R})$, any original Wald statistic is an extended Wald statistic so $\mathscr{W} \subset \mathscr{E}\mathscr{W}$.

Let $\Omega_0^* = \{\theta \,|\, q(\theta) = 0,\ \theta \in \Omega\}$, $Q(\theta) = \partial q(\theta)/\partial\theta^\top$ be the $r \times p$ matrix of derivatives with rank r for all $\theta \in \Omega$, $\hat{Q} = Q(\hat{\theta}_n)$, and $Q_0 = Q(\theta_0)$ where $q(\theta)$ is such that $q(\theta) = 0$ if and only if $r(\theta) = 0$. Then, $\Omega_0^* = \Omega_0$ so $H_0^* : \theta \in \Omega_0^*$ is a reformulation of $H_0 : \theta \in \Omega_0$; cf. Dagenais and Dufour (1991, p. 1605) where $\psi(\theta)$ and $\bar{\psi}(\theta)$ correspond to $r(\theta)$ and $q(\theta)$, respectively. For testing H_0^*, an original Wald statistic is

$$W_n^*(\hat{J}_n) = nq(\hat{\theta}_n)^\top \{\hat{Q}\hat{J}_n^{-1}\hat{Q}^\top\}^{-1} q(\hat{\theta}_n), \tag{3}$$

which is asymptotically equivalent to $W_n(\hat{J}_n)$ under H_0,

$$\mathscr{W}^* = \left\{ W_n^*(\hat{J}_n) \,|\; W_n^*(\hat{J}_n) \text{ is given by (3)},\;\; \hat{J}_n \in \hat{\mathscr{J}} \right\}$$

is the original family of asymptotically-equivalent Wald statistics,

$$W_n^*(J_n, Q_n) = nq(\hat{\theta}_n)^\top \{Q_n J_n^{-1} Q_n^\top\}^{-1} q(\hat{\theta}_n) \tag{4}$$

is an extended Wald statistic, and the extended family of asymptotically-equivalent Wald statistics is

$$\mathscr{E}\mathscr{W}^* = \left\{ W_n^*(J_n, Q_n) \,|\; W_n^*(J_n, Q_n) \text{ is given by (4)},\;\; J_n \in \mathscr{J},\;\; Q_n \in \mathscr{Q} \right\}$$

where $\mathscr{Q} = \{Q_n \,|\, Q_n \xrightarrow{p} Q_0,\ Q_n \text{ has rank } r\}$ and $W_n^*(\hat{J}_n) = W_n^*(\hat{J}_n, \hat{Q}) \in \mathscr{W}^* \subset \mathscr{E}\mathscr{W}^*$.

For later reference, it is useful to note the following results, which are proved in the appendix. First, in general, there exist two $r \times r$ nonsingular matrices P_0 and $\bar{P}_n$ such that

$$Q_0 = P_0 R_0, \tag{5}$$

$$q(\hat{\theta}_n) = \bar{P}_n r(\hat{\theta}_n), \tag{6}$$

and $\bar{P}_n \xrightarrow{p} P_0$. Second, consider the special case of $q(\theta) = Pr(\theta)$ where P is an $r \times r$ non-stochastic nonsingular matrix whose elements do not depend on θ; i.e., $q(\theta)$ is a nonsingular linear transformation of $r(\theta)$. In this special case,

$$q(\hat{\theta}_n) = Pr(\hat{\theta}_n), \qquad \hat{Q} = P\hat{R}, \qquad Q_0 = PR_0, \qquad \text{and} \qquad \bar{P}_n = P_0 = P. \tag{7}$$

