Identification Robust Confidence Sets Methods for Inference on Parameter Ratios and their Application to Estimating Value-of-Time and Elasticities

Denis Bolduc[1] Université Laval

Lynda Khalaf[2] Université Laval

Clément Yélou[3] Université Laval

May 30, 2005

[1] *Groupe de recherche en économie de l'énergie, de l'environnement et des ressources naturelles* [GREEN], Université Laval. Mailing address: Pavillon J.-A.-De Sève, St. Foy, Québec, Canada, G1K 7P4. TEL: (418) 656-5427; FAX: (418) 656-2707; Email: denis.bolduc@ecn.ulaval.ca.

[2] *Groupe de recherche en économie de l'énergie, de l'environnement et des ressources naturelles* [GREEN], Université Laval. Mailing address: Pavillon J.-A.-De Sève, St. Foy, Québec, Canada, G1K 7P4. TEL: (418) 656 2131-2409; FAX: (418) 656 7412; Email: lkha@ecn.ulaval.ca.

[3] *Groupe de recherche en économie de l'énergie, de l'environnement et des ressources naturelles* [GREEN], Université Laval. Mailing address: Pavillon J.-A.-De Sève, St. Foy, Québec, Canada, G1K 7P4. Email: cyelou@ecn.ulaval.ca.

Abstract

The problem of constructing confidence set estimates for parameter ratios arises in a variety of econometrics contexts; these include value-of-time estimation in transportation research and inference on elasticities given several model specifications. Even when the model under consideration is identifiable, parameter ratios involve a possibly discontinuous parameter transformation that becomes ill-behaved as the denominator parameter approaches zero. More precisely, the parameter ratio is not identified over the whole parameter space: it is locally almost unidentified or (equivalently) weakly identified over a subset of the parameter space. It is well known that such situations can strongly affect the distributions of estimators and test statistics, leading to the failure of standard asymptotic approximations, as shown by Dufour (1997). Here, we provide explicit solutions for projection-based simultaneous confidence sets for ratios of parameters when the joint confidence set is obtained through a generalized Fieller approach. The procedures are applied and compared in illustrative simulated and empirical examples, with a focus on choice models.

Key words: confidence set; generalized Fieller's theorem; delta-method; Weak identification; parameter transformation.

Journal of Economic Literature classification: C10, C35, R40.

Contents

List of Tables

1 Introduction

The problem of constructing confidence set estimates for parameter ratios arises in a variety of econometrics contexts; these include value-of-time estimation in transportation research, or inference on elasticities in demand or cost analysis. Even when the model under consideration is identifiable, parameter ratios involve a possibly discontinuous parameter transformation that becomes ill-behaved as the denominator parameter approaches zero. More precisely, the parameter ratio is not identified over the whole parameter space: it is locally almost unidentified over a nonidentification subset of the parameter space. Important examples include inference on elasticities (t-statistics and confidence intervals) in demand systems [Deaton and Muellbauer (1980); Banks, Blundell and Lewbel (1997)], and inference on fixed value of time in discret choice transportation models (Bolduc (1999)). It is well known that such situations can strongly affect the distributions of estimators and test statistics, leading to the failure of standard asymptotic approximations, as shown by Dufour (1997, 2003).

The delta-method, which is an asymptotically justified Wald-type method, provides a common procedure to construct Wald-type confidence sets (CI) for ratios of parameters or ratios of linear combinations of parameters in econometric models. In the statistics literature, Fieller's theorem [Fieller (1940, 1954)] gives a simple way to obtain an exact confidence interval (CI) for the ratio of two means of normal variates. Scheffé (1970) proposes a modification of Fieller's procedure, which avoids *trivial* confidence set, *i.e.* confidence sets which cover the entire real line.[1] Zerbe, Laska, Meisner and Kushner (1982) extend Fieller's theorem in two directions. First, they focus on ratios of parameters in the normal linear regression model. Secondly, they construct multivariate confidence regions and simultaneous confidence sets for several ratios of linear combinations of parameters. In this case, normality still guarantees exact confidence levels. Young, Zerbe and Hay (1997) applies Zerbe et al. (1982)'s results to the context of linear and nonlinear mixed-effects models, in which case the distribution of estimators and test statistics are asymptotic.

