DEPARTMENT OF
ECONOMICS
UNIVERSITY OF ALBERTA

Working Paper No. 2009-25

The perceived framework of a classical statistic: Is the non-invariance of a Wald statistic much ado about null thing?

Naorayex K. Dastoor
University of Alberta

August, 2009

The perceived framework of a classical statistic: Is the non-invariance of a Wald statistic much ado about null thing?

Naorayex K Dastoor[1]

University of Alberta

July 2006

(August 2009: Colours added to Figures 1 and 2)

Abstract

The distinction between a nominal framework for the three classical statistics and a perceived framework for each classical statistic provides more ways to interpret these statistics, and intuitively explains as well as more easily shows some well-known results. In particular, each classical statistic can be viewed in terms of a length in each of four spaces and, since the classical procedures per se are equivalent in a perceived framework, two statistics are identical if their perceived frameworks are identical. This helps to integrate the normally separately treated issues of a reformulation of a null hypothesis and of locally equivalent alternatives. For example, a Wald statistic is not invariant if a reformulation changes its perceived framework, and an appropriate score statistic is invariant as its perceived framework is unaffected by considering a locally equivalent alternative. [During the thirty-four months this paper was under consideration at *The Econometrics Journal*, the Editor-in-charge (Professor Stéphane Gregoir) did not reply to three (of the author's four) requests about the status of the submission, and provided neither a referee's report nor a first decision. Also, when asked to intervene by the author, the new Managing Editor (Professor Richard J Smith) offered the author the possibility of submitting the paper (as a new submission) to the new editorial regime, at which point, the author withdrew the paper.]

Keywords: Classical statistic, Likelihood ratio statistic, Nominal framework, Perceived framework, Score statistic, Wald statistic.

JEL classification: C12

[1] *Address:* Department of Economics, University of Alberta, Edmonton, Alberta T6G 2H4, Canada

Email: noxy.dastoor@ualberta.ca

Chronology of Submission to *The Econometrics Journal*

Summary: During the thirty-four months this paper was under consideration at *The Econometrics Journal*, the Editor-in-charge (Professor Stéphane Gregoir) did not reply to three (of the author's four) requests about the status of the submission, and provided neither a referee's report nor a first decision. Also, when asked to intervene by the author, the new Managing Editor (Professor Richard J Smith) offered the author the possibility of submitting the paper (as a new submission) to the new editorial regime, at which point, the author withdrew the paper.

Note: Below, in the left margin, a number in square brackets denotes the approximate number of months elapsed since submission of the paper.

31 July 2006: Paper submitted to *The Econometrics Journal*.

1 August 2006: Editorial Assistant informs author that the paper has been passed on to one of the Editors, Professor Stéphane Gregoir.

24 August 2006: Professor Gregoir acknowledges receipt of paper.

[8] **4 April 2007:** Author requests status of submission from Professor Gregoir, no reply received.

[9] **9 May 2007:** Author requests status of submission from Professor Gregoir.

[10] **25 May 2007:** Professor Gregoir replies that he will get in touch as soon as he has got the referees' reports.

[17] **7 January 2008:** Author requests status of submission from Professor Gregoir, no reply received.

[18] **17 February 2008:** Author requests status of submission from Professor Gregoir, no reply received.

[20] **31 March 2008:** Author contacts the new Managing Editor, Professor Richard J Smith, expressing concern about (and requesting status of) submission.

[20] **1 April 2008:** Professor Smith replies that he will make enquiries regarding the status of the paper and to contact him if, in two weeks, no response is received from either the editorial office or Professor Gregoir.

[20] **7 April 2008:** Professor Gregoir informs author that he has received only one referee's report, got in touch with three other potential referees but none delivered a report, has asked a fourth one, and hopes to be able to get back with some news soon.

[33] **3 May 2009:** Author requests Professor Richard J Smith to take over the processing of the submission.

[33] **4 May 2009:** Professor Smith replies that he has contacted Professor Gregoir and instructs the author to contact Professor Smith if nothing is heard within two weeks.

[34] **25 May 2009:** Author informs Professor Smith about not hearing anything.

[34] **2 June 2009:** Professor Smith writes, "In the circumstances given the length of time since the paper was submitted I would like to offer you the possibility to submit your paper to The Econometrics Journal under the new editorial regime; ... I should emphasise that I cannot give you any guarantee of eventual success as the paper would need to be treated as a new submission."

[35] **29 June 2009:** Author withdraws submission.

1. Introduction

In a model specified by a likelihood function where the usual regularity conditions hold, the validity of a set of restrictions (on a vector of unknown parameters) is commonly tested by using the classical statistics, namely, the likelihood ratio, a score (or Lagrange Multiplier), and a Wald statistic. Numerous results on the behaviour of these statistics are well known and presented in most graduate texts; for example, see Davidson and MacKinnon (1993, 2004). This paper presents a novel/different approach to examining the behaviour of these statistics by distinguishing between a nominal framework for all three statistics and a perceived framework for each statistic. In general, a (testing) framework is defined by three quantities: a parameter space, a maximand that provides an estimate for an unknown vector of parameters, and a restricted parameter space defined by a set of restrictions on the vector of parameters. In particular, in the nominal framework common to all three classical statistics, the parameter space is Ω, the maximand is a log-likelihood function $L(\theta)$ where θ is a vector of unknown parameters, and the restricted parameter space is defined by a vector of restrictions $r(\theta) = 0$. The standard asymptotic theory, when applied to the classical statistics, uses Taylor series approximations (or mean value expansions) of $L(\theta)$, the score function $g(\theta)$, and $r(\theta)$; for example, see Davidson and MacKinnon (1993, pp. 446-7). These approximations do not only provide the asymptotic distribution of a statistic, but also play a critical role in determining the form of a statistic. For example, these approximations provide the variance-covariance matrix of the asymptotic distribution of $n^{-1/2}g(\tilde{\theta})$ where n is the sample size and $\tilde{\theta}$ is the restricted maximum likelihood (ML) estimator of θ. Then, a score statistic is constructed as a quadratic form in $n^{-1/2}g(\tilde{\theta})$ where the weighting matrix is a consistent estimator of a generalized inverse of this variance-covariance matrix. Similarly, these approximations provide the factor 2 in the likelihood ratio statistic and a Wald statistic as an appropriate quadratic form in $\sqrt{n}r(\hat{\theta})$ where $\hat{\theta}$ is the unrestricted ML estimator of θ.

The perceived framework of a statistic is constructed such that it explicitly accounts for the approximations used when the standard asymptotic theory is applied to the classical statistics. Therefore, in the perceived framework of each classical statistic, the parameter space is Θ with $\Omega \subseteq \Theta$, the maximand is a quadratic approximation of $L(\theta)$ (which gives a linear approximation of $g(\theta)$), and the restricted parameter space is defined by a linear approximation of $r(\theta)$ being zero. In the perceived framework of the likelihood ratio statistic, the quadratic approximation of $L(\theta)$ is at $\hat{\theta}$ and the linear approximation of $r(\theta)$ is at $\tilde{\theta}$ and, in the perceived framework of a score (Wald) statistic, both approximations are at $\tilde{\theta}$ ($\hat{\theta}$). For each statistic, the particular approximations are chosen such that (for a given sample) the statistic is reproduced by applying the corresponding classical procedure to its perceived framework. Hence, notionally, a perceived framework is that as seen by a statistic for a given sample, whereas, the nominal framework is that as intended for the statistics.

Quadratic approximations of $L(\theta)$ have been used to argue that a score and a Wald statistic can be viewed as approximations to the likelihood ratio statistic. For example, consider Figure 3 in Newey and McFadden (1994, p. 2221) where $Q_n(\theta)$, DM_n, LM_n, and W_n correspond to $L(\theta)$, the likelihood ratio, a score, and a Wald statistic, respectively. Then, this figure shows that (for a scalar parameter with a linear restriction) a score and a Wald statistic are approximations to the likelihood ratio statistic in the sense that they can be reproduced by applying the likelihood ratio procedure to appropriate quadratic approximations of $L(\theta)$. In this example, a quadratic approximation of $L(\theta)$ for the likelihood ratio statistic is not considered as the focus is on the likelihood ratio procedure. However, there is no need to focus solely on the likelihood ratio procedure. Also, in general, a Wald statistic cannot be reproduced by considering only a quadratic approximation of $L(\theta)$; below, another figure will show the role played by a linear approximation of $r(\theta)$ in reproducing a Wald statistic for a nonlinear restriction. Therefore, by considering both a quadratic approximation of $L(\theta)$ and a linear approximation of $r(\theta)$ for each statistic, a perceived framework does not focus attention on any particular classical procedure and provides a more complete framework for examining the behaviour of a statistic.

The analysis with a perceived framework provides two general results. First, in a perceived framework, the classical procedures per se are equivalent so a difference between two statistics arises as a result of applying the equivalent procedures to different perceived frameworks, and (as expected, given the notion of a perceived framework) two statistics are identical if their perceived frameworks are identical. The second general result is an extension of two popular interpretations of the three statistics. On the one hand, based on Buse (1982) and Engle (1984), each statistic can be viewed in terms of a length in the space associated with the statistic. The likelihood ratio statistic is a length in the range of $L(\theta)$ and a score (Wald) statistic is a squared length in the range of $g(\theta)$ ($r(\theta)$). On the other hand, based on Newey and McFadden (1994), all three statistics can be viewed in terms of lengths in the range of $L(\theta)$. Using a perceived framework, each statistic can be interpreted in terms of a length in each of four possible spaces; the parameter space Θ and the spaces that contain the ranges of $L(\theta)$, $g(\theta)$, and $r(\theta)$. The interpretation of a classical statistic in terms of a length in a parameter space does not appear to have been considered in the literature. With this interpretation, the analysis of the classical statistics is simplified so some well-known results are more easily seen. These general results help to shed further light in some special cases. For example, if $L(\theta)$ is a quadratic function of θ and $r(\theta)$ is a linear function of θ, then the equality of the likelihood ratio statistic and appropriate versions of a score and a Wald statistic is intuitively explained by the fact that their perceived frameworks are identical, and it is seen that a quadratic $L(\theta)$ per se is sufficient for the equality of only the likelihood ratio and a score statistic. The general results are also applicable to classical-type statistics; those obtained by replacing a log-likelihood function with an appropriate maximand associated with some other method of estimation where a type of information matrix equality holds, for example, as in Gourieroux and Monfort (1989, Section 2.2).

Certain properties of a Wald and a score statistic are well-known. First, in general, a Wald statistic is not invariant to a reformulation of a nominal null hypothesis where $r(\theta) = 0$ is rewritten in an algebraically equivalent form; for the behaviour of the statistics to different types of invariance, see Dagenais and Dufour (1991). Numerous papers have examined various aspects of this non-invariance. For example, Gregory and Veall (1985) provide Monte Carlo evidence of the effect of a reformulation, Lafontaine and White (1986) and Breusch and Schmidt (1988) show how this non-invariance could be exploited to obtain a desired numerical value for a Wald statistic, Phillips and Park (1988) examine the effect of a reformulation on the small sample distribution of a Wald statistic, Davidson (1990) and Critchley, Marriott and Salmon (1996) apply the methods of differential geometry to explain the non-invariance, and Kemp (2001) provides a justification for ruling out certain reformulations. Second, a score statistic (evaluated appropriately) is invariant to certain alternative hypotheses. In a linear regression model, this result was first shown by Breusch (1978) and Godfrey (1978) in the case of testing for white noise error terms against the alternative of either autoregressive or moving average errors of the same order. Subsequently, Godfrey (1981) and Godfrey and Wickens (1982) have shown the invariance of an appropriate score statistic to what they define as a locally equivalent alternative (LEA).

In the literature, the analysis for a reformulation of a nominal null has been treated as a separate issue from the analysis with an LEA. This paper helps to integrate these two analyses by explaining the invariance or non-invariance properties of a statistic with respect to either a reformulation of a nominal null or an LEA. In particular, the likelihood ratio and a score statistic are invariant to a reformulation of a nominal null as their perceived frameworks are unaffected by a reformulation, whereas, a Wald statistic is not invariant if a reformulation changes its perceived framework. The analysis with respect to an LEA considers another nominal framework where (ceteris paribus) the maximand is a log-likelihood function $L^*(\theta)$ with the score vector $g^*(\theta)$. Using the definition of Godfrey and Wickens (1982), this other nominal framework is an LEA when, inter alia, $g(\tilde{\theta}) = g^*(\tilde{\theta})$ and $\tilde{\theta}$ is the restricted estimator of θ in both nominal frameworks. Davidson and MacKinnon (1993) have argued that $g(\tilde{\theta}) = g^*(\tilde{\theta})$ is too strong a requirement as an appropriate score statistic is also invariant if $g^*(\tilde{\theta})$ is a nonsingular linear transformation of $g(\tilde{\theta})$. Therefore, it is first shown that the use of the definition of Godfrey and Wickens (1982) is not as restrictive as may first appear, provided $L^*(\theta)$ is appropriately specified. Then, it is easily seen that an appropriate score statistic is invariant as its perceived framework is unaffected by considering an LEA, whereas, the likelihood ratio and a Wald statistic are not invariant as their perceived frameworks are affected by the use of an LEA.

The rest of the paper proceeds as follows. In the next section, the three statistics are first presented in their usual forms for a given nominal framework. Then, a generic perceived framework is constructed such that it provides as special cases the perceived frameworks of the likelihood ratio, a score, and a Wald statistic. Section 3 shows how each statistic can be expressed in terms of a length in each of four spaces and presents some further results. Section 4 provides the analysis with respect to a reformulation of a nominal null and discusses certain aspects of the non-invariance of a Wald statistic. Results with respect to an LEA are presented in Section 5 and, finally, some concluding remarks are stated in Section 6.

2. Nominal and perceived frameworks

The nominal framework common to all three classical statistics is defined by the triplet $(\Omega, L(\theta), \Omega_0)$ where the parameter space $\Omega \subseteq \mathbb{R}^p$, the maximand $L(\theta)$ is a log-likelihood function for n observations, $\theta \in \Omega$ is a $p \times 1$ vector of unknown parameters, and the restricted parameter space $\Omega_0 = \{\theta \,|\, r(\theta) = 0,\, \theta \in \Omega\}$ with $r(\theta) = 0$ being an $r \times 1$ vector of known restrictions and $r \leq p$. For notational simplicity, the dependence of quantities on the data and on n will be suppressed. The null and alternative hypotheses are $H_0 : \theta \in \Omega_0$ and $H_1 : \theta \in \Omega_1$, respectively, where Ω_0 and Ω_1 constitute a partition of Ω. Then, the unrestricted and restricted ML estimators of θ are $\hat{\theta} = \text{argmax}_{\theta\in\Omega} L(\theta)$ and $\tilde{\theta} = \text{argmax}_{\theta\in\Omega_0} L(\theta)$, respectively. Let $g(\theta) = \partial L(\theta)/\partial\theta$, $H(\theta) = -n^{-1}\partial^2 L(\theta)/\partial\theta\partial\theta^\top$, $\hat{H} = H(\hat{\theta})$, $R(\theta) = \partial r(\theta)/\partial\theta^\top$ (an $r \times p$ matrix), $\hat{R} = R(\hat{\theta})$, $\tilde{R} = R(\tilde{\theta})$, $J(\theta)$ be a $p \times p$ symmetric matrix, $\hat{J} = J(\hat{\theta})$, and $\tilde{J} = J(\tilde{\theta})$. Throughout, asymptotic results are obtained under H_0 and the following two assumptions hold.