3. A SIMPLE EXPLANATION

The original families $\mathscr{W}$ and $\mathscr{W}^*$ differ, unless $q(\theta) = 0$ is a particular type of reformulation of $r(\theta) = 0$. For example, if $q(\theta)$ is a nonsingular linear transformation of $r(\theta)$, then $\mathscr{W} = \mathscr{W}^*$ as (1), (3), and (7) yield $W_n(\hat{J}_n) = W_n^*(\hat{J}_n)$; cf. Davidson and MacKinnon (1993, p. 469). Although $\mathscr{W} \neq \mathscr{W}^*$ in general, it can be shown that

$$\mathscr{E}\mathscr{W} = \mathscr{E}\mathscr{W}^*, \tag{8}$$

which is proved in the appendix. Basically, the extended Wald statistics use estimators of J_0, R_0, and Q_0 that have the flexibility to exploit the relationship between $r(\hat{\theta}_n)$ and $q(\hat{\theta}_n)$ in (6), which results in the equality of the extended families, whereas, the estimators used by the original Wald statistics are only those evaluated at $\hat{\theta}_n$, which cannot always exploit (6) so the original families differ in general. The equality of the extended families shows that, *for a given sample*, any extended Wald statistic for testing H_0^* is *identical* to some extended Wald statistic for testing H_0 (and vice versa) so (8) implies (but is not implied by) the asymptotic equivalence of $W_n(J_n, R_n)$ and $W_n^*(J_n, Q_n)$ under H_0. Therefore, the original Wald statistics will now be viewed as members of $\mathscr{E}\mathscr{W}$; i.e., $W_n(\hat{J}_n) = W_n(\hat{J}_n, \hat{R})$ and

$$W_n^*(\hat{J}_n) = W_n(\hat{J}_n, R^*) = W_n(J_n^*, \hat{R}) \tag{9}$$

where $R^* = \bar{P}_n^{-1}\hat{Q} \in \mathscr{R}$ and $J_n^* \in \mathscr{J}$ is a particular matrix whose form is given in the proof of (9) in the appendix. Also, it can be shown that $R^*\hat{J}_n^{-1}R^{*\top} = \hat{R}(J_n^*)^{-1}\hat{R}^\top$.

Let $\hat{V}_n = \hat{R}\hat{J}_n^{-1}\hat{R}^\top \xrightarrow{p} V_0$ and $V_n^* = R^*\hat{J}_n^{-1}R^{*\top} \xrightarrow{p} V_0$ where $V_0 = R_0 J_0^{-1} R_0^\top$ is the asymptotic variance-covariance matrix of $\sqrt{n}r(\hat{\theta}_n)$ under H_0. Then, $W_n(\hat{J}_n)$ and $W_n^*(\hat{J}_n)$ are quadratic forms in $\sqrt{n}r(\hat{\theta}_n)$ with weighting matrices $\hat{V}_n^{-1}$ and $(V_n^*)^{-1}$, respectively. Here, the estimation of V_0 (instead of just R_0 or just J_0) is relevant since (9) shows that $W_n^*(\hat{J}_n)$ can be obtained from (2) by setting either $R_n = R^*$ with $J_n = \hat{J}_n$ or $J_n = J_n^*$ with $R_n = \hat{R}$. Since $W_n(\hat{J}_n) = W_n^*(\hat{J}_n)$ if $\hat{V}_n = V_n^*$, a Wald statistic is invariant if a reformulation of H_0 as H_0^* does not result in $W_n(\hat{J}_n)$ and $W_n^*(\hat{J}_n)$ using different consistent estimators of V_0. For example, if $q(\theta)$ is a nonsingular linear transformation of $r(\theta)$, then (7) holds and $R^* = \bar{P}_n^{-1}\hat{Q}$ reduces to $R^* = \hat{R}$ so $\hat{V}_n = V_n^*$. In this case, $W_n(\hat{J}_n)$ and $W_n^*(\hat{J}_n)$ use the same consistent estimator of V_0 so a Wald statistic is invariant or, equivalently, these two statistics are identical extended statistics for testing H_0. However, if $W_n(\hat{J}_n) \neq W_n^*(\hat{J}_n)$, then $\hat{V}_n \neq V_n^*$. This provides a simple explanation for the non-invariance of a Wald statistic; i.e., when H_0 is replaced with H_0^*, the non-invariance of a Wald statistic is equivalent to replacing one extended statistic for testing H_0 with a different extended statistic for testing H_0 and, in addition, this non-invariance implies that $W_n(\hat{J}_n)$ and $W_n^*(\hat{J}_n)$ use different estimators $\hat{V}_n$ and V_n^*, respectively, as consistent estimators of V_0. Also, in the case where $r = 1$ with $r(\hat{\theta}_n) \neq 0$, it is easily seen that $W_n(\hat{J}_n) = W_n^*(\hat{J}_n)$ if and only if $\hat{V}_n = V_n^*$. Therefore, when testing a single restriction, the non-invariance of a Wald statistic is also equivalent to using different consistent estimators of V_0.