Athough the solution provided by Fieller's method has been analyzed to some extent in the statistics litterature on location-scale, ANOVA and regression models [Darby (1980); Selwyn and Hall (1984); Buonaccorsi (1985); Bucephala and Gatsonis (1988); Zerbe (1978); Zerbe et al. (1982); Young et al. (1997)], its application to discret choice or limited dependent variable models is rather little documented. There is substantial evidence that standard asymptotics provides poor approximation to the sampling distribution of estimators and test statistics in discret choice or limited dependent variable models, even when linear hypothesis tests are of concern [see Davidson and MacKinnon (1999*b*), Davidson and MacKinnon (1999*a*), Davidson and MacKinnon (2000) and Savin and Würtz (1998)]. Furthermore, Dufour (1997) shows that most Wald-type confidence sets for a locally almost unidentified parameter in econometric models where the parameter space contains a nonidentification subset deviate arbitrarily from their nominal level, since they are almost surely bounded. In view of the recent literature on weak

[1] Scheffé (1970)'s procedure proceeds as follows: first test the null hypothesis that the numerator and the denominator are jointly equal to zero. If this test does not reject, state that both the numerator and the denominator jointly take on zero value. If the test rejects, state that the ratio is inside or outside a finite modified Fieller confidence set.

identification and weak instruments [Dufour (1997), Stock, Wright and Yogo (2002), Stock and Yogo (2002), Dufour (2003), Stock and Wright (2000)], there has been a renewed interest in an alternative method based on generalizing Fieller's theorem [Fieller (1940, 1954)]. In this paper, we consider Fieller-type simultaneous confidence sets for multiple ratio functions in econometric models under (2.1)-(2.2) below, with a focus on discret choice models. Our contributions can be classified into three categories.

First, we provide evidence based on two simulation studies that the delta method based confidence set for one parameter ratio in a discret choice model performs very poorly when the denominator approaches zero. One simulation study is based on a simple binary probit model, and the other one is based on a more complex model, a multinomial probit model with first-order generalized autoregressive errors [Bolduc (1992, 1999)].

Second, we use projection techniques to derive explicit form for simultaneous confidence sets for scalar linear transformations of a finite number of parameter ratios in general econometric models. Our characterization result shows that the confidence sets are not necessarily bounded, which implies that they will not suffer from the fundamental limitations documented in Dufour (1997). Our results hold asymptotically under mild regularity conditions and exactly for special cases. This extends work by Zerbe et al. (1982) beyond the normal linear regression model.

Third, the proposed procedures are applied to the transportation behavior analysis using the multinomial probit model specified and estimated in Bolduc (1999), where inference for three value of time ratios was relies on t-statistics based on the delta method.

The paper is organized as follows. In section 2, we set notation and introduce the statistical framework. In section 3, we discuss two methods for constructing a confidence interval for one parameters ratio and examine their statistical performance in illustrative discrete choice models. In section 4, we construct a Fieller-type joint confidence set for a finite number of ratios and then we derive projection-based simultaneous confidence sets. Empirical applications of the procedures are presented in section 5. Section 6 concludes.

2 Statistical Framework

In this section, we set notation and introduce the statistical framework. Consider the general parametric model

$$\left(\mathcal{Y}, \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p, p \geq 1\}\right), \tag{2.1}$$

where $\mathcal{Y}$ is the observations set, P_θ is a probability distribution over $\mathcal{Y}$ and $\theta = (\theta_1, \theta_2, ..., \theta_p)'$ is the parameter vector. The model is regular and identifiable, so based on a sample of size T, there exists a consistent and asymptotically normal estimator $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_p)'$of θ :

$$\left(\hat{\theta} - \theta\right) \longrightarrow \mathrm{N}\left(\mathbf{0}, \Sigma_{\hat{\theta}}\right), \quad T \to \infty, \tag{2.2}$$

where $\det\left(\Sigma_{\hat{\theta}}\right) \neq 0$. Let $\hat{\Sigma}_{\hat{\theta}}$ denote a consistent estimate of $\Sigma_{\hat{\theta}}$. For any constinuously differentiable function $g : \Theta^* \longrightarrow \mathbb{R}^q$, $(q \geq 1)$, where $\Theta^* \subseteq \Theta$, if $\hat{\theta} \in \Theta^*$, then $g\left(\hat{\theta}\right)$ is asymptotically normal with mean $g(\theta)$ and estimated variance matrix $\hat{\Sigma}_{g(\hat{\theta})}$ given by

$$\hat{\Sigma}_{g(\hat{\theta})} = \frac{\partial g\left(\hat{\theta}\right)}{\partial \theta'} \hat{\Sigma}_{\hat{\theta}} \frac{\partial g'\left(\hat{\theta}\right)}{\partial \theta}. \tag{2.3}$$