Assumption 1. For a given sample: (a) $\hat{\theta} \in \Omega_1$, $g(\hat{\theta}) = 0$, and $L(\hat{\theta}) > L(\tilde{\theta})$; (b) $\hat{R}$ and $\tilde{R}$ have rank r; and (c) $\hat{H}$, $\hat{J}$, and $\tilde{J}$ are positive definite matrices.

Assumption 2. Under H_0: (a) $\hat{\theta} \xrightarrow{p} \theta_0$ and $\tilde{\theta} \xrightarrow{p} \theta_0$ where $\theta_0 \in \Omega_0$; (b) $\hat{J} \xrightarrow{p} J_0$ and $\tilde{J} \xrightarrow{p} J_0$ where J_0 is the (positive definite) limiting information matrix; (c) $H(\breve{\theta}) \xrightarrow{p} J_0$ and $R(\breve{\theta}) \xrightarrow{p} R_0$ where $\breve{\theta} \xrightarrow{p} \theta_0$ and $R_0 = R(\theta_0)$ has rank r; and (d) $\sqrt{n}(\hat{\theta} - \tilde{\theta}) \overset{a}{\sim} N(0,\, J_0^{-1}R_0^\top V_0^{-1}R_0 J_0^{-1})$, $n^{-1/2}g(\tilde{\theta}) \overset{a}{\sim} N(0,\, R_0^\top V_0^{-1} R_0)$, and $\sqrt{n}r(\hat{\theta}) \overset{a}{\sim} N(0,\, V_0)$ where $V_0 = R_0 J_0^{-1} R_0^\top$.

The ensuing analysis will present algebraic results for a given sample so Assumption 1 will ensure that all sample quantities are well defined and have appropriate properties. Assumption 2 simply states relevant standard asymptotic results obtained under H_0 and under the usual regularity conditions where the (limiting) information matrix equality holds. Rigorous statements of the appropriate conditions required and formal derivations of these standard asymptotic results can be found in, for example, Davidson and MacKinnon (1993) and Newey and McFadden (1994).

Consider the likelihood ratio, a score, and a Wald statistic given by

$$LR = 2\{L(\hat{\theta}) - L(\tilde{\theta})\}, \tag{1}$$

$$S = \frac{1}{n} g(\tilde{\theta})^\top \tilde{J}^{-1} g(\tilde{\theta}), \tag{2}$$

and

$$W = nr(\hat{\theta})^\top \{\hat{R}\hat{J}^{-1}\hat{R}^\top\}^{-1} r(\hat{\theta}), \tag{3}$$

respectively. Under H_0, each of these statistics is asymptotically distributed as a $\chi^2(r)$ variate. Since the Lagrangean associated with $\tilde{\theta}$ is $\mathcal{L}(\theta, \lambda) = L(\theta) - \lambda^\top r(\theta)$ where λ is an $r \times 1$ vector of Lagrange multipliers, the first-order condition $\partial \mathcal{L}(\tilde{\theta}, \tilde{\lambda})/\partial\theta = 0$ gives

$$g(\tilde{\theta}) = \tilde{R}^\top \tilde{\lambda} \tag{4}$$

so (2) is often written as $S = n^{-1}\tilde{\lambda}^\top \tilde{R}\tilde{J}^{-1}\tilde{R}^\top\tilde{\lambda}$, a Lagrange Multiplier statistic. In the case where $L(\theta)$ is a quadratic function of θ (such that $H(\theta) = \bar{H}$ is a positive definite nonstochastic matrix with $\lim_{n\to\infty} \bar{H} = J_0$) and $r(\theta)$ is a linear function of θ, Buse (1982, p. 156) and Engle (1984, pp. 782-4) showed that

$$LR = S = W \tag{5}$$

where S and W are evaluated with $\tilde{J} = \bar{H}$ and $\hat{J} = \bar{H}$, respectively.

The perceived frameworks of LR, S, and W can be obtained as special cases of a generic perceived framework as follows. Let $C = LR, S, W$. Then, the perceived framework of C is defined for a given sample by the triplet $(\Theta, L_C(\theta), \Theta_{0C})$ where the parameter space $\Theta = \mathbb{R}^p$ with $\Omega \subseteq \Theta$, the maximand $L_C(\theta)$ is a quadratic approximation of $L(\theta)$ at a chosen point $\bar{\theta}_C$ such that $\bar{\theta}_C \xrightarrow{p} \theta_0$, and the restricted parameter space Θ_{0C} is obtained by replacing $r(\theta)$ in Ω_0 with $r_C(\theta)$, a linear approximation of $r(\theta)$ at a chosen point $\ddot{\theta}_C$ such that $\ddot{\theta}_C \xrightarrow{p} \theta_0$. Formally,

$$L_C(\theta) = L(\bar{\theta}_C) + g(\bar{\theta}_C)^\top(\theta - \bar{\theta}_C) - \frac{n}{2}(\theta - \bar{\theta}_C)^\top J_C(\theta - \bar{\theta}_C), \tag{6}$$

$$r_C(\theta) = r(\ddot{\theta}_C) + R_C(\theta - \ddot{\theta}_C), \tag{7}$$

and

$$\Theta_{0C} = \{\theta \,|\, r_C(\theta) = 0, \, \theta \in \Theta\} \tag{8}$$

where J_C is a symmetric positive definite matrix such that $J_C \xrightarrow{p} J_0$, and R_C is an $r \times p$ matrix with rank r such that $R_C \xrightarrow{p} R_0$. Basically, $L_C(\theta)$ is a second-order Taylor series approximation of $L(\theta)$ at $\bar{\theta}_C$ with $H(\bar{\theta}_C)$ replaced by J_C to cater for different estimators of J_0. Similarly, $r_C(\theta)$ is a first-order Taylor series approximation of $r(\theta)$ at $\ddot{\theta}_C$ with $R(\ddot{\theta}_C)$ replaced by R_C to cater for different estimators of R_0. Here, the unrestricted and restricted estimators of θ are $\hat{\theta}_C = \text{argmax}_{\theta\in\Theta} L_C(\theta)$ and $\tilde{\theta}_C = \text{argmax}_{\theta\in\Theta_{0C}} L_C(\theta)$, respectively. Note that, throughout, a 'hat' ('tilde') on θ is used for an unrestricted (restricted) estimator, and the subscript C will indicate a quantity from a perceived framework where, unless stated otherwise, $C = LR, S, W$. Let $g_C(\theta) = \partial L_C(\theta)/\partial\theta$. Then, it can be shown that

$$g_C(\theta) = g(\bar{\theta}_C) - nJ_C(\theta - \bar{\theta}_C), \tag{9}$$

$$\hat{\theta}_C = \bar{\theta}_C + \frac{1}{n}J_C^{-1}g(\bar{\theta}_C), \tag{10}$$

$$\tilde{\theta}_C = \hat{\theta}_C - J_C^{-1} R_C^\top \{R_C J_C^{-1} R_C^\top\}^{-1} r_C(\hat{\theta}_C), \tag{11}$$

and

$$g_C(\tilde{\theta}_C) = n R_C^\top \{R_C J_C^{-1} R_C^\top\}^{-1} r_C(\hat{\theta}_C) = n J_C(\hat{\theta}_C - \tilde{\theta}_C) \tag{12}$$

where $\hat{\theta}_C \xrightarrow{p} \theta_0$ and $\tilde{\theta}_C \xrightarrow{p} \theta_0$.

Since the special cases of $L_C(\theta)$ and $r_C(\theta)$ are determined by appropriately choosing the sample quantities $\bar{\theta}_C$, J_C, $\ddot{\theta}_C$, and R_C, it is convenient to collect these quantities as the quadruplet

$$\mathcal{P}_C \equiv (\bar{\theta}_C, J_C, \ddot{\theta}_C, R_C). \tag{13}$$

Initially, $\mathcal{P}_C$ is treated as fixed for the purpose of evaluating a statistic and random variables are obtained in the usual manner by later accounting for repeated sampling; cf. a ML estimator is obtained by first evaluating a ML estimate and later accounting for repeated sampling. Therefore, in the perceived framework of C, the likelihood ratio, a score, and a Wald statistic are

$$LR(\mathcal{P}_C) = 2\{L_C(\hat{\theta}_C) - L_C(\tilde{\theta}_C)\}, \tag{14}$$

$$S(\mathcal{P}_C) = \frac{1}{n} g_C(\tilde{\theta}_C)^\top J_C^{-1} g_C(\tilde{\theta}_C), \tag{15}$$

and

$$W(\mathcal{P}_C) = n r_C(\hat{\theta}_C)^\top \{R_C J_C^{-1} R_C^\top\}^{-1} r_C(\hat{\theta}_C), \tag{16}$$

respectively. Then, the perceived frameworks of LR, S, and W are obtained by choosing $\mathcal{P}_C$ such that $C = C(\mathcal{P}_C)$ for a given sample; i.e., a statistic in the nominal framework is reproduced by a corresponding statistic in its perceived framework. This paper focusses on the classical statistics, but the generic perceived framework could also be used to determine the perceived frameworks of other asymptotic statistics such as classical-type, $C(\alpha)$, and Hausman statistics. For the classical statistics, the following proposition (proved in

Appendix A) provides the particular choices for the components of $\mathcal{P}_C$. Table 1 (on page 37) presents these choices as well as the special cases of (6), (7), and (9) to (11), which are easily seen except, perhaps, $\tilde{\theta}_{LR} = \tilde{\theta}$ and $\tilde{\theta}_S = \tilde{\theta}$ in the last row; the derivations of these two special cases are indicated in the proof of the proposition.

Proposition 1. *Let $\mathcal{P}_{LR} = (\hat{\theta}, b\hat{H}, \tilde{\theta}, \tilde{R}Z)$, $\mathcal{P}_S = (\tilde{\theta}, \tilde{J}, \tilde{\theta}, \tilde{R})$, and $\mathcal{P}_W = (\hat{\theta}, \hat{J}, \hat{\theta}, \hat{R})$ where b is a particular positive scalar such that $b \xrightarrow{p} 1$, and Z is a particular $p \times p$ nonsingular matrix such that $Z \xrightarrow{p} I_p$. Then, $LR = LR(\mathcal{P}_{LR})$, $S = S(\mathcal{P}_S)$, and $W = W(\mathcal{P}_W)$.*

This proposition (and its proof) shows that, in the perceived framework of LR, the maximand $L_{LR}(\theta)$ is a second-order Taylor's series approximation of $L(\theta)$ at $\hat{\theta}$ where the usual curvature provided by $\hat{H}$ is adjusted by the factor b to force the quadratic approximation to go through the point $(\tilde{\theta},\ L(\tilde{\theta})) \in \Theta \times \mathbb{R}$, and the restricted parameter space Θ_{0LR} is given by a particular linear approximation of $r(\theta)$ at $\tilde{\theta}$ such that $\tilde{\theta}_{LR} = \tilde{\theta}$. Therefore, $L_{LR}(\hat{\theta}_{LR}) = L(\hat{\theta})$, $L_{LR}(\tilde{\theta}_{LR}) = L(\tilde{\theta})$, and the unrestricted (restricted) estimator of θ in the perceived framework of LR coincides with the unrestricted (restricted) estimator of θ in the nominal framework. In the perceived framework of S (W), the maximand $L_S(\theta)$ ($L_W(\theta)$) is a quadratic approximation of $L(\theta)$ at $\tilde{\theta}$ ($\hat{\theta}$) with curvature provided by $\tilde{J}$ ($\hat{J}$), and the restricted parameter space Θ_{0S} (Θ_{0W}) is given by a first-order Taylor series approximation of $r(\theta)$ also at $\tilde{\theta}$ ($\hat{\theta}$) so the restricted (unrestricted) estimator of θ in the perceived framework of S (W) coincides with the restricted (unrestricted) estimator of θ in the nominal framework.

3. Lengths in $\mathfrak{L}$, $\mathfrak{G}$, $\mathfrak{R}$, and Θ

3.1. General results

Let $\mathfrak{L} = \mathbb{R}$, $\mathfrak{G} = \mathbb{R}^p$, and $\mathfrak{R} = \mathbb{R}^r$ denote the spaces that contain the ranges of the functions $L(\theta)$, $g(\theta)$, and $r(\theta)$, respectively. Then, the two popular interpretations mentioned above can be stated formally as follows. First, based on Buse (1982) and Engle (1984), LR is the length of $L(\hat{\theta}) - L(\tilde{\theta})$ in $\mathfrak{L}$ where a length is twice the Euclidean length, S is the squared length of $g(\tilde{\theta})$ in $\mathfrak{G}$ with the metric $(n\tilde{J})^{-1}$, and W is the squared length of $r(\hat{\theta})$ in $\mathfrak{R}$ with the metric $n\{\hat{R}\hat{J}^{-1}\hat{R}^\top\}^{-1}$; see Greene (2003, Figure 17.2, p. 485). Second, based on Newey and McFadden (1994), all three statistics can be viewed as lengths in $\mathfrak{L}$ where, again, a length is twice the Euclidean length. Now, in a perceived framework, $L_C(\theta)$ is a quadratic function of θ and $r_C(\theta)$ is a linear function of θ. Therefore, by analogy with (5),

$$LR(\mathcal{P}_C) = S(\mathcal{P}_C) = W(\mathcal{P}_C),$$

which, using $C = C(\mathcal{P}_C)$ and (12), can be extended to

$$C = LR(\mathcal{P}_C) = S(\mathcal{P}_C) = W(\mathcal{P}_C) = n(\hat{\theta}_C - \tilde{\theta}_C)^\top J_C(\hat{\theta}_C - \tilde{\theta}_C) \quad \textbf{(17)}$$

where the last expression is a Hausman statistic in a perceived framework. By defining lengths and metrics in a given space, (17) will enable C to be expressed in terms of a length in each of the four spaces $\mathfrak{L}$, $\mathfrak{G}$, $\mathfrak{R}$, and Θ. In addition, (17) shows that the classical procedures per se are equivalent in a perceived framework, and any two of LR, S, and W are identical if their perceived frameworks are identical. Therefore, in finite samples, a difference between two statistics arises as a result of applying the equivalent procedures to different perceived frameworks and, asymptotically, the statistics are equivalent as the difference between their perceived frameworks vanishes.