4. CONCLUDING REMARKS

Given the simple explanation, the results of Lafontaine and White (1986) and Breusch and Schmidt (1988) can be interpreted as showing how an estimator of V_0 can be easily chosen such that $W_n^*(\hat{J}_n)$ has a desired numerical value. In principle, some criterion could be used either to choose among estimators or to rule out certain estimators of V_0; indirectly, this would either provide an optimal formulation of the restrictions or rule out certain formulations, respectively. For example, the results of Phillips and Park (1988) and Kemp (2001) could be interpreted as providing some guidance on choosing an estimator and on ruling out certain estimators of V_0, respectively. Now, if two extended Wald statistics for testing H_0 use different consistent estimators of V_0, then it is reasonable to expect (if not require) the statistics to be different for a given sample. Therefore, since the non-invariance of a Wald statistic implies the use of different consistent estimators of V_0, this non-invariance should (contrary to econometrics folklore) not be viewed as an undesirable property of a Wald statistic, and especially in the case of testing a single restriction where the non-invariance is equivalent to using different consistent estimators of V_0.

APPENDIX

Proof of Equation (5). Let $R(\theta) = [R_1(\theta),\, R_2(\theta)]$ and $Q(\theta) = [Q_1(\theta),\, Q_2(\theta)]$ be conformably partitioned with $\theta = (\theta_1^\top, \theta_2^\top)^\top$ where θ_1 is an $r \times 1$ vector, and $R_1(\theta)$ and $Q_1(\theta)$ are $r \times r$ nonsingular matrices for $\theta \in \Omega_0$. Then, as shown by Dagenais and Dufour (1991, p. 1606), the implicit function theorem ensures that (for $\theta \in \Omega_0$) there exists a differentiable function h such that $\theta_1 = h(\theta_2)$ so

$$\frac{\partial \theta_1}{\partial \theta_2^\top} = \frac{\partial h}{\partial \theta_2^\top} = -R_1(\theta)^{-1} R_2(\theta) = -Q_1(\theta)^{-1} Q_2(\theta)$$

where the last equality follows as $q(\theta) = 0$ if and only if $r(\theta) = 0$. This last equality provides $Q_2(\theta_0) = Q_1(\theta_0) R_1(\theta_0)^{-1} R_2(\theta_0)$, which can be substituted into $Q_0 = [Q_1(\theta_0), Q_2(\theta_0)]$ to yield (5) where

$$P_0 = Q_1(\theta_0) R_1(\theta_0)^{-1} \qquad \textbf{(A.1)}$$

is an $r \times r$ nonsingular matrix.

Proof of Equation (6). Let $s_n(\theta) = \partial L_n(\theta)/\partial\theta$, λ be an $r \times 1$ vector of Lagrange multipliers, and $\tilde{R} = R(\tilde{\theta}_n)$ where $\tilde{\theta}_n$ is the restricted estimator of θ under H_0. Then, from the Lagrangean $\mathfrak{L}(\theta, \lambda) = L_n(\theta) - \lambda^\top r(\theta)$, the first-order condition $\partial \mathfrak{L}(\tilde{\theta}_n, \tilde{\lambda}_n)/\partial\theta = 0$ gives $s_n(\tilde{\theta}_n) = \tilde{R}^\top \tilde{\lambda}_n$. Another equation for $s_n(\tilde{\theta}_n)$ can be obtained from a mean-value expansion of $s_n(\tilde{\theta}_n)$ at $\hat{\theta}_n$ so, as $s_n(\hat{\theta}_n) = 0$, $s_n(\tilde{\theta}_n) = n\bar{J}_n(\hat{\theta}_n - \tilde{\theta}_n)$ where $\bar{J}_n$ is the matrix $-n^{-1}\partial^2 L_n(\theta)/\partial\theta\partial\theta^\top$ with each of its rows evaluated at a (possibly different) mean value given by a convex combination of $\hat{\theta}_n$ and $\tilde{\theta}_n$. Assuming that $\bar{J}_n$ is nonsingular, the two equations for $s_n(\tilde{\theta}_n)$ provide