As a special case, any linear combination $L'\hat{\theta}$ of the elements of $\hat{\theta}$, where L is a known $p \times 1$ vector, is asymptotically normal with estimated variance

$$\hat{\Sigma}_{L'\hat{\theta}} = L'\hat{\Sigma}_{\hat{\theta}} L. \tag{2.4}$$

We consider Fieller-type confidence sets for ratio functions in econometric models under (2.1)-(2.2), with a focus on discret choice models. We first examine, through illustrative empirical models and Monte Carlo simulation studies, the poor performance of the delta method-based confidence set for one parameter ratio. The latter confidence set is a Wald-type method based on (2.3). Then, we consider the problem of *simultaneous confidence sets* for multiple ratios, in which case we propose to use a theory of quadric confidence sets in order to derive the explicit form of the simultaneous confidence limits for any scalar linear combination of these ratios.

So, our purpose is to build *simultaneous confidence sets* for scalar linear transformations of the components of vector-valued ratio functions $h : \Theta \longrightarrow \mathbb{R}^q, h(\theta) = (h_1(\theta), h_2(\theta), ..., h_q(\theta))'$. Individual confidence sets for several parameters are said to be simultaneous if they are constructed ensuring an overall confidence level control; see Miller (1981), Dufour (1989), Abdelkhalek and Dufour (1998).

Definition 1 *In the framework of model* 2.1-2.2*, let* $\{g_i(\theta) : i \in I\}$ *be a set of parameters defined as functions of* θ*, where the index set* I *may be finite or infinite and* $g_i(\theta) \in \mathbb{R}, \forall\ i \in I$ *and let* $CS_i \subset \mathbb{R}$ *be a confidence set for* $g_i(\theta), \forall\ i \in I$*. The sets* CS_i*,* $i \in I$*, constitute simultaneous confidence sets with level* $1-\alpha$ *for* $g_i(\theta), i \in I$ *if and only if*

$$\Pr\left(g_i(\theta) \in CS_i,\ i \in I\right) \geq 1-\alpha. \tag{2.5}$$

A key feature of a ratio function, e.g. $h_i(\theta), i = 1, ..., q$, is that it may display discontinuities in its domain Θ, so a reliable confidence set should be immune to such possible discontinuity problems. Specifically, the coverage probability should be close to the nominal confidence level, even when the true value of the parameter vector is in a discontinuity boundary.

We consider the case where $h_i(\theta) = L_i'\theta / K'\theta$, where L_i and K are known $p \times 1$ vectors, $i = 1, ..., q$. Ratios with the same denominator are encountered in many fields in economics; these include long run elasticities in dynamic demand models, and the economic value of time for several use-specific portions of travel time in transportation research. In this context, the discontinuity set for any h_i, $i = 1, ..., q$, is the set of all $\theta \in \Theta$ such that $K'\theta = 0$. Our setup covers simultaneous confidence sets for the individual ratios $h_i(\theta)$ or for linear combinations of $h_i(\theta), i = 1, ..., q$. For these cases, we apply projection techniques to a joint confidence region constructed for the vector $h(\theta) = (h_1(\theta), ..., h_q(\theta))'$.

3 Confidence Set for One Ratio of Parameters

In this section, we illustrate statistical problems associated with a confidence set constructed for one parameter ratio using the delta method. For convenience, we first give a brief discussion of two confidence set procedures, one based on the delta method and the other based on the Fieller's theorem, as they apply to the ratio $\delta(\theta) = \theta_1/\theta_2$ defined from model (2.1). Let

$$\hat{\Sigma}_{12} = \begin{bmatrix} \hat{v}_1 & \hat{v}_{12} \\ \hat{v}_{12} & \hat{v}_2 \end{bmatrix}$$

denote the submatrix of $\hat{\Sigma}_{\hat{\theta}}$ that corresponds to $\left(\hat{\theta}_1, \hat{\theta}_2\right)$.