As above, let the length of a scalar $\mathfrak{l} \in \mathfrak{L}$ be defined as twice its Euclidean length, and let the squared lengths of the vectors $\mathfrak{g} \in \mathfrak{G}$ and $\mathfrak{r} \in \mathfrak{R}$ be denoted as

$$||\mathfrak{g}||^2_{\mathfrak{G}} \equiv \frac{1}{n}\mathfrak{g}^\top \tilde{J}^{-1}\mathfrak{g} \qquad \text{and} \qquad ||\mathfrak{r}||^2_{\mathfrak{R}} \equiv n\mathfrak{r}^\top\{\hat{R}\hat{J}^{-1}\hat{R}^\top\}^{-1}\mathfrak{r}, \quad \textbf{(18)}$$

respectively. In $\mathfrak{G}$, the metric $(n\tilde{J})^{-1}$ is used just to reproduce the interpretation of S as the squared length of $g(\tilde{\theta})$. Similarly, the metric $n\{\hat{R}\hat{J}^{-1}\hat{R}^\top\}^{-1}$ in $\mathfrak{R}$ will reproduce the interpretation of W as the squared length of $r(\hat{\theta})$. For squared lengths in $\mathfrak{G}$ and $\mathfrak{R}$, the metrics will always be $(n\tilde{J})^{-1}$ and $n\{\hat{R}\hat{J}^{-1}\hat{R}^\top\}^{-1}$, respectively, so the subscripts $\mathfrak{G}$ and $\mathfrak{R}$ in (18) identify the spaces being considered. For lengths in Θ, it is convenient to allow for different metrics for different statistics. Therefore, in Θ,

$$\left|\left|\hat{\theta}_C - \tilde{\theta}_C\right|\right|^2_{nJ_C} \equiv n(\hat{\theta}_C - \tilde{\theta}_C)^\top J_C(\hat{\theta}_C - \tilde{\theta}_C) \quad \textbf{(19)}$$

will denote the squared length of the vector $\hat{\theta}_C - \tilde{\theta}_C$ where the subscript nJ_C specifies the metric being used. Given Assumption 1, the estimates of J_0 and R_0 considered above have the same (positive definiteness and rank r) properties as those of J_0 and R_0, respectively, so the metrics in $\mathfrak{G}$, $\mathfrak{R}$, and Θ are well defined in the sense of being positive definite matrices. In practice, this may not be the case. For example, when S is evaluated with $\tilde{J} = H(\tilde{\theta})$, Dagenais and Dufour (1991, Table 1, p. 1611) provide cases where $S < 0$ as $H(\tilde{\theta})$ is not positive definite for the given sample. To cater for such cases, a perceived framework and the metrics in $\mathfrak{G}$, $\mathfrak{R}$, and Θ could be defined by first using J_0 and R_0 instead of their estimates. This would ensure that the resulting statistics from such a perceived framework could be viewed as squared lengths where the metrics are well defined. Then, J_0 and R_0 in these resulting statistics could be replaced with appropriate estimates to provide a classical statistic, which could be interpreted as a 'feasible' squared length with any problems arising (such as $S < 0$ for a given sample) viewed as a shortcoming of the estimates used. For the case considered here, the following proposition (proved in Appendix B) provides the expressions for each classical statistic in terms of a length in each of the four spaces.

Proposition 2. *In terms of a length in each of $\mathfrak{L}$, $\mathfrak{G}$, $\mathfrak{R}$, and Θ, the special cases of (17) are*

$$LR = 2\{L_{LR}(\hat{\theta}) - L_{LR}(\tilde{\theta})\} = \left|\left|A_{LR}g(\tilde{\theta})\right|\right|^2_{\mathfrak{G}} = \left|\left|B_{LR}r(\hat{\theta})\right|\right|^2_{\mathfrak{R}} = \left|\left|\hat{\theta} - \tilde{\theta}\right|\right|^2_{nJ_{LR}}, \quad \textbf{(20)}$$

$$S = 2\{L_S(\hat{\theta}_S) - L_S(\tilde{\theta})\} = \left|\left|g(\tilde{\theta})\right|\right|^2_{\mathfrak{G}} = \left|\left|B_S r(\hat{\theta})\right|\right|^2_{\mathfrak{R}} = \left|\left|\hat{\theta}_S - \tilde{\theta}\right|\right|^2_{n\tilde{J}}, \quad \textbf{(21)}$$

and

$$W = 2\{L_W(\hat{\theta}) - L_W(\tilde{\theta}_W)\} = \left|\left|A_W g(\tilde{\theta})\right|\right|^2_{\mathfrak{G}} = \left|\left|r(\hat{\theta})\right|\right|^2_{\mathfrak{R}} = \left|\left|\hat{\theta} - \tilde{\theta}_W\right|\right|^2_{n\hat{J}} \quad \textbf{(22)}$$

where (for $C = LR, W$) A_C is a particular $p \times p$ nonsingular matrix such that $A_C \xrightarrow{p} I_p$, and (for $C = LR, S$) B_C is a particular $r \times r$ nonsingular matrix such that $B_C \xrightarrow{p} I_r$.

With respect to viewing a classical statistic in terms of a length in a given space, this proposition provides all possible interpretations, including the two popular ones. In (20), the first equality reproduces the interpretation based on Buse (1982) and Engle (1984) of LR as twice the Euclidean length of $L(\hat{\theta}) - L(\tilde{\theta}) = L_{LR}(\hat{\theta}) - L_{LR}(\tilde{\theta})$ in $\mathfrak{L}$ and, given appropriate metrics, the remaining equalities enable LR to be viewed as squared lengths of the vectors $A_{LR}g(\tilde{\theta})$, $B_{LR}r(\hat{\theta})$, and $\hat{\theta} - \tilde{\theta}$ in $\mathfrak{G}$, $\mathfrak{R}$, and Θ, respectively. Similar interpretations also apply to each of (21) and (22). Now consider LR, S, and W in a given space. In $\mathfrak{L}$, the interpretation provided by the first equality in each of (20) to (22) is as based on Newey and McFadden (1994) except that, as will be seen below, their illustration for a Wald statistic is only appropriate for cases where $\tilde{\theta}_W = \tilde{\theta}$. In $\mathfrak{G}$ with the metric $(n\tilde{J})^{-1}$, S is (by construction) the squared length of $g(\tilde{\theta})$, whereas, LR and W are squared lengths of nonsingular linear transformations of $g(\tilde{\theta})$. Since the matrices associated with these transformations converge in probability to an identity matrix, the well-known asymptotic equivalence of the statistics is easily seen. Similarly, in $\mathfrak{R}$ with the metric $n\{\hat{R}\hat{J}^{-1}\hat{R}^{\top}\}^{-1}$, W is the squared length of $r(\hat{\theta})$, whereas, LR and S are squared lengths of nonsingular linear transformations of $r(\hat{\theta})$ where the matrices associated with these transformations converge in probability to an identity matrix. The interpretation of LR, S, and W as squared lengths in Θ is given by the last expression in (20) to (22), respectively, which will be discussed in more detail below. Finally, if $L(\theta)$ is an appropriate maximand associated with some other method of estimation where a type of information matrix equality holds, then (20) to (22) show that classical-type statistics can also be expressed in terms of a length in each of the four spaces.

3.2. Example of classical statistics as lengths in $\mathfrak{L}$

For a diagrammatic exposition, $\mathfrak{L}$ is an appealing space to consider as it can also show the relative magnitudes of statistics. Therefore, let S and W be evaluated with $\tilde{J} = H(\tilde{\theta})$ and $\hat{J} = \hat{H}$, respectively, and (for $r = p = 1$) let $L(\theta) = -10\left(\ln\theta + \frac{1}{\theta}\right)$ and $r(\theta) = \theta^3 - (0.4)^3$; here, $\theta > 0$ is the mean of an exponential distribution and $L(\theta)$ is the log-likelihood function for a random sample of ten observations with a sample mean of one. Then, the interpretation of LR, S, and W as lengths in $\mathfrak{L}$ is illustrated in Figure 1 (on page 38) where $L_S(\theta)$ and $L_W(\theta)$ are second-order Taylor series approximations of $L(\theta)$ at $\tilde{\theta}$ and $\hat{\theta}$, respectively, $\tilde{L} = L(\tilde{\theta})$,

$\hat{L} = L(\hat{\theta})$, $\tilde{L}_S = L_S(\tilde{\theta})$, $\hat{L}_W = L_W(\hat{\theta})$, and (on the vertical axis) the spaces $\mathfrak{L}$ and $\mathfrak{R}$ have different scales with the value of zero being with respect to $\mathfrak{R}$. To simplify the diagram, $L_{LR}(\theta)$ and $r_{LR}(\theta)$ are not displayed; in $\Theta \times \mathfrak{L}$, $L_{LR}(\theta)$ would be the quadratic curve through the point $(\tilde{\theta}, \tilde{L})$ with a maximum of $\hat{L}$ at $\hat{\theta}$ and, in $\Theta \times \mathfrak{R}$, $r_{LR}(\theta)$ would be a straight line through the point $(\tilde{\theta}, 0)$ with slope approximately half that of $r_S(\theta)$.

Figure 1 clearly shows the role played by $r_W(\theta)$ in the interpretation of a Wald statistic when $\tilde{\theta}_W \neq \tilde{\theta}$. The role of $r_W(\theta)$ is obscured in examples that consider linear restrictions where $\tilde{\theta}_W$ coincides with $\tilde{\theta}$ when $r = p$. For example, Newey and McFadden (1994, Figure 3, p. 2221) depict a situation where $\tilde{\theta}_W = \tilde{\theta}$, and Davidson and MacKinnon (2004, Figure 10.3, p. 433) illustrate a case where $\tilde{\theta}_W = \tilde{\theta}$ and $\hat{\theta}_S = \hat{\theta}$. It may be interesting to note that $\hat{\theta}_S$ coincides with $\hat{\theta}$ when $r(\theta) = \theta - \theta^0$ and $L_S(\theta)$ is obtained by setting $\bar{\theta}_S = \theta^0$ and $J_S = J_1(\theta^0)$ where the scalar θ is the mean of the univariate exponential family of distributions, θ^0 is known, $J_1(\theta) = E[H(\theta)]$, and given random sampling. Another interpretation of a Wald statistic, provided by Pagan (1982) and reproduced in Poirier (1995, Exercise 7.3.3, p. 374), is also based on an example where $\tilde{\theta}_W = \tilde{\theta}$. In particular, for $r(\theta) = \theta - \theta^0$ as above and W evaluated with $\hat{J} = \hat{H}$, Pagan (1982, Figure 1, p. 259) shows that $W = 2\int_{\tilde{\theta}}^{\hat{\theta}} g_W(\theta)d\theta$ where $\tilde{\theta} = \theta^0$, whereas, the first equality in (22) and $g_W(\theta) = dL_W(\theta)/d\theta$ yield $W = 2\int_{\tilde{\theta}_W}^{\hat{\theta}} g_W(\theta)d\theta$, which caters for cases where $\tilde{\theta}_W \neq \tilde{\theta}$. If desired, the graphs of $g(\theta)$, $g_S(\theta)$, and $g_W(\theta)$ could be added to Figure 1. Then, the areas $\int_{\tilde{\theta}}^{\hat{\theta}} g(\theta)d\theta$, $\int_{\tilde{\theta}}^{\hat{\theta}_S} g_S(\theta)d\theta$, and $\int_{\tilde{\theta}_W}^{\hat{\theta}} g_W(\theta)d\theta$ would equal $LR/2$, $S/2$, and $W/2$ respectively; the first two integrals would reproduce appropriate areas in Pagan (1982, Figure 2, p. 260) where, on the horizontal axis, the points A and D correspond to $\tilde{\theta}$ and $\hat{\theta}_S$, respectively.

3.3. Equalities among statistics

Suppose that $L(\theta)$ is a quadratic function of θ such that $H(\theta) = H$ is a positive definite (stochastic or nonstochastic) matrix with $H \xrightarrow{p} J_0$. In this case, $R_{LR} = \tilde{R}Z$ and $J_{LR} = b\hat{H}$ reduce to $R_{LR} = \tilde{R}$ and $J_{LR} = H$, respectively, as $Z = I_p$, $b = 1$, and $\hat{H} = H$; see the note at the end of Appendix A. Let $L_S(\theta)$ and $L_W(\theta)$ be second-order Taylor series approximations of $L(\theta)$ at $\tilde{\theta}$ and $\hat{\theta}$, respectively. Then, S and W are evaluated with $\tilde{J} = H$ and $\hat{J} = H$, respectively. Here, it is useful to separate the two cases where $r(\theta)$ is either a nonlinear or a linear function of θ. First, let $r(\theta)$ be a nonlinear function of θ. Then,

$$L(\theta) = L_{LR}(\theta) = L_S(\theta) = L_W(\theta) \qquad \text{and} \qquad r_{LR}(\theta) = r_S(\theta), \tag{23}$$

which show that the perceived frameworks of LR and S are identical (but not identical to the nominal framework) so $\hat{\theta} = \hat{\theta}_S$. Therefore, using the last expression in each of (20) to (22),

$$LR = S = \left|\left|\hat{\theta} - \tilde{\theta}\right|\right|^2_{nH} \qquad \text{and} \qquad W = \left|\left|\hat{\theta} - \tilde{\theta}_W\right|\right|^2_{nH} \tag{24}$$

where $LR \neq W$ in general. In Figure 1, if $L(\theta)$ and $L_S(\theta)$ are ignored and $L_W(\theta)$ is viewed to be as in (23), then W is as given in the diagram and $LR = S = 2\{L_W(\hat{\theta}) - L_W(\tilde{\theta})\} > W$. Second, let $r(\theta)$ be a linear function of θ. Then,

$$L(\theta) = L_{LR}(\theta) = L_S(\theta) = L_W(\theta) \qquad \text{and} \qquad r(\theta) = r_{LR}(\theta) = r_S(\theta) = r_W(\theta) \tag{25}$$

so, here, the perceived frameworks of LR, S, and W (and the nominal framework) are identical, and

$$LR = S = W = \left|\left|\hat{\theta} - \tilde{\theta}\right|\right|^2_{nH}. \tag{26}$$

Basically, (26) reproduces the equalities of Buse (1982) and Engle (1984); strictly speaking, (5) is a special case of (26) obtained by setting $H = \bar{H}$.

From examples demonstrating (26), particularly in textbooks, it is possible to get the impression that a quadratic $L(\theta)$ is sufficient for the equality of the three statistics. Such examples invariably consider linear restrictions so the difference between (24) and (26) is not seen. Therefore, it is important to note that a quadratic $L(\theta)$ per se is sufficient for the equality of only LR and S. The relationships in (23) and (25) help to explain the equalities in (24) and (26), respectively; i.e., two statistics are identical if their perceived frameworks are identical. For classical-type statistics, the results here can be used to either determine equalities or easily show and explain known equalities. For example, let $L(\theta)$ be the maximand associated with the efficient generalized method of moments estimator of θ for orthogonality conditions that are linear in θ. Then, (24) provides (and (23) explains) the equality of D and LM in Newey and West (1987, Proposition 3, p. 785). Similarly, (26) provides (and (25) explains) the equality of D, LM, and W in Newey and West (1987, Proposition 4, p. 785).