$$\hat{\theta}_n - \tilde{\theta}_n = n^{-1} \bar{J}_n^{-1} \tilde{R}^\top \tilde{\lambda}_n. \qquad \textbf{(A.2)}$$

Since $r(\tilde{\theta}_n) = 0$, a mean value expansion of $r(\hat{\theta}_n)$ at $\tilde{\theta}_n$ gives $r(\hat{\theta}_n) = \bar{R}(\hat{\theta}_n - \tilde{\theta}_n)$ where $\bar{R}$ is the matrix $R(\theta)$ with each of its rows evaluated at a (possibly different) mean value given by a convex combination of $\hat{\theta}_n$ and $\tilde{\theta}_n$. Then, assuming that $\bar{R}\bar{J}_n^{-1}\tilde{R}^\top$ is nonsingular, substituting (A.2) into $r(\hat{\theta}_n) = \bar{R}(\hat{\theta}_n - \tilde{\theta}_n)$ gives $\tilde{\lambda}_n = n\{\bar{R}\bar{J}_n^{-1}\tilde{R}^\top\}^{-1}r(\hat{\theta}_n)$ so (A.2) can be written as

$$\hat{\theta}_n - \tilde{\theta}_n = \bar{J}_n^{-1}\tilde{R}^\top\{\bar{R}\bar{J}_n^{-1}\tilde{R}^\top\}^{-1}r(\hat{\theta}_n). \quad \textbf{(A.3)}$$

Since $q(\tilde{\theta}_n) = 0$, a mean value expansion of $q(\hat{\theta}_n)$ at $\tilde{\theta}_n$ gives $q(\hat{\theta}_n) = \bar{Q}(\hat{\theta}_n - \tilde{\theta}_n)$ where $\bar{Q}$ is the matrix $Q(\theta)$ with each of its rows evaluated at a (possibly different) mean value given by a convex combination of $\hat{\theta}_n$ and $\tilde{\theta}_n$. Finally, substituting (A.3) into $q(\hat{\theta}_n) = \bar{Q}(\hat{\theta}_n - \tilde{\theta}_n)$ gives (6) where

$$\bar{P}_n = \bar{Q}\bar{J}_n^{-1}\tilde{R}^\top\{\bar{R}\bar{J}_n^{-1}\tilde{R}^\top\}^{-1} \quad \textbf{(A.4)}$$

is an $r \times r$ nonsingular matrix (assuming that $\bar{Q}\bar{J}_n^{-1}\tilde{R}^\top$ is nonsingular) and $\bar{P}_n \xrightarrow{p} P_0$ as $\bar{Q} \xrightarrow{p} Q_0 = P_0R_0$, $\bar{J}_n \xrightarrow{p} J_0$, $\tilde{R} \xrightarrow{p} R_0$, and $\bar{R} \xrightarrow{p} R_0$.

Proof of the equations in (7). Since $q(\theta) = Pr(\theta)$ and $Q(\theta) = PR(\theta)$, the first three equations in (7) are obvious, and the last two equalities are easily seen as, in this special case, $Q_1(\theta_0) = PR_1(\theta_0)$ and $\bar{Q} = P\bar{R}$ so (A.1) and (A.4) reduce to $P_0 = P$ and $\bar{P}_n = P$, respectively.