3.1 The delta method and the Fieller-type confidence sets

The well-known delta-method relies on a first order Taylor series approximation for the ratio function $\delta(\theta) = \theta_1/\theta_2$ to obtain an estimate for the asymptotic variance of its maximum likelihood estimator $\hat{\delta} = \hat{\theta}_1/\hat{\theta}_2$. This estimated asymptotic variance is

$$\hat{\Sigma}_{\delta(\hat{\beta})} = \hat{G}'\hat{\Sigma}_{12}\hat{G},$$

where

$$\hat{G} = \left[\frac{1}{\hat{\theta}_2}, \quad -\frac{\hat{\theta}_1}{\hat{\theta}_2^2}\right]'.$$

To get a $(1-\alpha)$ level confidence set, the delta method yields the following Wald-type confidence set, using the asymptotic normal distribution critical point $z_{\alpha/2}$:

$$\mathrm{DCS}(\delta; 1-\alpha) = \left[\frac{\hat{\theta}_1}{\hat{\theta}_2} - z_{\alpha/2}\hat{\Sigma}^{1/2}_{\delta(\hat{\theta})}, \quad \frac{\hat{\theta}_1}{\hat{\theta}_2} + z_{\alpha/2}\hat{\Sigma}^{1/2}_{\delta(\hat{\theta})}\right]. \tag{3.6}$$

This confidence set is bounded and may therefore have zero *coverage probability*. In other words, the probability that this confidence set misses the true ratio may be practically one (Dufour (1997)).

On the other hand, Fieller' theorem, introduced in the context of the ratio of two means of normal variates where it leads to exact confidence sets, *inverts* a t-test of a linear restriction associated to the ratio. Inverting a test with respect to a parameter actually means that we collect all the values of this parameter for which the test is not significant. For the ratio $\delta(\theta) = \theta_1/\theta_2$ in the context of model (2.1)-2.2, it applies as follows. For each possible value δ_0 of the ratio, define the auxiliary hypothesis H_{δ_0}:

$$H_{\delta_0}: \quad \theta_1 - \delta_0\theta_2 = 0. \tag{3.7}$$

Then, a $(1-\alpha)$ level confidence set corresponds to the set of δ_0 for which an t-test of H_{δ_0} is not significant at level α. The test statistic in question is defined by:

$$t(\delta_0) = \frac{\left(\hat{\theta}_1 - \delta_0\hat{\theta}_2\right)}{\sigma_{\left(\hat{\theta}_1 - \delta_0\hat{\theta}_2\right)}}, \tag{3.8}$$

where

$$\sigma_{\left(\hat{\theta}_1-\delta_0\hat{\theta}_2\right)} = \left(\delta_0^2\hat{v}_2 - 2\delta_0\hat{v}_{12} + \hat{v}_1\right)^{1/2}$$

is an estimate of the variance of $\left(\hat{\theta}_1 - \delta_0\hat{\theta}_2\right)$. Under H_{δ_0},

$$t(\delta_0) \overset{asy}{\sim} \mathrm{N}(0,1).$$

So, Fieller' theorem gives a $(1-\alpha)$ level confidence set as the set of δ_0 such that

$$|t(\delta_0)| \leq z_{\alpha/2},$$

which leads to the following confidence set

$$\mathrm{FCS}(\delta; 1-\alpha) = \left\{\delta_0 : \left(\hat{\theta}_1 - \delta_0\hat{\theta}_2\right)^2 \leq z_{\alpha/2}^2\left(\hat{v}_1 + \delta_0^2\hat{v}_2 - 2\delta_0\hat{v}_{12}\right)\right\}. \tag{3.9}$$

This requires solving the following second degree polynomial inequality for δ_0:

$$A\delta_0^2 + 2B\delta_0 + C \leq 0, \tag{3.10}$$

where

$$\left\{\begin{array}{lcl} A & = & \hat{\theta}_2^2 - z_{\alpha/2}^2\hat{v}_2 \\ B & = & -\hat{\theta}_1\hat{\theta}_2 + z_{\alpha/2}^2\hat{v}_{12} \\ C & = & \hat{\theta}_1^2 - z_{\alpha/2}^2\hat{v}_1. \end{array}\right. \tag{3.11}$$

In appendix A, we present explicit solutions to the Fieller-type confidence set for one ratio of parameters, as defined by 3.9–3.11. These solutions show that the Fieller-type confidence set shares two basic properties. First, $\mathrm{FCS}(\delta; 1-\alpha)$ cannot be an empty set,[2] which is a useful property. Second, the Fieller-type confidence set for one ratio of parameters is either a bounded interval, an unbounded interval, or the entire real line $]-\infty, +\infty[$. The confidence set $\mathrm{FCS}(\delta; 1-\alpha)$ is an unbounded interval or the entire real line only when $\left|\hat{\theta}_2/\left(\hat{v}_2\right)^{1/2}\right| < z_{\alpha/2}$, i.e. when the Student's t-test of $H_0 : \theta_2 = 0$ is not significant is not significant at level α. Therefore, when the denominator is close to zero, the Fieller-type confidence set will give unbounded solutions, whereas the delta method still yields bounded confidence sets.