3.4. Perceived null

Henceforth, an appropriate set will be referred to as a type of null or alternative. For example, both H_0 and Ω_0 (H_1 and Ω_1) will be referred to as the nominal null (nominal alternative). Let Θ_{0C} and Θ_{1C} constitute a partition of Θ and, using (19) and the last expression in (17), consider C written as

$$C = \left|\left|\hat{\theta}_C - \tilde{\theta}_C\right|\right|^2_{nJ_C} \tag{27}$$

where $\hat{\theta}_C \xrightarrow{p} \theta_0$, $\tilde{\theta}_C \xrightarrow{p} \theta_0$, and (from the note at the end of Appendix B) $\hat{\theta}_C \in \Theta_{1C}$. Then, in Θ with the metric nJ_C, C can be viewed either as the squared length of $\hat{\theta}_C - \tilde{\theta}_C$ or, more usefully, as the squared distance from the point $\hat{\theta}_C \in \Theta_{1C}$ to the point $\tilde{\theta}_C \in \Theta_{0C}$. Now, $\tilde{\theta}_C$ is the optimal point in Θ_{0C} where optimality is with respect to the perceived maximand $L_C(\theta)$. Therefore, for a given sample, Θ_{0C} is a perceived null in the sense that (in the perceived framework of C) the restricted estimate of θ is chosen from points in Θ_{0C}. Similarly, Θ_{1C} is a perceived alternative.

The use of 'perceived' should avoid any possible confusion with 'implicit' as used in the literature. For example, Mizon and Richard (1986) define an implicit null, Davidson and MacKinnon (1987) define an implicit null and an implicit alternative, and White (1987) defines a 'model implicitly tested'. These 'implicit' definitions account for repeated sampling, whereas, the 'perceived' definitions are for a given sample. Here, by accounting for repeated sampling, $\Theta_0 = \{\theta \,|\, R_0(\theta - \theta_0) = 0, \ \theta \in \Theta\}$ will be viewed as the implicit null of C as $\Theta_{0C} \xrightarrow{p} \Theta_0$ in the sense that if $\breve{\theta} \in \Theta_{0C}$ and $\breve{\theta} \xrightarrow{p} \theta_*$, then $\theta_* \in \Theta_0$ since $r_C(\breve{\theta}) = 0$ and (7) shows that $r_C(\breve{\theta}) \xrightarrow{p} R_0(\theta_* - \theta_0)$. Hence, LR, S, and W have identical nominal nulls and identical implicit nulls, but their perceived nulls could differ depending on the form of $r(\theta)$. For example, in Figure 1, $\Omega_0 = \Theta_0 = \Theta_{0LR} = \Theta_{0S} = \{\tilde{\theta}\}$ and $\Theta_{0W} = \{\tilde{\theta}_W\}$. A nonlinear restriction considered by Gregory and Veall (1985) provides an example where Ω_0, Θ_0, and Θ_{0C} differ. In particular, let $\theta = (\theta_1, \theta_2)^\top$ and $r(\theta) = \theta_1\theta_2 - 1$ where θ_1 and θ_2 are positive scalars, and let $\tilde{\theta}$, $\hat{\theta}$, $\tilde{\theta}_W$, R_{LR}, $r(\hat{\theta})$, and $\hat{R}$ be provided by the sample. Then, in the positive quadrant of $\mathbb{R}^2$, Ω_0 is the set of points on the hyperbola given by $\theta_1\theta_2 = 1$, Θ_0 is the (set of points on the) straight line tangential to this hyperbola at the unknown point θ_0, Θ_{0S} is the straight line tangential to the hyperbola at $\tilde{\theta}$, Θ_{0LR} is the straight line (through the point $\tilde{\theta}$) given by $R_{LR}(\theta - \tilde{\theta}) = 0$, and Θ_{0W} is the straight line (through the point $\tilde{\theta}_W$) given by $r(\hat{\theta}) + \hat{R}(\theta - \hat{\theta}) = 0$.

4. Reformulation of a nominal null

4.1. General results

Following Davidson and MacKinnon (1993, p. 468), let $p : \mathbb{R}^r \to \mathbb{R}^r$ be a mapping such that, for any $x \in \mathbb{R}^r$, $p(x) = 0$ if and only if $x = 0$. Now, let $x = r(\theta)$, $q(\theta) = p(r(\theta))$, and $\Omega_0^q = \{\theta \,|\, q(\theta) = 0,\ \theta \in \Omega\}$, and recall that $\Omega_0 = \{\theta \,|\, r(\theta) = 0,\ \theta \in \Omega\}$. Then, as $\Omega_0^q = \Omega_0$, $H_0^q : \theta \in \Omega_0^q$ is a reformulation of the nominal null $H_0 : \theta \in \Omega_0$. For later reference, it is useful to note some results related to the mapping $p(x)$. First, let $Q(\theta) = \partial q(\theta)/\partial\theta^\top$ be the $r \times p$ matrix of derivatives, $\tilde{Q} = Q(\tilde{\theta})$, $\hat{Q} = Q(\hat{\theta})$, and $P_0 = P(0)$ where $P(x) = \partial p(x)/\partial x^\top$ is the $r \times r$ Jacobian matrix, which is assumed to be nonsingular for all $x \in \mathbb{R}^r$ and hence nonsingular for all $\theta \in \Omega$. Then, $Q(\theta) = P(x)R(\theta)$,

$$\tilde{Q} = P_0\tilde{R}, \tag{28}$$

and

$$\hat{Q} = \hat{P}\hat{R} \tag{29}$$

where $\hat{P} = P(r(\hat{\theta})) \xrightarrow{p} P_0$, $\tilde{Q} \xrightarrow{p} P_0R_0$, $\hat{Q} \xrightarrow{p} P_0R_0$, and (given the assumptions on $P(x)$, R_0, $\tilde{R}$, and $\hat{R}$) the matrices $\tilde{Q}$, $\hat{Q}$, and P_0R_0 have full row rank. Second, a mean value expansion of $p(x)$ at the zero vector is $p(x) = P_* x$ where P_* is $P(x)$ with its i-th row evaluated at $x_i^* = \alpha_i x$ for some $\alpha_i \in [0,1]$. This mean value expansion with $x = r(\hat{\theta})$ yields

$$q(\hat{\theta}) = \hat{P}_* r(\hat{\theta}) \tag{30}$$

where $\hat{P}_*$ is P_* with $x_i^* = \alpha_i r(\hat{\theta})$. Here, $x_i^* \xrightarrow{p} 0$ so $\hat{P}_* \xrightarrow{p} P_0$ and it is assumed that $\hat{P}_*$ is nonsingular. Third, if $p(x)$ is a linear mapping, then $q(\theta) = Pr(\theta)$, $\hat{Q} = P\hat{R} \xrightarrow{p} PR_0$, and

$$\hat{P}_* = \hat{P} = P \tag{31}$$

where P is a nonstochastic nonsingular matrix whose elements do not depend on θ.

The nominal frameworks $(\Omega, L(\theta), \Omega_0)$ and $(\Omega, L(\theta), \Omega_0^q)$ are identical as $\Omega_0 = \Omega_0^q$ so $\hat{\theta}$ $(\tilde{\theta})$ is the unrestricted (restricted) estimator of θ in both these nominal frameworks. Let C_q denote a statistic for testing H_0^q in the nominal framework $(\Omega, L(\theta), \Omega_0^q)$. Then, as is well known, $LR = LR_q$, $S = S_q$, and (in general) $W \neq W_q$. Now, by construction, a perceived maximand does not depend on the form of the restrictions so the perceived framework of C_q is defined for a given sample by $(\Theta, L_C(\theta), \Theta_{0C}^q)$ where $L_C(\theta)$ is as above, $\Theta_{0C}^q = \{\theta \,|\, q_C(\theta) = 0,\ \theta \in \Theta\}$, and $q_C(\theta)$ is an appropriate linear approximation of $q(\theta)$. Since the perceived frameworks of C and C_q can only differ in their perceived nulls, the unrestricted and restricted estimators of θ in the perceived framework of C_q are $\hat{\theta}_C$ as above and $\tilde{\theta}_C^q = \operatorname{argmax}_{\theta \in \Theta_{0C}^q} L_C(\theta)$, respectively. Then, by analogy with (27),

$$C_q = \left|\left|\hat{\theta}_C - \tilde{\theta}_C^q\right|\right|_{nJ_C}^2 . \tag{32}$$

Clearly, if $\Theta_{0C} = \Theta_{0C}^q$ (or, equivalently, if the perceived frameworks of C and C_q are identical) for a given sample, then $\tilde{\theta}_C = \tilde{\theta}_C^q$ and (27) and (32) yield $C = C_q$.

In the case of the likelihood ratio statistics, the row for (7) in Table 1 gives $r_{LR}(\theta) = R_{LR}(\theta - \tilde{\theta})$ so (8) provides the perceived null of LR as

$$\Theta_{0LR} = \{\theta \,|\, R_{LR}(\theta - \tilde{\theta}) = 0,\ \theta \in \Theta\}$$

and (by analogy) the perceived null of LR_q is

$$\Theta_{0LR}^q = \{\theta \,|\, Q_{LR}(\theta - \tilde{\theta}) = 0,\ \theta \in \Theta\}$$

where $Q_{LR} = \tilde{Q}Z$ and Z is as above; (A.4) in Appendix A shows that Z depends only on J_{LR} and an appropriately evaluated form of $H(\theta)$ so both R_{LR} and Q_{LR} use the same Z. Now, $R_{LR} = \tilde{R}Z$ and (28) provide $Q_{LR} = P_0 R_{LR}$ where P_0 is nonsingular so $R_{LR}(\theta - \tilde{\theta}) = 0$ is equivalent to $Q_{LR}(\theta - \tilde{\theta}) = 0$; i.e., $\Theta_{0LR} = \Theta^q_{0LR}$. Similarly, in the case of the score statistics, the perceived nulls of S and S_q are

$$\Theta_{0S} = \{\theta \,|\, \tilde{R}(\theta - \tilde{\theta}) = 0,\, \theta \in \Theta\} \qquad \text{and} \qquad \Theta^q_{0S} = \{\theta \,|\, \tilde{Q}(\theta - \tilde{\theta}) = 0,\, \theta \in \Theta\},$$

respectively, so $\Theta_{0S} = \Theta^q_{0S}$ given (28) where P_0 is nonsingular. Then, for $C = LR, S$, it follows that $C = C_q$ as $\Theta_{0C} = \Theta^q_{0C}$. Therefore, the likelihood ratio (a score) statistic is invariant to a reformulation of a nominal null as the perceived frameworks of LR and LR_q (S and S_q) are identical. Note that, since $\Theta^q_{0C} = \Theta_{0C} \xrightarrow{p} \Theta_0$ for $C = LR, S$, the implicit nulls of the likelihood ratio and a score statistic are invariant to a reformulation of H_0 as H^q_0. This implies that the implicit null of a Wald statistic is also invariant to the reformulation as the three statistics have identical implicit nulls.

For the Wald statistics, the perceived null of W is

$$\Theta_{0W} = \{\theta \,|\, r(\hat{\theta}) + \hat{R}(\theta - \hat{\theta}) = 0,\, \theta \in \Theta\}$$

and the perceived null of W_q is $\Theta^q_{0W} = \{\theta \,|\, q(\hat{\theta}) + \hat{Q}(\theta - \hat{\theta}) = 0,\, \theta \in \Theta\}$, which (using (29) and (30)) can be written as

$$\Theta^q_{0W} = \{\theta \,|\, \hat{P}_* r(\hat{\theta}) + \hat{P}\hat{R}(\theta - \hat{\theta}) = 0,\, \theta \in \Theta\}.$$

Now, consider two cases for the mapping $p(x)$. First, suppose that $p(x)$ is a linear mapping. Then, (31) holds where P is nonsingular so $\Theta_{0W} = \Theta^q_{0W}$, which provides $W = W_q$. In this case, a Wald statistic is invariant to a reformulation of a nominal null as the perceived frameworks of W and W_q are identical. Second, suppose that $p(x)$ is a mapping such that $\hat{P}_* \neq \hat{P}$. Then, $\Theta_{0W} \neq \Theta^q_{0W}$ so (in general) $\tilde{\theta}_W \neq \tilde{\theta}^q_W$ and $W \neq W_q$. In this case, a Wald statistic is not invariant to a reformulation of a nominal null as the different perceived nulls of W and W_q provide different restricted estimates of θ in their respective perceived frameworks.

4.2. Example of perceived nulls of Wald statistics

Different perceived nulls for Wald statistics can be illustrated by considering an example in Lafontaine and White (1986). To provide an accurate illustration of this example, it is necessary to show that Θ can be the space associated with a subset of the parameters in a model. To see this, let $L_*(\psi) = L_*(\delta, \theta)$ be a log-likelihood function for an $l \times 1$ vector of parameters $\psi = \left(\delta^\top, \theta^\top\right)^\top$ where δ and θ are $(l-p) \times 1$ and $p \times 1$ vectors, respectively, and corresponding versions of Assumptions 1 and 2 hold with J_{*0} denoting the limiting information matrix. Also, let the restrictions be given by $f(\psi) = 0$, $\hat{\psi}$ be the unrestricted ML estimator of ψ, and $\hat{J}_*$ be a matrix evaluated at $\hat{\psi}$ such that $\hat{J}_* \xrightarrow{p} J_{*0}$. Then, a Wald statistic for testing $f(\psi) = 0$ is $W_f = nf(\hat{\psi})^\top \{\hat{F}\hat{J}_*^{-1}\hat{F}^\top\}^{-1} f(\hat{\psi})$ where $\hat{F} = \partial f(\hat{\psi})/\partial \psi^\top$. Now, let $\hat{J}_*^{-1}$ and J_{*0}^{-1} be conformably partitioned with $\psi = \left(\delta^\top, \theta^\top\right)^\top$ as

$$\hat{J}_*^{-1} = \begin{bmatrix} \hat{J}^{\delta\delta} & \hat{J}^{\delta\theta} \\ \hat{J}^{\theta\delta} & \hat{J}^{\theta\theta} \end{bmatrix} \qquad \text{and} \qquad J_{*0}^{-1} = \begin{bmatrix} J_0^{\delta\delta} & J_0^{\delta\theta} \\ J_0^{\theta\delta} & J_0^{\theta\theta} \end{bmatrix},$$

respectively, and let $\hat{J}^{-1} \equiv \hat{J}^{\theta\theta} \xrightarrow{p} J_0^{\theta\theta} \equiv J_0^{-1}$. If $f(\psi) = r(\theta)$, then $\hat{F} = [\mathbf{0}_{r\times(l-p)}, \hat{R}]$ and $W_f = W = \left|\left|\hat{\theta} - \tilde{\theta}_W\right|\right|_{n\hat{J}}^2$. Similarly, if $f(\psi) = q(\theta)$, then $W_f = W_q = \left|\left|\hat{\theta} - \tilde{\theta}_W^q\right|\right|_{n\hat{J}}^2$. These Wald statistics are appropriate for illustrating the example in Lafontaine and White (1986).