Proof of Equation (8). It will be shown that $\mathscr{EW}^* \subseteq \mathscr{EW}$ and $\mathscr{EW} \subseteq \mathscr{EW}^*$, which imply (8); throughout this proof, $R_n \in \mathscr{R}$, $Q_n \in \mathscr{Q}$, and $J_n \in \mathscr{J}$. Let $R_{Qn} = \bar{P}_n^{-1}Q_n$ where $\bar{P}_n$ is as in (6). Then, R_{Qn} has rank r with $R_{Qn} \xrightarrow{p} P_0^{-1}Q_0 = R_0$ as $Q_0 = P_0R_0$. Therefore, $R_{Qn} \in \mathscr{R}$ so $W_n(J_n, R_{Qn}) \in \mathscr{EW}$. Now, (2), (4), and (6) show that

$$W_n(J_n, R_{Qn}) = W_n^*(J_n, Q_n), \quad \textbf{(A.5)}$$

which provides $\mathscr{EW}^* \subseteq \mathscr{EW}$ as $W_n^*(J_n, Q_n)$ is an arbitrary member of $\mathscr{EW}^*$. Similarly, let $Q_{Rn} = \bar{P}_nR_n$. Then, $Q_{Rn} \in \mathscr{Q}$ so $W_n^*(J_n, Q_{Rn}) \in \mathscr{EW}^*$. Here, (2), (4), and (6) show that $W_n^*(J_n, Q_{Rn}) = W_n(J_n, R_n)$ so $\mathscr{EW} \subseteq \mathscr{EW}^*$ as $W_n(J_n, R_n)$ is an arbitrary member of $\mathscr{EW}$. Hence, $\mathscr{EW} = \mathscr{EW}^*$.

The equality in (8) can also be obtained by showing that there exist two matrices $J_{Qn} \in \mathscr{J}$ and $J_{Rn} \in \mathscr{J}$ such that $W_n(J_{Qn}, R_n) = W_n^*(J_n, Q_n)$ and $W_n^*(J_{Rn}, Q_n) = W_n(J_n, R_n)$. To see this, let R_{Qn} be as above and $M_{Rn} = I_p - R_n^\top\{R_n J_n^{-1} R_n^\top\}^{-1} R_n J_n^{-1}$. Then,

$$D_n = J_n^{-1} R_n^\top \{R_n J_n^{-1} R_n^\top\}^{-1} R_{Qn} J_n^{-1} R_{Qn}^\top \{R_n J_n^{-1} R_n^\top\}^{-1} R_n J_n^{-1} + J_n^{-1} M_{Rn} \qquad \textbf{(A.6)}$$

is a symmetric matrix such that $D_n \xrightarrow{p} J_0^{-1}$. A proof by contradiction shows that D_n^{-1} exists. Therefore, suppose that D_n is singular. Then, there exists a $p \times 1$ vector $\xi \neq 0$ such that $D_n\xi = 0$, which (using (A.6)) can be written as

$$J_n^{-1} R_n^\top \{R_n J_n^{-1} R_n^\top\}^{-1} R_{Qn} J_n^{-1} R_{Qn}^\top \{R_n J_n^{-1} R_n^\top\}^{-1} R_n J_n^{-1}\xi + J_n^{-1} M_{Rn}\xi = 0. \qquad \textbf{(A.7)}$$

Since $\{R_{Qn} J_n^{-1} R_{Qn}^\top\}^{-1}$ exists, premultiplying (A.7) by $R_n J_n^{-1} R_n^\top \{R_{Qn} J_n^{-1} R_{Qn}^\top\}^{-1} R_n$ gives $R_n J_n^{-1}\xi = 0$ (which implies $M_{Rn}\xi = \xi$) so (A.7) reduces to $J_n^{-1}\xi = 0$, which provides the contradiction that $\xi = 0$. Hence, D_n is a symmetric nonsingular matrix. Now, let $J_{Qn} = D_n^{-1}$ and note that (A.6) gives $R_n J_{Qn}^{-1} R_n^\top = R_{Qn} J_n^{-1} R_{Qn}^\top$. Then, $J_{Qn} \in \mathscr{J}$ and, using (2), it is easily seen that $W_n(J_{Qn}, R_n) = W_n(J_n, R_{Qn})$ so, given (A.5),