It is interesting to note that $\mathrm{FCS}(\delta; 1-\alpha)$ can remain informative even if it is unbounded. In particular, if we test $H_0 : \delta = r$, where r is any known scalar, and consider a decision rule which rejects H_0 when $r \notin \mathrm{FCS}(\delta; 1-\alpha)$, H_0 will be rejected at level α for all values of r not enclosed by the unbounded $\mathrm{FCS}(\delta; 1-\alpha)$.

3.2 Motivating experiments

To explore the feasibility of the Fieller-type and the delta method based confidence sets in discret choice contexts , we examine two illustrative examples. First, we present an empirical example based on Ben-Akiva and Lerman (1985, Chapter 7), and then we run a simulation study in a binary probit model.

[2] In section 4.2, we prove in corollary 4 the non-emptyness property for the general case of simultaneous confidence sets for scalar linear transformations of ratios of parameters.

3.2.1 A trinomial logit model of travel demand

The first example we present is an application of both procedures to estimating the value of time in a three-alternative logit mode choice model analyzed in Ben-Akiva and Lerman (1985, Chapters 3, 5 and 7.).

The model is specified as follows. The universal choice set consists of three modes to work: driving alone, sharing a ride, transit bus. Each worker n has a feasible choice set, denoted by C_n, that has $J_n \leq 3$ feasible choices.[3] Let $U_{in} = V_{in} + \varepsilon_{in}$ denote the real-valued utility index associated with alternative $i \in C_n$ for individual n, where V_{in} is the systematic component of the utility and ε_{in} is the random component. Alternative $i \in C_n$ is choosen by individual n if and only if $U_{in} \geq U_{jn}$ for all $j \neq i$, $j \in C_n$. The probability that alternative $i \in C_n$ is choosen by individual n is given by

$$\begin{aligned} P_n(i) &= \Pr(U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i) \\ &= \Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in C_n, j \neq i) \\ &= \Pr(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}, \forall j \in C_n, j \neq i). \end{aligned}$$

The multinomial logit (MNL) model is obtained as

$$P_n(i) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}}, \forall i \in C_n \tag{3.12}$$

and corresponds to independently and identically Gumbel-distributed ε_{in}, $i \in C_n$, with a scale parameter equal to one. This model is estimated assuming linear-in-parameters functions for the deterministic components V_{in} and a single vector of coefficients θ that applies to all the utility functions. The utility U_{in} takes the following form

$$U_{in} = \theta' X_{in} + \varepsilon_{in}, \tag{3.13}$$

where X_{in} is a vector describing the attributes of alternative i for individual n. This leads to

$$P_n(i) = \frac{e^{\theta' X_{in}}}{\sum_{j \in C_n} e^{\theta' X_{jn}}}, \forall i \in C_n.$$

The variables X_{in} include two alternative-specific constants, three generic attributes of the travel modes and seven alternative-specific socioeconomic and locational characteristics of worker n. The three variables for generic attributes of the travel modes are:[4] round trip travel time (the sum of in-vehicle and out-of vehicle times), (round trip out-of vehicle time)/(one-way distance), and (round trip travel cost)/(household income). Their coefficients in the functions V_{in} will be denoted respectively by $\theta_3, \theta_4, \theta_5$. This model was estimated by maximum likelihood using data for a sample of 1136 workers taken from a 1968 survey in the Washington, D.C., metropolitan area.

[3] The rules used for determining which subset of the three potential alternatives is feasible for each worker were entirely judgmental. See Ben-Akiva and Lerman (1985, chapter 5) for details.

[4] See Ben-Akiva and Lerman (1985, Table 7.1, page 158) for details on the other variables.