Let $l > p = r = 1$, $\Omega = \{\theta \,|\, \theta > 0\}$, and $r(\theta) = \theta - 1$. Then, $\Omega \subset \Theta = \mathbb{R}$, $\Omega_0 = \Theta_{0W} = \{1\}$, and $W = \left|\left|\hat{\theta} - \tilde{\theta}\right|\right|_{n\hat{J}}^2$ as $\tilde{\theta}_W = \tilde{\theta}$ $(= 1)$. Lafontaine and White (1986) consider different forms of $q(\theta)$ given by $q^{(k)}(\theta) = \theta^k - 1$ for $k \in \{-1, \pm 2, \pm 3, \dots\}$. Therefore, to indicate quantities related to the k-th form of $q(\theta)$, the subscript and superscript q will be replaced with k and (k), respectively. Then, for the k-th form of $q(\theta)$, a Wald statistic is $W_k = \left|\left|\hat{\theta} - \tilde{\theta}_W^{(k)}\right|\right|_{n\hat{J}}^2$ and its perceived null is $\Theta_{0W}^{(k)} = \{\theta \,|\, q_W^{(k)}(\theta) = 0,\, \theta \in \Theta\} = \{\tilde{\theta}_W^{(k)}\}$ where $q_W^{(k)}(\theta) = \hat{a}_k\theta - \hat{b}_k$ is a first-order Taylor series approximation of $q^{(k)}(\theta)$ at $\hat{\theta}$, $\hat{a}_k = k\hat{\theta}^{k-1}$, $\hat{b}_k = 1 - (1-k)\hat{\theta}^k$, and $\tilde{\theta}_W^{(k)} = \hat{b}_k/\hat{a}_k$. Figure 2 (on page 39) illustrates the perceived nulls of W and W_k for $k = -16, -8, 8, 16$ with $\hat{\theta} = 1.1$. In this figure, the value of zero is for $\mathfrak{R}$ on the vertical axis so the perceived null of a Wald statistic is the point where the linear approximation of a restriction intersects the horizontal axis Θ; to avoid unnecessary lines, the graphs of $r(\theta)$, $q^{(k)}(\theta)$ and $q_W^{(k)}(\theta)$ are appropriately truncated. Therefore, the perceived nulls of W and W_k are the points $\tilde{\theta}$ and $\tilde{\theta}_W^{(k)}$, respectively, on the horizontal axis. Since $\tilde{\theta}_W^{(k)}$ moves further away from $\hat{\theta}$ as k decreases and as all the Wald statistics use the same scalar metric $n\hat{J}$, the qualitative effect on $W_k = \left|\left|\hat{\theta} - \tilde{\theta}_W^{(k)}\right|\right|_{n\hat{J}}^2$ as k varies is given by the change in the Euclidean distance between $\hat{\theta}$ and $\tilde{\theta}_W^{(k)}$. Hence, Figure 2 demonstrates that W_k increases as k decreases, a result shown numerically by Lafontaine and White (1986, Table 1, p. 37). This qualitative effect is the same for any metric $n\hat{J}$ and, if $\hat{\theta} < 1$, a similar diagram would show that W_k increases as k increases.

4.3. Interpreting a Wald statistic as a squared distance in Θ

It could be argued that a reformulation of a nominal null is justifiable if the perceived null of a Wald statistic is not an adequate representation of the nominal null, the hypothesis of interest. To see this, let Ω_0, Ω_1, and Ω_2 constitute a partition of Θ where $\Omega_2 = \varnothing$ if $\Omega = \Theta$, and consider $W = \left|\left|\hat{\theta} - \tilde{\theta}_W\right|\right|_{n\hat{J}}^2$ as the squared distance from the point $\hat{\theta}$ to the point $\tilde{\theta}_W$. Now, $\hat{\theta}$ is a point in both the perceived alternative Θ_{1W} and the nominal alternative Ω_1, whereas, $\tilde{\theta}_W$ is a point in the perceived null Θ_{0W} and not necessarily a point in the nominal null Ω_0. Therefore, if $\tilde{\theta}_W \notin \Omega_0$, then Θ_{0W} is an inadequate representation of Ω_0 in the sense that W is then the squared distance from a point in Ω_1 to either another point in Ω_1 or a point in Ω_2. For example, in Figure 1, $\Theta_{0W} = \{\tilde{\theta}_W\}$ is an inadequate representation of $\Omega_0 = \{\tilde{\theta}\}$ as $\tilde{\theta}_W$ is a point in Ω_1. Similarly, in Figure 2, $\Theta_{0W}^{(k)} = \{\tilde{\theta}_W^{(k)}\}$ is also an inadequate representation of Ω_0 as W_k is the squared distance from $\hat{\theta}$ to $\tilde{\theta}_W^{(k)}$ where $\tilde{\theta}_W^{(k)}$ is either a point in Ω_1 or a point not even in the nominal parameter space Ω; i.e., in Figure 2, $\tilde{\theta}_W^{(k)} \in \Omega_2 = \{\theta \,|\, \theta \leq 0\}$ if $k \leq -39$. To ensure that Θ_{0W} is an adequate representation of Ω_0, it would seem reasonable to require $\tilde{\theta}_W \in \Omega_0$. Then, W would be the squared distance from $\hat{\theta} \in \Omega_1$ to $\tilde{\theta}_W \in \Omega_0$. Examples of statistics based on such a distance are FG in Critchley, Marriott and Salmon (1996) and MC in Newey and West (1987). On the one hand, as seen above, Figure 1 is an example where Θ_{0W} is an inadequate representation of Ω_0 so a reformulation of $H_0 : \theta^3 = (0.4)^3$ as $H_0^q : \theta = 0.4$ provides W_q as the squared distance from

$\hat{\theta} \in \Omega_1$ to $\tilde{\theta}_W^q \in \Omega_0$; in Figure 1, $W_q = 2\{L_W(\hat{\theta}) - L_W(\tilde{\theta})\}$ where W_q is evaluated with $\hat{J} = \hat{H}$. On the other hand, Figure 2 provides an example where Θ_{0W} is an adequate representation of Ω_0 as $\Theta_{0W} = \Omega_0 = \{\tilde{\theta}\}$ so, here, there is no need to consider a reformulation of $H_0 : \theta = 1$ as $H_0^{(k)} : \theta^k = 1$.

It is important to note two points concerning the argument above that a reformulation of H_0 is justifiable if $\tilde{\theta}_W \notin \Omega_0$. First, the examples in Figures 1 and 2 may suggest that a Wald statistic should be used only when the restrictions are linear. However, a Wald statistic is invariant to a one-to-one transformation of the entire parameter space with the restrictions also appropriately transformed; see Dagenais and Dufour (1991, p. 1607). In particular, if $(\Gamma, L_\gamma(\gamma), \Gamma_0)$ is an appropriately transformed version of the framework $(\Omega, L(\theta), \Omega_0)$, then a Wald statistic for nonlinear restrictions in θ can be obtained as a Wald statistic for linear restrictions in γ. For a parameter space, convention may suggest a particular specification that is in some sense natural or convenient for purposes of interpretation but, ultimately, even a conventional specification is arbitrary. Therefore, as Θ is obtained with reference to Ω, the argument above depends on the specification of the parameters in a nominal framework. Second, the form $W = \left|\left|\hat{\theta} - \tilde{\theta}_W\right|\right|^2_{n\hat{J}}$ is such that a reformulation of H_0 can only affect $\tilde{\theta}_W$. It is also possible to write $W = \left|\left|\hat{\theta} - \tilde{\theta}\right|\right|^2_{nJ_r}$ where only the metric nJ_r is affected by a reformulation of H_0. Formally, this metric can be obtained from the equality of $T_1(A, R)$ and $T_6(A_{61}, R_{61})$ as given in the proof of the theorem in Dastoor (2003). In particular, it can be shown that $T_1(\hat{J}, \hat{R}) = W$ and $T_6(A_{61}, R_{61}) = n(\hat{\theta} - \tilde{\theta})^\top A_{61}(\hat{\theta} - \tilde{\theta})$ where A_{61} is nonsingular. With $A = \hat{J}$ and $R = \hat{R}$, it can also be shown that A_{61} is positive definite so $W = \left|\left|\hat{\theta} - \tilde{\theta}\right|\right|^2_{nJ_r}$ where $J_r = A_{61}$. Using this form for a Wald statistic, both W and W_q can be viewed as the squared distance from $\hat{\theta}$ to $\tilde{\theta}$ where each statistic uses a different metric. However, in order to compare W and W_q, both statistics should be expressed as squared distances using the same metric. For Wald statistics, the choice of $n\hat{J}$ for the metric in Θ is particularly useful as it provides the simple interpretation of W as the squared distance from $\hat{\theta}$ to $\tilde{\theta}_W$ where a reformulation of H_0 can only affect $\tilde{\theta}_W$. This simple interpretation would not necessarily be provided by using another metric. For example, using the Euclidean metric in Θ when $\hat{J} \neq I_p$, W would be viewed as the squared distance from $\sqrt{n}\hat{J}^{1/2}\hat{\theta}$ to $\sqrt{n}\hat{J}^{1/2}\tilde{\theta}_W$ where $\hat{J}^{1/2}$ is the matrix square root of $\hat{J}$. Therefore, the argument that a reformulation of H_0 is justifiable if $\tilde{\theta}_W \notin \Omega_0$ is also dependent on the choice of $n\hat{J}$ for the metric in Θ.

5. Locally equivalent alternative

Consider another nominal framework given by $(\Omega, L^*(\theta), \Omega_0)$ where $L^*(\theta)$ is a log-likelihood function not equal to $L(\theta)$ for all $\theta \in \Omega$, asymptotic results are obtained under H_0, and corresponding versions of Assumptions 1 and 2 hold with J_0^* denoting the limiting information matrix. In this nominal framework, the unrestricted and restricted ML estimators of θ are $\hat{\theta}^* = \text{argmax}_{\theta\in\Omega} L^*(\theta)$ and $\tilde{\theta}^* = \text{argmax}_{\theta\in\Omega_0} L^*(\theta)$, respectively. Let $g^*(\theta) = \partial L^*(\theta)/\partial\theta$. Then, following Godfrey and Wickens (1982), $L^*(\theta)$ is an LEA to $L(\theta)$ with respect to H_0 if

$$L^*(\theta) = L(\theta) \text{ for only } \theta \in \Omega_0, \qquad g^*(\tilde{\theta}) = g(\tilde{\theta}), \qquad \text{and} \qquad J_0^* = J_0. \tag{33}$$

The three conditions in (33) will be referred to as the G-W conditions; the first two are explicitly stated by Godfrey and Wickens (1982, p. 76) and the third is implicit in the specification of the models in their equations (2.1) and (2.9). The first condition ensures that the two nominal frameworks $(\Omega, L^*(\theta), \Omega_0)$ and $(\Omega, L(\theta), \Omega_0)$ have identical restricted models with different unrestricted models, and the last two conditions ensure that S is identical to a score statistic for testing H_0 in the nominal framework $(\Omega, L^*(\theta), \Omega_0)$ when $\tilde{J}$ is used as an estimator of J_0^*. Godfrey (1981, 1988) and Godfrey and Wickens (1982) provide numerous examples of LEAs, which yield the equality of appropriate score statistics.

The analysis to be carried out is simplified if $L^*(\theta)$ satisfies the G-W conditions, which may appear to be quite restrictive at first sight. For example, with respect to the second condition, Davidson and MacKinnon (1993, p. 470) remark that 'this requirement is too strong: It is enough if the components of $\tilde{g}^2$ are all linear combinations of those of $\tilde{g}^1$ and vice versa.' However, the G-W conditions are not too strong provided $L^*(\theta)$ is appropriately specified. To see this, let β be a $p\times 1$ vector and consider another nominal framework given by

$(B, L^+(\beta), B_0)$ such that the two frameworks $(\Omega, L(\theta), \Omega_0)$ and $(B, L^+(\beta), B_0)$ have identical restricted models under their respective nominal nulls; i.e., $L^+(\beta) = L(\theta)$ either for a $\theta \in \Omega_0$ and an appropriate $\beta \in B_0$, or for a $\beta \in B_0$ and an appropriate $\theta \in \Omega_0$. Then, there must be a one-to-one transformation between the points in Ω_0 and those in B_0; without such a transformation, the two frameworks could yield different restricted models. Therefore, let $\beta = \phi(\theta)$ where $\phi(.)$ is a one-to-one transformation with the Jacobian $\Phi(\theta) = \partial\beta/\partial\theta^\top$, a nonsingular matrix for all $\theta \in \Omega$. Now, let $L^*(\theta) = L^+(\phi(\theta))$, $g^+(\beta) = \partial L^+(\beta)/\partial\beta$, and corresponding versions of Assumptions 1 and 2 hold with J_0^+ denoting the limiting information matrix in the nominal framework $(B, L^+(\beta), B_0)$. Then, applying the G-W conditions to this $L^*(\theta)$ shows that $L^+(\beta)$ is an LEA to $L(\theta)$ with respect to H_0 if

$$L^+(\phi(\theta)) = L(\theta) \text{ for only } \theta \in \Omega_0, \quad \tilde{\Phi}^\top g^+(\tilde{\beta}) = g(\tilde{\theta}), \quad \text{and} \quad \Phi_0^\top J_0^+ \Phi_0 = J_0 \qquad \textbf{(34)}$$

where $\tilde{\beta} = \phi(\tilde{\theta})$, $\tilde{\Phi} = \Phi(\tilde{\theta})$, and $\Phi_0 = \Phi(\theta_0)$. The proof of (34) is given in Appendix C and a simple example that illustrates the specification of an appropriate $L^*(\theta)$ is provided in Appendix D. In (34), the first condition ensures that the two frameworks $(B, L^+(\beta), B_0)$ and $(\Omega, L(\theta), \Omega_0)$ have identical restricted models with different unrestricted models and, in the spirit of Davidson and MacKinnon (1993), the second condition requires $g^+(\tilde{\beta})$ to be a nonsingular linear transformation of $g(\tilde{\theta})$. Therefore, provided $L^*(\theta)$ is appropriately specified, the G-W conditions are not as restrictive as may first appear.