$$W_n^*(J_n, Q_n) = W_n(J_n, R_{Qn}) = W_n(J_{Qn}, R_n). \qquad \textbf{(A.8)}$$

Similarly, it can be shown that $W_n(J_n, R_n) = W_n^*(J_n, Q_{Rn}) = W_n^*(J_{Rn}, Q_n)$ where Q_{Rn} is as above and

$$J_{Rn} = \left[J_n^{-1} Q_n^\top \{Q_n J_n^{-1} Q_n^\top\}^{-1} Q_{Rn} J_n^{-1} Q_{Rn}^\top \{Q_n J_n^{-1} Q_n^\top\}^{-1} Q_n J_n^{-1} + J_n^{-1} M_{Qn}\right]^{-1} \in \mathscr{J}$$

with $M_{Qn} = I_p - Q_n^\top \{Q_n J_n^{-1} Q_n^\top\}^{-1} Q_n J_n^{-1}$.

Proof of Equation (9). Let $R^* = \bar{P}_n^{-1}\hat{Q} \in \mathscr{R}$ and

$$J_n^* = \left[\hat{J}_n^{-1} \hat{R}^\top \{\hat{R} \hat{J}_n^{-1} \hat{R}^\top\}^{-1} R^* \hat{J}_n^{-1} R^{*\top} \{\hat{R} \hat{J}_n^{-1} \hat{R}^\top\}^{-1} \hat{R} \hat{J}_n^{-1} + \hat{J}_n^{-1} \hat{M}_{Rn}\right]^{-1} \in \mathscr{J}$$

where $\hat{M}_{Rn} = I_p - \hat{R}^\top \{\hat{R} \hat{J}_n^{-1} \hat{R}^\top\}^{-1} \hat{R} \hat{J}_n^{-1}$; i.e., R^*, J_n^*, and $\hat{M}_{Rn}$ are special cases of R_{Qn}, J_{Qn}, and M_{Rn}, respectively, obtained by setting $J_n = \hat{J}_n$, $R_n = \hat{R}$, and $Q_n = \hat{Q}$. Then, (9) is obtained from (A.8) by noting that $W_n^*(\hat{J}_n, \hat{Q}) = W_n^*(\hat{J}_n)$.

REFERENCES

Breusch, T.S., and P. Schmidt (1988) "Alternative forms of the Wald test: How long is a piece of string?" *Communications in Statistics - Theory and Methods* **17**, 2789-2795.

Critchley, F., P. Marriott, and M. Salmon (1996) "On the differential geometry of the Wald test with nonlinear restrictions" *Econometrica* **64**, 1213-1222.

Dagenais, M.G., and J.-M. Dufour (1991) "Invariance, nonlinear models, and asymptotic tests" *Econometrica* **59**, 1601-1615.

Dastoor, N.K. (2003) "The equality of comparable extended families of classical-type and Hausman-type statistics" *Journal of Econometrics* **117**, 313-330.

Davidson, R. (1990) "The geometry of the Wald test" Discussion Paper 800, Queen's University.

Davidson, R., and J.G. MacKinnon (1993) *Estimation and Inference in Econometrics*, Oxford University Press: New York.

Gourieroux, C., and A. Monfort (1989) "A general framework for testing a null hypothesis in a 'mixed' form" *Econometric Theory* **5**, 63-82.

Gregory, A.W., and M.R. Veall (1985) "Formulating Wald tests of nonlinear restrictions" *Econometrica* **53**, 1465-1468.

Kemp, G.C.R. (2001) "Invariance and the Wald test" *Journal of Econometrics* **104**, 209-217.

Lafontaine, F., and K.J. White (1986) "Obtaining any Wald statistic you want" *Economics Letters* **21**, 35-40.

Newey, W.K., and D. McFadden (1994) "Large sample estimation and hypothesis testing" in *Handbook of Econometrics, Volume IV* by R.F. Engle and D.L. McFadden, Eds., North-Holland: Amsterdam, 2111-2245.

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