In this model, the ratio of two coefficients θ_i and θ_j of the utility function (3.13) provides information about the marginal rate of substitution between the corresponding variables. The economic value of travel time can then be defined as the marginal rate of substitution between the time and cost variables. In particular, since round trip travel time is the sum of in-vehicle and out-of vehicle times, the value of total travel time is equal to that of in-vehicle time and is given by

$$\delta_{\text{tot}} = \frac{\theta_3}{\theta_5} \times (\text{household income})\,. \tag{3.14}$$

Similarly, the value of out-of-vehicle time is

$$\delta_{\text{out}} = \left[\frac{\theta_3}{\theta_5} + \frac{\theta_4}{\theta_5 \times (\text{one-way distance})}\right] \times (\text{household income})\,. \tag{3.15}$$

Ben-Akiva and Lerman (1985) computed point estimates for the two parameter functions δ_{tot} and δ_{out}. Let $h_1(\theta) = \theta_3/\theta_5$ and $h_2(\theta) = \theta_4/\theta_5$. If θ_5 is close to zero, then the functions δ_{tot} and δ_{out} will be weakly identified. Here, we give 95%-level confidence sets for the ratios $h_1(\theta)$ and $h_2(\theta)$, using the delta method and the Fieller-type procedures.

The delta method yields

$$\begin{aligned} \text{DCS}(\theta_3/\theta_5; .95) &= [-.0002089, \quad .0023483] \\ \text{DCS}(\theta_4/\theta_5; .95) &= [-.1974734, \quad 1.1382400], \end{aligned} \tag{3.16}$$

whereas the Fieller-type method gives

$$\begin{aligned} \text{FCS}(\theta_3/\theta_5; .95) &= [-\infty, \quad -.0151209] \cup [.0003947, \quad +\infty] \\ \text{FCS}(\theta_4/\theta_5; .95) &= [-\infty, \quad -7.2190631] \cup [.0826500, \quad +\infty]\,. \end{aligned} \tag{3.17}$$

This example illustrates a situation where a Fieller type confidence set is unbounded and is in conflict with the one based on the delta-method. We emphasize that although $\text{FCS}(\theta_3/\theta_5; .95)$ and $\text{FCS}(\theta_4/\theta_5; .95)$ are unbounded, they remains informative. For instance, if we test $H_0 : \theta_3/\theta_5 = 0$ or $H_0 : \theta_4/\theta_5 = 0$ using the derived confidence sets as mentionned above, the unbounded Fieller-type confidence sets (3.17) are indeed quite informative and lead to rejection of H_0 as expected for the economic value of travel time. In contrast, using the confidence set based on the delta-method, H_0 is not rejected, which is counter intuitive since this implies that travel time may have a zero economic value. As pointed out above, the former method is more likely to give confidence sets robust to severe size problems as documented in Dufour (1997). As a result, an unbounded confidence set may be quite informative and reliable, whereas a bounded confidence set may fail to cover the true parameter value.

3.2.2 Simulation study I: a binary probit model

In this example, we consider the simple binary probit model specified as follows:

$$
\begin{aligned}
y_n^* &= \theta_1 + \theta_2 x_{2n} + \theta_3 x_{3n} + u_n \\
y_n &= \begin{cases} 0 & \text{if} \quad y_n^* < 0 \\ 1 & \text{otherwise} \end{cases} \qquad (3.18) \\
u_n &\sim i.i.d.\ N(0,1),
\end{aligned}
$$

where for individual n, x_{2n} and x_{3n} are observations on explanatory variables, y_n^* is the latent (unobservable) variable that may represent utility, y_n is the observed choice, u_n is the error term assumed to be identically and independently distributed as a standard normal, and $\theta = (\theta_1, \theta_2, \theta_3)'$ is the parameter vector. The model (3.18) is estimated by the method of maximum likelihood, which gives a consistent and asymptotically normal estimator $\hat{\theta}$ for θ. The aim is to run a simulation study to assess the coverage rate properties of the two confidence set procedures, the delta method and the Fieller-type method, as they apply to the ratio $\delta = \theta_2/\theta_3$ in model (3.18) when the denominator approaches zero.

The design used in this simulation study is as follows. The regressors x_2 and x_3 and the error term u are drawn from three independent $\mathrm{N}(0,1)$ variates. The parameters are set to $\theta_1 = 1$, $\theta_2 = 3.3$ and θ_3 varies from 2 to 0.0001; the sample size T is set to $T = 100, 250, 1000, 5000$, and 10000. We construct 95%-level confidence sets for δ, using the delta method and the Fieller-type method. Based on 10000 replications, we compute the empirical coverage rate for both procedures. Simulation results are shown in Table **??**.

These results show that the empirical coverage rate of the delta method based confidence set deteriorates rapidly as the denominator becomes close to zero, no matter how large the sample size. Especially, when the denominator value is lower than 0.1, the empirical coverage rate deviates markedly from the nominal confidence level. In contrast, the Fieller-type method, although it is approximate in our application, does not suffer from such problems. The poor performance of the delta method based confidence set might be more serious in many empirical models where specified discret choice models are more complex. As a result, confidence sets based on the delta method should be avoided, while the Fieller-type method is more appealing.