Henceforth, it is assumed that $L^*(\theta)$ satisfies the G-W conditions in (33). Therefore, the nominal frameworks $(\Omega, L^*(\theta), \Omega_0)$ and $(\Omega, L(\theta), \Omega_0)$ have identical restricted models with different unrestricted models so $\tilde{\theta}^* = \tilde{\theta}$ and (in general) $\hat{\theta}^* \neq \hat{\theta}$. Let C^* denote a statistic for testing H_0 in the nominal framework $(\Omega, L^*(\theta), \Omega_0)$ where $S^* \equiv n^{-1} g^*(\tilde{\theta}^*)^\top \tilde{J}^{-1} g^*(\tilde{\theta}^*)$; i.e., here, $J_0^* = J_0$ so $\tilde{J}$ is used in both S^* and S. Then, it is well known that $S = S^*$ and (in general) $LR \neq LR^*$ and $W \neq W^*$. All quantities associated with C^* can be obtained by analogy with the results in Sections 2 and 3. In particular, the perceived framework of C^* is defined for a given sample by $(\Theta, L_C^*(\theta), \Theta_{0C}^*)$, $\mathcal{P}_C^* \equiv (\bar{\theta}_C^*, J_C^*, \ddot{\theta}_C^*, R_C^*)$,

$$L_C^*(\theta) = L^*(\bar{\theta}_C^*) + g^*(\bar{\theta}_C^*)^\top(\theta - \bar{\theta}_C^*) - \frac{n}{2}(\theta - \bar{\theta}_C^*)^\top J_C^*(\theta - \bar{\theta}_C^*), \qquad \textbf{(35)}$$

$$\Theta^*_{0C} = \{\theta \,|\, r(\ddot{\theta}^*_C) + R^*_C(\theta - \ddot{\theta}^*_C) = 0,\, \theta \in \Theta\}, \qquad \textbf{(36)}$$

and

$$C^* = \left|\left|\hat{\theta}^*_C - \tilde{\theta}^*_C\right|\right|^2_{nJ^*_C} \qquad \textbf{(37)}$$

where $\hat{\theta}^*_C = \text{argmax}_{\theta\in\Theta} L^*_C(\theta)$ and $\tilde{\theta}^*_C = \text{argmax}_{\theta\in\Theta^*_{0C}} L^*_C(\theta)$ are the unrestricted and restricted estimators of θ, respectively, in the perceived framework of C^*. Then, the perceived maximands (perceived nulls) of LR^*, S^*, and W^* are obtained from (35) ((36)) by setting $\mathcal{P}^*_{LR} = (\hat{\theta}^*, J^*_{LR}, \tilde{\theta}, R^*_{LR})$, $\mathcal{P}^*_S = (\tilde{\theta}, \tilde{J}, \tilde{\theta}, \tilde{R})$, and $\mathcal{P}^*_W = (\hat{\theta}^*, \hat{J}^*, \hat{\theta}^*, \hat{R}^*)$, respectively. Unlike the case in the previous section where the perceived frameworks of C and C_q could only differ in their perceived nulls, here, the perceived frameworks of C and C^* could differ in their perceived maximands and in their perceived nulls. However, if the perceived frameworks of C and C^* are identical for a given sample, then $\hat{\theta}_C = \hat{\theta}^*_C$, $\tilde{\theta}_C = \tilde{\theta}^*_C$, $J_C = J^*_C$, and (27) and (37) yield $C = C^*$.

For $C = LR, W$, it is easily seen that $L_C(\theta) \neq L^*_C(\theta)$ for all $\theta \in \Theta$, and $\Theta_{0C} \neq \Theta^*_{0C}$ (although $\tilde{\theta}_{LR} = \tilde{\theta}^*_{LR} = \tilde{\theta}$) so, in general, $C \neq C^*$. Therefore, the likelihood ratio (a Wald) statistic is not invariant to an LEA as the perceived frameworks of LR and LR^* (W and W^*) differ. Finally, in the case of the score statistics, the first two conditions in (33) show that $L_S(\theta) = L^*_S(\theta)$ for all $\theta \in \Theta$ and it is easily seen that $\Theta_{0S} = \Theta^*_{0S}$ so $S = S^*$. Therefore, an appropriate score statistic is invariant to an LEA as the perceived frameworks of S and S^* are identical.

6. Concluding remarks

This paper has shown that the distinction between a nominal framework for the three classical statistics and a perceived framework for each classical statistic provides more ways to interpret the statistics and intuitively explains as well as more easily shows some well-known results. In a perceived framework, the classical procedures per se are equivalent so, in finite samples, a difference between two statistics arises as a result of applying the equivalent procedures to different perceived frameworks and, asymptotically, the statistics are equivalent as the difference between their perceived frameworks vanishes. Then, as a perceived framework is meant to represent the framework as seen by a statistic, it is not surprising that two statistics are identical if their perceived frameworks are identical; this must be the case for a perceived framework to be meaningful. For the normally separately treated issues of a reformulation of a nominal null hypothesis and of an LEA, the identicalness of (or different) perceived frameworks explains the invariance or non-invariance properties of the statistics and also provides a useful view of these properties. In particular, the non-invariance of a Wald statistic (to a reformulation of a nominal null) is, on the one hand, desirable as W and W_q have different perceived frameworks and, on the other hand, undesirable as two identical nominal frameworks provide these different perceived frameworks. Similarly, the invariance of a score statistic (to an LEA) is desirable as S and S^* have identical perceived frameworks, and undesirable as two different nominal frameworks provide the identical perceived frameworks. In the case of the likelihood ratio statistic, both its invariance (to a reformulation of a nominal null) and its non-invariance (to an LEA) are desirable; is it surprising that the 'father' of the 'holy trinity' is omniscient? Finally, although this paper has focussed on the behaviour of classical statistics, the analysis is applicable to appropriate classical-type statistics, and the concept of a perceived framework could also be used to examine the behaviour of other asymptotic test statistics.

References

Breusch, T. S. (1978), Testing for autocorrelation in dynamic linear models, *Australian Economic Papers*, 17(31), 334–355.

Breusch, T. S. and A. R. Pagan (1979), A simple test for heteroscedasticity and random coefficient variation, *Econometrica*, 47(5), 1287–1294.

Breusch, Trevor S. and Peter Schmidt (1988), Alternative forms of the Wald test: How long is a piece of string?, *Communications in Statistics - Theory and Methods*, 17(8), 2789–2795.

Buse, A. (1982), The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note, *The American Statistician*, 36(3), 153–157.

Critchley, Frank, Paul Marriott and Mark Salmon (1996), On the differential geometry of the Wald test with nonlinear restrictions, *Econometrica*, 64(5), 1213–1222.

Dagenais, Marcel G. and Jean-Marie Dufour (1991), Invariance, nonlinear models, and asymptotic tests, *Econometrica*, 59(6), 1601–1615.

Dastoor, Naorayex K. (2003), The equality of comparable extended families of classical-type and Hausman-type statistics, *Journal of Econometrics*, 117(2), 313–330.

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

Davidson, Russell and James G. MacKinnon (1987), Implicit alternatives and the local power of test statistics, *Econometrica*, 55(6), 1305–1329.

Davidson, Russell and James G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York.

Davidson, Russell and James G. MacKinnon (2004), *Econometric Theory and Methods*, Oxford University Press, New York.

Engle, Robert F. (1984), Wald, likelihood ratio, and Lagrange multiplier tests in econometrics, in *Handbook of Econometrics, Volume II*, eds. Zvi Griliches and Michael D. Intriligator, chapter 13, North-Holland, Amsterdam, 775–826.

Godfrey, L. G. (1978), Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables, *Econometrica*, 46(6), 1293–1301.

Godfrey, L. G. (1981), On the invariance of the Lagrange multiplier test with respect to certain changes in the alternative hypothesis, *Econometrica*, 49(6), 1443–1455.

Godfrey, L. G. (1988), *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches*, Cambridge University Press, Cambridge.

Godfrey, L. G. and M. R. Wickens (1982), Tests of misspecification using locally equivalent alternative models, in *Evaluating the Reliability of Macro-Economic Models*, eds. Gregory C. Chow and Paolo Corsi, chapter 6, John Wiley and Sons Ltd, 71–99.

Gourieroux, C. and A. Monfort (1989), A general framework for testing a null hypothesis in a "mixed" form, *Econometric Theory*, 5(1), 63–82.

Greene, William H. (2003), *Econometric Analysis*, Prentice Hall, Upper Saddle River, New Jersey, fifth edition.

Gregory, Allan W. and Michael R. Veall (1985), Formulating Wald tests of nonlinear restrictions, *Econometrica*, 53(6), 1465–1468.

Kemp, Gordon C. R. (2001), Invariance and the Wald test, *Journal of Econometrics*, 104(2), 209–217.

Lafontaine, Francine and Kenneth J. White (1986), Obtaining any Wald statistic you want, *Economics Letters*, 21(1), 35–40.

Mizon, Grayham E. and Jean-Francois Richard (1986), The encompassing principle and its application to testing non-nested hypotheses, *Econometrica*, 54(3), 657–678.

Newey, Whitney K. and Daniel McFadden (1994), Large sample estimation and hypothesis testing, in *Handbook of Econometrics, Volume IV*, eds. Robert F. Engle and Daniel L. McFadden, chapter 36, North-Holland, Amsterdam, 2111–2245.

Newey, Whitney K. and Kenneth D. West (1987), Hypothesis testing with efficient method of moments estimation, *International Economic Review*, 28(3), 777–787.

Pagan, A. (1982), Estimation and control of linear econometric models, *IHS-Journal*, 6, 247–268.

Phillips, P. C. B. and Joon Y. Park (1988), On the formulation of Wald tests of nonlinear restrictions, *Econometrica*, 56(5), 1065–1083.

Poirier, Dale J. (1995), *Intermediate Statistics and Econometrics: A Comparative Approach*, MIT Press, Cambridge, Massachusetts.

White, Halbert (1987), Specification testing in dynamic models, in *Advances in Econometrics, Fifth World Congress, Volume I*, ed. Truman F. Bewley, chapter 1, Cambridge University Press, Cambridge, 1–58.

Appendix A

Proof of Proposition 1. To show $LR = LR(\mathcal{P}_{LR})$, three preliminary results are required. First, let $d = \sqrt{n}(\hat{\theta} - \tilde{\theta})$ and recall that $H(\theta) = -n^{-1}\partial^2 L(\theta)/\partial\theta\partial\theta^\top$. Then, as $g(\hat{\theta}) = 0$, an exact second-order Taylor series expansion of $L(\tilde{\theta})$ at $\hat{\theta}$ yields

$$L(\hat{\theta}) - L(\tilde{\theta}) = \frac{1}{2} d^\top H_{LR} d \tag{A.1}$$

where $H_{LR} = H(\theta_{LR})$, $\theta_{LR} = \alpha\hat{\theta} + (1-\alpha)\tilde{\theta}$ for some $\alpha \in [0,1]$, and $H_{LR} \xrightarrow{p} J_0$. Now, let

$$b = \frac{d^\top H_{LR} d}{n(\hat{\theta} - \tilde{\theta})^\top \hat{H}(\hat{\theta} - \tilde{\theta})} \tag{A.2}$$

where (although H_{LR} is not necessarily positive definite) the numerator is positive given (A.1) with $L(\hat{\theta}) > L(\tilde{\theta})$, the denominator is also positive as $\hat{\theta} \neq \tilde{\theta}$ and $\hat{H}$ is positive definite, and $b - 1 = (d^\top \hat{H} d)^{-1} d^\top (H_{LR} - \hat{H}) d = o_p(1)$; it can be shown that $d^\top \hat{H} d \overset{a}{\sim} \chi^2(r)$. Therefore, $b > 0$ and $b \xrightarrow{p} 1$.

Second, a mean value expansion of $g(\hat{\theta}) = 0$ at $\tilde{\theta}$ gives

$$g(\tilde{\theta}) = nH_*(\hat{\theta} - \tilde{\theta}) \tag{A.3}$$

where H_* is $H(\theta)$ with each of its rows evaluated at a (possibly different) mean value given by a convex combination of $\hat{\theta}$ and $\tilde{\theta}$, and H_* is assumed to be nonsingular; in general, H_* is not necessarily a symmetric matrix although $H_* \xrightarrow{p} J_0$ and $H_*^\top \xrightarrow{p} J_0$. Let

$$Z = (H_*^\top)^{-1} J_{LR} \tag{A.4}$$

where $J_{LR} = b\hat{H}$. Then, Z is a $p \times p$ nonsingular matrix such that $Z \xrightarrow{p} I_p$.

Third, (A.4) and $R_{LR} = \tilde{R}Z$ can be written as

$$H_*^{-1} = J_{LR}^{-1} Z^\top \qquad \text{and} \qquad \tilde{R}^\top = (Z^\top)^{-1} R_{LR}^\top, \tag{A.5}$$

respectively. Then, substituting (4) into (A.3) gives

$$\hat{\theta} - \tilde{\theta} = \frac{1}{n} H_*^{-1} \tilde{R}^\top \tilde{\lambda}, \tag{A.6}$$

which, using (A.5), provides

$$\hat{\theta} - \tilde{\theta} = \frac{1}{n} J_{LR}^{-1} R_{LR}^{\top} \tilde{\lambda}. \tag{A.7}$$

Now, to see that $LR = LR(\mathcal{P}_{LR})$, let $C = LR$ and $\mathcal{P}_{LR} \equiv (\bar{\theta}_{LR}, J_{LR}, \ddot{\theta}_{LR}, R_{LR}) = (\hat{\theta}, b\hat{H}, \tilde{\theta}, \tilde{R}Z)$ where b and Z are given by (A.2) and (A.4), respectively. Then, all the quantities in the third column of Table 1 are easily obtained; for the last entry in this column, the row for (7) provides $r_{LR}(\hat{\theta}_{LR}) = R_{LR}(\hat{\theta}_{LR} - \tilde{\theta})$ so substituting this into (11) and then sequentially using $\hat{\theta}_{LR} = \hat{\theta}$ and (A.7) gives $\tilde{\theta}_{LR} = \tilde{\theta}$. Using $J_{LR} = b\hat{H}$ and (A.2), the row for (6) in Table 1 provides

$$L_{LR}(\theta) = L(\hat{\theta}) - \frac{1}{2} d^{\top} H_{LR} d \frac{(\theta - \hat{\theta})^{\top} \hat{H} (\theta - \hat{\theta})}{(\hat{\theta} - \tilde{\theta})^{\top} \hat{H} (\hat{\theta} - \tilde{\theta})},$$

which (given $\hat{\theta}_{LR} = \hat{\theta}$, $\tilde{\theta}_{LR} = \tilde{\theta}$, and (A.1)) shows that $L_{LR}(\hat{\theta}_{LR}) = L(\hat{\theta})$ and $L_{LR}(\tilde{\theta}_{LR}) = L(\tilde{\theta})$ so (1) and (14) yield $LR = LR(\mathcal{P}_{LR})$.

To see that $S = S(\mathcal{P}_S)$, let $C = S$ and $\mathcal{P}_S \equiv (\bar{\theta}_S, J_S, \ddot{\theta}_S, R_S) = (\tilde{\theta}, \tilde{J}, \tilde{\theta}, \tilde{R})$. Then, all the quantities in the fourth column of Table 1 are easily obtained; for the last entry in this column, the row for (7) provides $r_S(\hat{\theta}_S) = \tilde{R}(\hat{\theta}_S - \tilde{\theta})$ so substituting this into (11) and then sequentially using $\hat{\theta}_S = \tilde{\theta} + \frac{1}{n}\tilde{J}^{-1} g(\tilde{\theta})$ and (4) gives $\tilde{\theta}_S = \tilde{\theta}$. In Table 1, the row for (9) provides $g_S(\tilde{\theta}_S) = g(\tilde{\theta})$ so (2) and (15) yield $S = S(\mathcal{P}_S)$.

Finally, to see that $W = W(\mathcal{P}_W)$, let $C = W$ and $\mathcal{P}_W \equiv (\bar{\theta}_W, J_W, \ddot{\theta}_W, R_W) = (\hat{\theta}, \hat{J}, \hat{\theta}, \hat{R})$. Then, all the quantities in the last column of Table 1 are easily obtained and the row for (7) provides $r_W(\hat{\theta}_W) = r(\hat{\theta})$ so (3) and (16) yield $W = W(\mathcal{P}_W)$. ■

Note. Suppose that $L(\theta)$ is a quadratic function of θ such that $H(\theta) = H$ is a positive definite (stochastic or nonstochastic) matrix with $H \xrightarrow{p} J_0$. Then, $H_{LR} = \hat{H} = H_*^{\top} = H_* = H$ where H_{LR} and H_* are as in (A.1) and (A.3), respectively. Therefore, (A.2), $J_{LR} = b\hat{H}$, and (A.4) reduce to $b = 1$, $J_{LR} = H$, and $Z = I_p$, respectively.

Appendix B

Proof of Proposition 2. To show (20) to (22), four preliminary results are required; below, a row refers to one in Table 1. First, the row for (9) provides

$$g_{LR}(\tilde{\theta}_{LR}) = nJ_{LR}(\hat{\theta} - \tilde{\theta}_{LR}) \tag{B.1}$$

so substituting $\tilde{\theta}_{LR} = \tilde{\theta}$ on the right-hand side of (B.1) and sequentially using (A.7), $R_{LR} = \tilde{R}Z$, and (4) gives

$$g_{LR}(\tilde{\theta}_{LR}) = Z^{\top}g(\tilde{\theta}). \tag{B.2}$$

Second, since $r(\tilde{\theta}) = 0$, a mean value expansion of $r(\hat{\theta})$ at $\tilde{\theta}$ is

$$r(\hat{\theta}) = R_*(\hat{\theta} - \tilde{\theta}) \tag{B.3}$$

where R_* is $R(\theta)$ with each of its rows evaluated at a (possibly different) mean value given by a convex combination of $\hat{\theta}$ and $\tilde{\theta}$. Substituting (A.6) into (B.3) yields

$$\tilde{\lambda} = nV_*^{-1}r(\hat{\theta}) \tag{B.4}$$

where $V_* = R_*H_*^{-1}\tilde{R}^{\top}$ is assumed to be nonsingular and $V_* \xrightarrow{p} R_0J_0^{-1}R_0^{\top} = V_0$. Also, the row for (7) provides

$$r_{LR}(\hat{\theta}_{LR}) = R_{LR}(\hat{\theta}_{LR} - \tilde{\theta}) \tag{B.5}$$

so substituting $\hat{\theta}_{LR} = \hat{\theta}$ on the right-hand side of (B.5) and sequentially using (A.7) and (B.4) gives

$$r_{LR}(\hat{\theta}_{LR}) = V_{LR}V_*^{-1}r(\hat{\theta}) \tag{B.6}$$

where $V_{LR} = R_{LR}J_{LR}^{-1}R_{LR}^{\top} \xrightarrow{p} V_0$.

Third, the row for (7) provides

$$r_S(\hat{\theta}_S) = \tilde{R}(\hat{\theta}_S - \tilde{\theta}) \tag{B.7}$$

so substituting $\hat{\theta}_S = \tilde{\theta} + n^{-1}\tilde{J}^{-1}g(\tilde{\theta})$ on the right-hand side of (B.7) and sequentially using (4) and (B.4) gives

$$r_S(\hat{\theta}_S) = V_S V_*^{-1} r(\hat{\theta}) \tag{B.8}$$

where $V_S = \tilde{R}\tilde{J}^{-1}\tilde{R}^\top \xrightarrow{p} V_0$.

Fourth, let $V_W = \hat{R}\hat{J}^{-1}\hat{R}^\top \xrightarrow{p} V_0$ and

$$D = \hat{R}^\top V_W^{-1} R_* H_*^{-1} + I_p - \tilde{R}^\top V_*^{-1} R_* H_*^{-1}. \tag{B.9}$$

Then, it is easily seen that $D \xrightarrow{p} I_p$ and

$$D\tilde{R}^\top = \hat{R}^\top V_W^{-1} V_*. \tag{B.10}$$

Since $r_W(\hat{\theta}_W) = r(\hat{\theta})$, the first equality in (12) with $C = W$, $R_W = \hat{R}$, and $J_W = \hat{J}$ provides

$$g_W(\tilde{\theta}_W) = \hat{R}^\top V_W^{-1} V_* [n V_*^{-1} r(\hat{\theta})],$$

which, sequentially using (B.10), (B.4), and (4), gives

$$g_W(\tilde{\theta}_W) = Dg(\tilde{\theta}). \tag{B.11}$$

Assuming that $R_* H_*^{-1}\hat{R}^\top$ is nonsingular, a proof by contradiction shows that D is nonsingular. Therefore, suppose that D is singular. Then, there exists a $p \times 1$ vector $\xi \neq 0$ such that $D\xi = 0$, which (using (B.9)) can be written as

$$\hat{R}^\top V_W^{-1} R_* H_*^{-1}\xi + \xi - \tilde{R}^\top V_*^{-1} R_* H_*^{-1}\xi = 0. \tag{B.12}$$

Since $R_* H_*^{-1}\hat{R}^\top$ is nonsingular, premultiplying (B.12) by $R_* H_*^{-1}$ implies $R_* H_*^{-1}\xi = 0$ so (B.12) reduces to $\xi = 0$, the contradiction sought. Therefore, D is nonsingular.

To see the special cases in (20) to (22), first, let $U^{1/2}$ and $U^{-1/2}$ be matrix square roots of U and U^{-1}, respectively, where $U^{-1/2} = (U^{1/2})^{-1}$ and all the U-matrices are symmetric and positive definite. Then, given the definitions in (18), equations (15) and (16) can be written as

$$S(\mathcal{P}_C) = \left|\left| \tilde{J}^{1/2} J_C^{-1/2} g_C(\tilde{\theta}_C) \right|\right|^2_{\mathfrak{G}} \tag{B.13}$$

and

$$W(\mathcal{P}_C) = \left|\left|V_W^{1/2} V_C^{-1/2} r_C(\hat{\theta}_C)\right|\right|_{\mathfrak{R}}^2, \tag{B.14}$$

respectively, where $V_C = R_C J_C^{-1} R_C^\top$. Finally, appropriately substituting (B.2), (B.6), (B.8), (B.11), and relevant quantities from Table 1 into (14), (B.13), (B.14), and (19), it can be shown that (17) provides the special cases in (20) to (22) where $A_{LR} = \tilde{J}^{1/2} J_{LR}^{-1/2} Z^\top$ and $A_W = \tilde{J}^{1/2} \hat{J}^{-1/2} D$ are $p \times p$ nonsingular matrices such that $A_C \xrightarrow{p} I_p$ (for $C = LR, W$), and $B_{LR} = V_W^{1/2} V_{LR}^{1/2} V_*^{-1}$ and $B_S = V_W^{1/2} V_S^{1/2} V_*^{-1}$ are $r \times r$ nonsingular matrices such that $B_C \xrightarrow{p} I_r$ (for $C = LR, S$). ■

Note: Given (B.6), (B.8), and $r_W(\hat{\theta}_W) = r(\hat{\theta})$, it is easily seen that $r_C(\hat{\theta}_C) = 0$ iff $r(\hat{\theta}) = 0$; i.e., $\hat{\theta}_C \in \Theta_{0C}$ iff $\hat{\theta} \in \Omega_0$. However, by Assumption 1(a), $\hat{\theta} \notin \Omega_0$ so $\hat{\theta}_C \notin \Theta_{0C}$.

Appendix C

Proof of (34). Let $L^+(\beta)$ be a log-likelihood function where $\beta = \phi(\theta)$, $\beta \in B \subseteq \mathbb{R}^p$, and $\phi : \Omega \to B$ is a one-to-one transformation with the Jacobian $\Phi(\theta) = \partial\beta/\partial\theta^\top$, a nonsingular matrix for all $\theta \in \Omega$. Then,

$$\beta \in B_0 = \{\beta \,|\, r(\phi^{-1}(\beta)) = 0,\, \beta \in B\} \quad \text{iff} \quad \theta \in \Omega_0 = \{\theta \,|\, r(\theta) = 0,\, \theta \in \Omega\} \qquad \textbf{(C.1)}$$

where $\phi^{-1}(.)$ is the inverse transformation of $\phi(.)$. Now, let

$$L^*(\theta) = L^+(\phi(\theta)) \qquad \text{for all } \theta \in \Omega. \qquad \textbf{(C.2)}$$

Then, the first condition in (33) can be written as

$$L^+(\phi(\theta)) = L(\theta) \qquad \text{for only } \theta \in \Omega_0, \qquad \textbf{(C.3)}$$

the first condition in (34). This condition yields $\tilde{\beta} = \phi(\tilde{\theta})$ given (C.1) and the invariance property of ML estimators. To obtain the remaining two conditions in (34), first note that (C.2) provides

$$g^*(\theta) = \Phi(\theta)^\top g^+(\beta) = \Phi(\theta)^\top g^+(\phi(\theta)) \qquad \text{for all } \theta \in \Omega, \qquad \textbf{(C.4)}$$

and

$$J_0^* = \Phi_0^\top J_0^+ \Phi_0; \qquad \textbf{(C.5)}$$

cf. Davidson and MacKinnon (1993, equations (13.65) and (13.68), pp. 464-5). Then, evaluating (C.4) at $\tilde{\theta}$ yields

$$g^*(\tilde{\theta}) = \tilde{\Phi}^\top g^+(\tilde{\beta}) \qquad \textbf{(C.6)}$$

as $\tilde{\beta} = \phi(\tilde{\theta})$ given (C.3). Finally, the last two conditions in (34) are obtained by substituting (C.6) and (C.5) into the last two conditions in (33), respectively. ■

Appendix D

An example illustrating the specification of an appropriate $L^*(\theta)$. Breusch and Pagan (1979) have shown that, for testing the null hypothesis of homoskedastic errors in a linear regression model, an appropriate score statistic is invariant to certain formulations of the alternative hypothesis of heteroskedasticity. A special case of their model provides a simple example that illustrates the specification of an $L^*(\theta)$ that satisfies the G-W conditions. Therefore, for $t = 1, 2, \ldots, n$, let y_t be independently distributed as $N(0,\, h_t)$ variates where h_t is to be defined. Let θ_1 and θ_2 be scalars, $\theta = (\theta_1, \theta_2)^\top$, $\Omega = \{\theta \,|\, \theta_1 > 0,\ \theta_2 \in \mathbb{R}\}$, $\Omega_0 = \{\theta \,|\, \theta_1 > 0,\ \theta_2 = 0\}$, $\theta_0 = (\theta_{10}, 0)^\top$, and let $L(\theta)$ be the log-likelihood function obtained with $h_t = \theta_1 + \theta_2 z_t$ where z_t is a nonstochastic scalar. Then,

$$L(\theta) = L(\theta_1, \theta_2) = -\frac{n}{2}\ln\{2\pi\} - \frac{1}{2}\sum_{t=1}^{n}\ln\{\theta_1 + \theta_2 z_t\} - \frac{1}{2}\sum_{t=1}^{n}\frac{y_t^2}{\theta_1 + \theta_2 z_t},$$

$$\begin{pmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \end{pmatrix} = \begin{pmatrix} s_y^2 \\ 0 \end{pmatrix},\ g(\tilde{\theta}) = \frac{1}{2\tilde{\theta}_1^2}\begin{pmatrix} 0 \\ n(\bar{w} - \tilde{\theta}_1 \bar{z}) \end{pmatrix},\ \text{and } J_0 = \frac{1}{2\theta_{10}^2}\begin{bmatrix} 1 & \mu_z \\ \mu_z & \upsilon_z^2 \end{bmatrix} \tag{D.1}$$

where $s_y^2 = n^{-1}\sum_{t=1}^{n} y_t^2$, $\bar{w} = n^{-1}\sum_{t=1}^{n} y_t^2 z_t$, $\bar{z} = n^{-1}\sum_{t=1}^{n} z_t$, $\mu_z = \lim_{n\to\infty}\bar{z}$, and $\upsilon_z^2 = \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} z_t^2$.

Suppose that $L^*(\theta)$ is the log-likelihood function obtained with $h_t = \exp\{\theta_1 + \theta_2 z_t\}$ where the unrestricted and restricted parameter spaces are $\Omega^* = \mathbb{R}^2$ and $\Omega_0^* = \{\theta \,|\, \theta_1 \in \mathbb{R},\ \theta_2 = 0\}$, respectively. Then,

$$L^*(\theta) = -\frac{n}{2}\ln\{2\pi\} - \frac{n}{2}\theta_1 - \frac{n\bar{z}}{2}\theta_2 - \frac{1}{2}\sum_{t=1}^{n}\frac{y_t^2}{\exp\{\theta_1 + \theta_2 z_t\}} \tag{D.2}$$

so, using appropriate quantities in (D.1), it can be seen that $\tilde{\theta}_1^* = \ln\tilde{\theta}_1$, $\tilde{\theta}_2^* = 0$,

$$\tilde{\Phi}^\top g^*(\tilde{\theta}^*) = g(\tilde{\theta}), \qquad \text{and} \qquad \Phi_0^\top J_0^* \Phi_0 = J_0 \tag{D.3}$$

where $\tilde{\Phi} = \tilde{\theta}_1^{-1} I_2$ and $\Phi_0 = \theta_{10}^{-1} I_2$. Here, $S^* = n^{-1} g^*(\tilde{\theta}^*)^\top (\tilde{J}^*)^{-1} g^*(\tilde{\theta}^*)$ is a score statistic (for testing $H_0^* : \theta \in \Omega_0^*$) where $\tilde{J}^* \xrightarrow{p} J_0^*$. Therefore, although this $L^*(\theta)$ does not satisfy any one of the three G-W conditions, (D.3) provides the invariance result shown by Breusch and Pagan (1979) as $S^* = S$ when $\tilde{J}^* = (\tilde{\Phi}^\top)^{-1}\tilde{J}\tilde{\Phi}^{-1}$.