3.2.3 Simulation study II: a multinomial probit model with logit kernel

We consider a more complex formulation of discrete choice, i.e. the *multinomial probit with logit kernel model*. Since the properties of standard asymptotics in this class of models are little documented, it is important to assess the performance of both confidence set procedures within this framework. The model can be described as follows.

Each individual denoted by $n = 1, \ldots, T$ in a population of size T faces J discrete alternatives (or choices) of a choice set $\mathcal{C}$. The observed choice made by individual n is denoted by $i_n \in \mathcal{C}$, and X_n is the $(\mathcal{J} \times \mathcal{K})$ matrix of explanatory variables associated with individual n; these variables include socio-economic variables, alternatives characteristics as well as different types of interactions. X_{jn} denotes the j-th row of X_n. The vector of unknown parameters to be

Table 1: Empirical coverage rates for the delta method- and the Fieller method- based confidence sets for a parameter ratio in a simple binary probit model.

T	100		250		1000		5000		10000	
θ_3	DCS	FCS	DCS	FCS	DCS	FCS	DCS	FCS	DCS	FCS
2	93.30	95.33	94.71	95.29	95.09	95.38	95.05	95.11	94.84	94.93
1	90.97	95.82	94.01	95.19	94.97	95.13	95.08	95.06	94.84	95.01
0.5	88.58	95.50	91.11	95.33	94.38	94.87	94.86	94.92	95.14	94.96
0.4	89.00	95.48	90.65	94.90	93.91	94.93	94.68	94.69	94.81	94.88
0.3	82.11	95.34	89.93	94.80	92.81	95.09	94.98	95.01	94.77	94.75
0.2	79.15	95.77	85.48	95.12	91.36	94.97	94.07	94.61	95.10	95.07
0.1	61.59	95.79	72.97	95.08	85.82	95.13	91.81	94.99	93.16	95.22
10^{-2}	22.03	95.66	30.44	94.89	41.07	95.50	56.86	95.09	64.47	94.96
10^{-3}	06.79	95.69	10.17	94.99	13.90	94.89	19.56	94.54	24.15	94.92
10^{-4}	02.42	95.67	02.98	95.19	04.29	94.79	06.38	95.41	07.71	95.34

Note: Numbers reported are empirical coverage rates for the confidence set based on the delta method [in the columns titled "DCS"] and for the one based on the Fieller's method [in the columns titled "FCS"]. θ_3 is the denominator of the ratio and T is the sample size. The nominal confidence level is 95%.

estimated is denoted by $\theta = (\beta', \bar{\beta}')'$, where the sub-vector β, of dimension $(\mathcal{K} \times 1)$, denotes the parameters associated with X_n and the sub-vector $\bar{\beta}$ contains the nuisance parameters. We write the discrete choice model for individual n as:

$$\varsigma_{i,n} = \begin{cases} 1 & \text{if individual } n \text{ chooses alternative } i \\ 0 & \text{otherwise.} \end{cases} \tag{3.19}$$

$$U_{in} = X_{in}\beta + \varepsilon_{in}, \quad i = 1, 2, \ldots, \mathcal{J}, \tag{3.20}$$

where U_{in} is the indirect utility indicator associated with alternative i for individual n. For convenience, we write this model in the following compact form:

$$U_n = X_n\beta + \varepsilon_n,$$

where $U_n = (U_{1,n}, U_{2,n}, \ldots, U_{\mathcal{J},n})'$ and $\varepsilon_n = (\varepsilon_{1,n}, \varepsilon_{2,n}, \ldots, \varepsilon_{\mathcal{J},n})'$ are $\mathcal{J} \times 1$ vectors. For further reference, let X denote the matrix that concatenates vertically the individual matrices X_n for $n = 1, \ldots, T$. The alternative i is chosen by individual n if and only if $U_{in} \geq U_{jn}, \forall j \in \mathcal{C}$; the vector $\varsigma_n = (\varsigma_{1,n}, \varsigma_{2,n}, \ldots, \varsigma_{\mathcal{J},n})'$ gives the observed choice made by individual n. Therefore, the choice probability $P_n(i)$ associated with the alternative i, $i \in \mathcal{C}$ chosen by individual n is defined by:

$$P_n(i) = P(U_{in} \geq U_{jn}, \forall j \in \mathcal{C}). \tag{3.21}$$

The computation burdens of the choice probability (3.21) depend on the distribution assumed for the error term ε_n. For example, assuming $\varepsilon_n \overset{i.i.d.}{\sim} N(0, \Sigma)$ gives the Multinomial Probit