An $L^*(\theta)$ that satisfies the G-W conditions can be obtained by first replacing $\theta = (\theta_1, \theta_2)^\top$ with $\beta = (\beta_1, \beta_2)^\top$ on the right-hand side of (D.2), and then by specifying an appropriate transformation $\beta = \phi(\theta)$. To see this, let $L^*(\theta) = L^+(\beta)$ where $L^+(\beta)$ is the log-likelihood function obtained with $h_t = \exp\{\beta_1 + \beta_2 z_t\}$ and let

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \ln\{\theta_1\} \\ \theta_2/\theta_1 \end{pmatrix} = \phi(\theta), \qquad \Phi(\theta) = \frac{1}{\theta_1} \begin{bmatrix} 1 & 0 \\ -\theta_2\theta_1^{-1} & 1 \end{bmatrix},$$

$B = \mathbb{R}^2$, and $B_0 = \{\beta \,|\, \beta_1 \in \mathbb{R},\, \beta_2 = 0\}$. Note that $\Phi(\theta)$ is such that $\tilde{\Phi}$ and Φ_0 in (D.3) are equal to $\Phi(\tilde{\theta})$ and $\Phi(\theta_0)$, respectively. Then,

$$L^*(\theta) = L^*(\theta_1, \theta_2) = -\frac{n}{2}\ln\{2\pi\} - \frac{n}{2}\ln\theta_1 - \frac{n\bar{z}\theta_2}{2\theta_1} - \frac{1}{2\theta_1}\sum_{t=1}^{n} \frac{y_t^2}{\exp\{\theta_2 z_t/\theta_1\}},$$

which is just the right-hand side of (D.2) with θ_1 and θ_2 replaced with $\ln\theta_1$ and θ_2/θ_1, respectively. Here, $\tilde{\theta}^* = \tilde{\theta}$, $g^*(\tilde{\theta}) = g(\tilde{\theta})$, $J_0^* = J_0$, and $L^*(\theta) = L(\theta)$ for only $\theta \in \Omega_0$ as $L^*(\theta_1, 0) = L(\theta_1, 0)$ and $L^*(\theta) \neq L(\theta)$ for $\theta \notin \Omega_0$. Therefore, as required, this $L^*(\theta)$ satisfies the G-W conditions. Alternatively, this $L^*(\theta)$ can also be obtained by first specifying $L^+(\beta)$ as the log-likelihood function obtained with $h_t = \beta_1\exp\{\beta_2 z_t\}$, $B = \Omega$ and $B_0 = \Omega_0$, and then by using the transformation $(\beta_1, \beta_2) = (\theta_1,\, \theta_2/\theta_1)$, an example where $\tilde{\beta} = \tilde{\theta}$ although $\beta \neq \theta$.

Table 1. Quantities in the perceived frameworks of LR, S, and W

Equation	Quantity	Special case of C		
		LR	S	W
(13)	$\bar{\theta}_C$	$\hat{\theta}$	$\tilde{\theta}$	$\hat{\theta}$
	J_C	$b\hat{H}$	$\tilde{J}$	$\hat{J}$
	$\ddot{\theta}_C$	$\tilde{\theta}$	$\tilde{\theta}$	$\hat{\theta}$
	R_C	$\tilde{R}Z$	$\tilde{R}$	$\hat{R}$
(6)	$L_C(\theta)$	$L(\hat{\theta})$ $-\frac{n}{2}(\theta-\hat{\theta})^\top J_{LR}(\theta-\hat{\theta})$	$L(\tilde{\theta})+g(\tilde{\theta})^\top(\theta-\tilde{\theta})$ $-\frac{n}{2}(\theta-\tilde{\theta})^\top \tilde{J}(\theta-\tilde{\theta})$	$L(\hat{\theta})$ $-\frac{n}{2}(\theta-\hat{\theta})^\top \hat{J}(\theta-\hat{\theta})$
(7)	$r_C(\theta)$	$R_{LR}(\theta-\tilde{\theta})$	$\tilde{R}(\theta-\tilde{\theta})$	$r(\hat{\theta})+\hat{R}(\theta-\hat{\theta})$
(9)	$g_C(\theta)$	$nJ_{LR}(\hat{\theta}-\theta)$	$g(\tilde{\theta})-n\tilde{J}(\theta-\tilde{\theta})$	$n\hat{J}(\hat{\theta}-\theta)$
(10)	$\hat{\theta}_C$	$\hat{\theta}$	$\tilde{\theta}+\frac{1}{n}\tilde{J}^{-1}g(\tilde{\theta})$	$\hat{\theta}$
(11)	$\tilde{\theta}_C$	$\tilde{\theta}$	$\tilde{\theta}$	$\hat{\theta}-\hat{J}^{-1}\hat{R}^\top$ $\times\{\hat{R}\hat{J}^{-1}\hat{R}^\top\}^{-1}r(\hat{\theta})$

Note: In the third column, b is a particular positive scalar such that $b \xrightarrow{p} 1$, and Z is a particular $p \times p$ nonsingular matrix such that $Z \xrightarrow{p} I_p$.

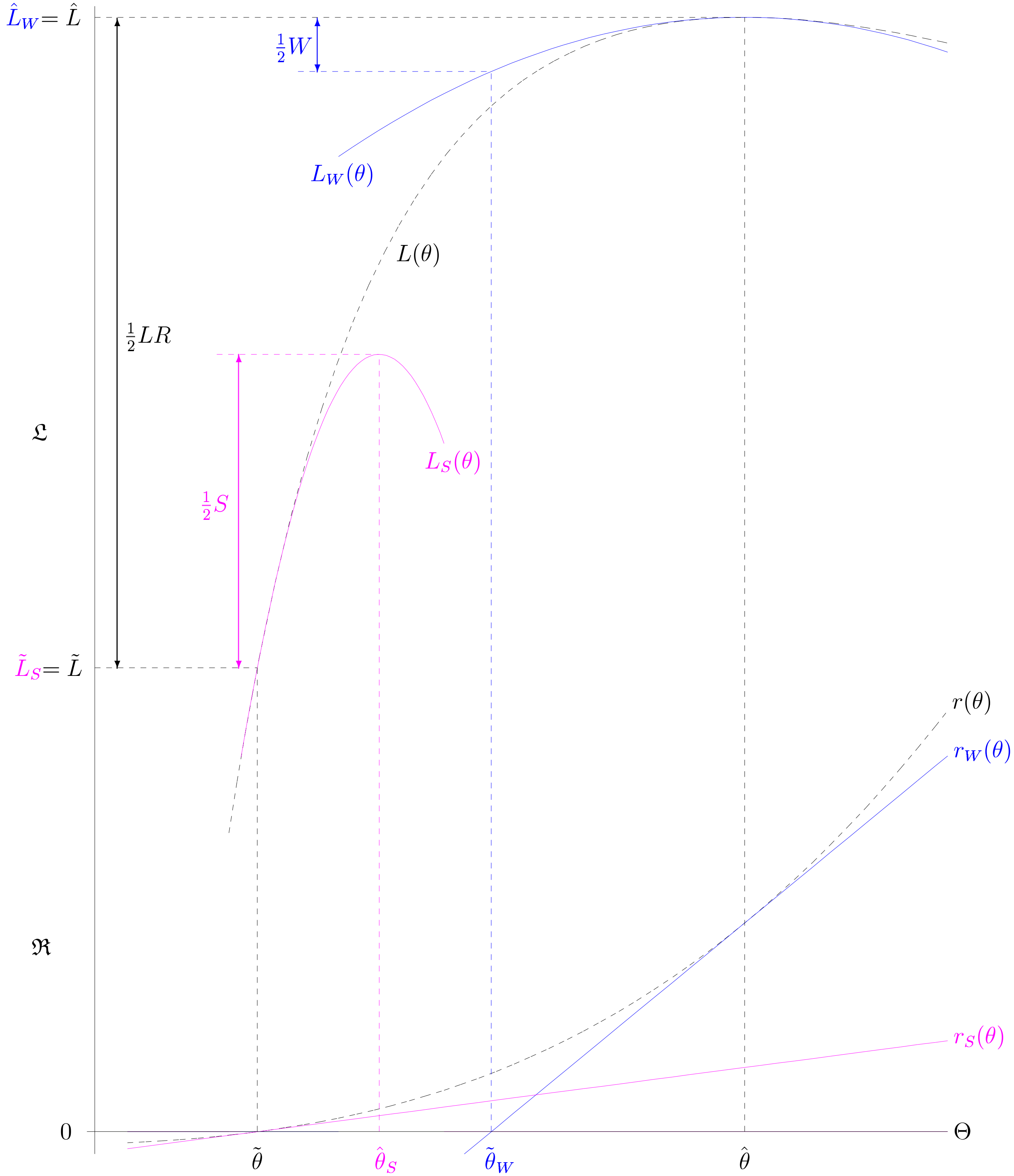

Figure 1. Classical statistics as lengths in $\mathfrak{L}$

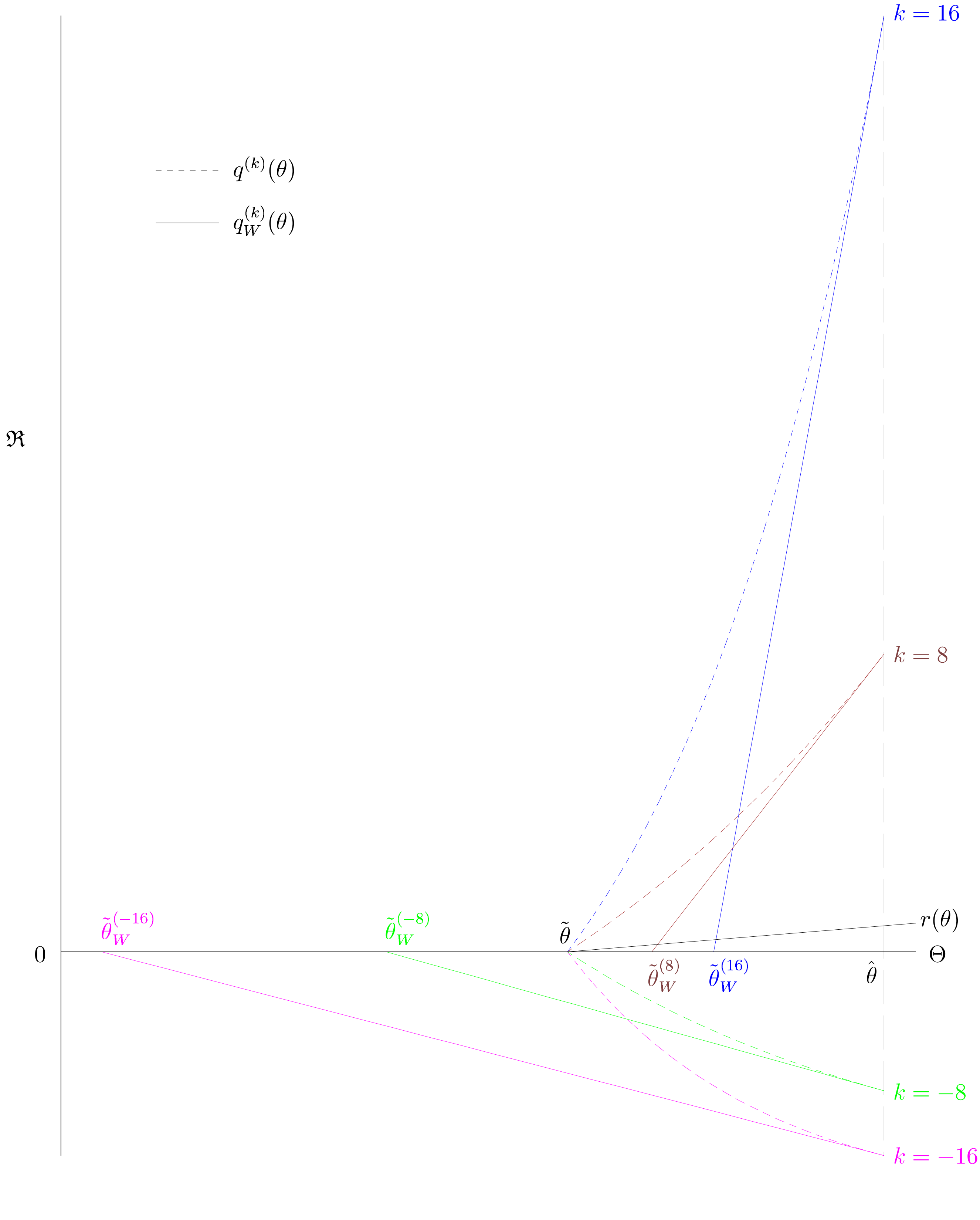

Figure 2. Perceived nulls of Wald statistics

Department of Economics, University of Alberta
Working Paper Series

http://www.economics.ualberta.ca/working_papers.cfm

2009-24: Tit-for-Tat Strategies in Repeated Prisoner's Dilemma Games: Evidence from NCAA Football – **Humphreys, Ruseski**
2009-23: Modeling Internal Decision Making Process: An Explanation of Conflicting Empirical Results on Behavior of Nonprofit and For-Profit Hospitals – **Ruseski,** Carroll
2009-22: Monetary and Implicit Incentives of Patent Examiners – **Langinier,** Marcoul
2009-21: Search of Prior Art and Revelation of Information by Patent Applicants – **Langinier,** Marcoul
2009-20: Fuel versus Food – **Chakravorty,** Hubert, Nøstbakken
2009-19: Can Nuclear Power Supply Clean Energy in the Long Run? A Model with Endogenous Substitution of Resources – **Chakravorty**, Magné, Moreaux
2009-18 Too Many Municipalities? – **Dahlby**
2009-17 The Marginal Cost of Public Funds and the Flypaper Effect – **Dahlby**
2009-16 The Optimal Taxation Approach to Intergovernmental Grants – **Dahlby**
2009-15 Adverse Selection and Risk Aversion in Capital Markets – Braido, da Costa, **Dahlby**
2009-14 A Median Voter Model of the Vertical Fiscal Gap – **Dahlby,** Rodden, **Wilson**
2009-13 A New Look at Copper Markets: A Regime-Switching Jump Model – Chan, **Young**
2009-12 Tort Reform, Disputes and Belief Formation – **Landeo**
2009-11 The Role of the Real Exchange Rate Adjustment in Expanding Service Employment in China – **Xu**, Xiaoyi
2009-10 "Twin Peaks" in Energy Prices: A Hotelling Model with Pollution and Learning – **Chakravorty**, Leach, Moreaux
2009-09 The Economics of Participation and Time Spent in Physical Activity – **Humphreys, Ruseski**
2009-08 Water Allocation Under Distribution Losses: Comparing Alternative Institutions – **Chakravorty**, Hochman, Umetsu, Zilberman
2009-07 A Comparative Analysis of the Returns on Provincial and Federal Canadian Bonds – **Galvani, Behnamian**
2009-06 Portfolio Diversification in Energy Markets – **Galvani**, **Plourde**
2009-05 Spanning with Zero-Price Investment Assets – **Galvani**, **Plourde**
2009-04 Options and Efficiency in Spaces of Bounded Claims – **Galvani**, Troitsky
2009-03 Identifying the Poorest Older Americans – Fisher, Johnson, **Marchand**, Smeeding, Boyle Torrey
2009-02 Time-Saving Innovations, Time Allocation, and Energy Use: Evidence from Canadian Households – **Brenčič, Young**
2009-01 Trigger Happy or Gun Shy? Dissolving Common-Value Partnerships with Texas Shootouts – Brooks, **Landeo**, Spier