(MNP) model. In this case $P_n(i)$ requires the evaluation of multi-dimensional integrals, which may be analytically untractable for large choice sets; in particular, when the choice set involves four or more alternatives, the choice probabilities are usually simulated. Assuming $\varepsilon_n \overset{i.i.d.}{\sim}$ Gumbel leads to the Multinomial Logit (MNL) model, in which case the choice probabilities have a simple to compute explicit form. In this simulation study, we consider the kernel logit model formulation that results from an attractive combination of MNP and MNL (see Ben-Akiva, Bolduc and Walker (2001)):

$$U_n = X_n\beta + \varepsilon_n \tag{3.22}$$

$$\varepsilon_n = W\xi_n + \nu_n,$$

$$W = FG, \tag{3.23}$$

$$\xi_n \overset{i.i.d}{\sim} N(0, I_{\mathcal{J}}) \text{ and } \nu_n \overset{i.i.d}{\sim} \text{ Gumbel},$$

where G is a diagonal matrix of dimension $(\mathcal{J}\times\mathcal{J})$ that have the standard deviation terms of the components ε_n of its main diagonal, and the matrix F captures the correlation structure among the error terms. When ξ_n is known, this model reduces to the usual MNL model specification; then the probability $P_n(i)$ that individual n chooses alternative i conditionnally to ξ_n can be written as:

$$P_n(i|\xi_n) = \frac{\mathrm{e}^{X_{in}\beta + W_i\xi_n}}{\sum_{j=1}^{\mathcal{J}} \mathrm{e}^{X_{jn}\beta + W_j\xi_n}}. \tag{3.24}$$

Then, we obtain the unconditional choice probability of alternative i by integrating $P_n(i|\xi_n)$ over the domain of ξ_n:

$$P_n(i) = \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{\mathcal{J}} \Lambda(i|\xi_n) n(\xi_n; 0, I_{\mathcal{J}})\, \mathrm{d}\xi_n. \tag{3.25}$$

Expression in equation (3.25) shows that $P_n(i)$ is a $\mathcal{J}$-dimensional unbounded integral. In our experiment, we consider $\mathcal{J} = 3$. So, we have been able to use numerical integration to compute this tri-dimensional integral, athough it is very computer-time demanding. McFadden (1989) suggested the simulated maximum likelihood approach, where the multivariate integral is replaced by an approximation obtained by simulations. This approach requires draws from the distribution of ξ_n. Using $\mathcal{S}$ independent draws, the empirical mean

$$\hat{P}_n(i) = \frac{1}{\mathcal{S}}\sum_{r=1}^{\mathcal{S}} \Lambda(i|\xi_n^r), \tag{3.26}$$

where ξ_n^r denotes a given draw r from the distribution of ξ_n, is an unbiased and consistent estimator for the choice probability $P_n(i)$. Then, replacing the choice probability in the log likelihood with the simulator $\hat{P}_n(i)$ and maximizing the simulated likelihood function

$$\hat{L}(\theta) = \sum_{n=1}^{T} \ln \hat{P}_n(i|\theta). \tag{3.27}$$

lead to simulated maximum likelihood (SML) estimator. SML estimators are known to be consistent and asymptotically efficient under mild regularity conditions. Efficiency requires that the sample size and the number of replications $\mathcal{S}$ used to compute the probability simulator both are large.

When estimating SML based logit kernel models, there may be two important problems namely the non-identification of the parameter vector, and the bias associated with simulating the log likelihood function. Walker (2001) highlights the fact that the Gumbel *i.i.d.* term leads to extra identification conditions that impose restrictions on the matrices F and G. On the other hand, using a large number of random draws $\mathcal{S}$ helps reducing the simulation bias; for instance Bolduc (1999) suggests that with $\mathcal{S} = 50$ draws, the estimation results are very close to those obtained with a larger number of draws.

The model is estimated using the method of maximum likelihood. The choice probabilities $P_n(i)$ are computed using the simulator defined by (3.26). The design of our simulation study is as follows. X_n is composed of $\mathcal{K} = 5$ variables that are drawn as 1.5 times independent $U_{[0 \quad 1]}$ where $U_{[a \quad b]}$ denotes the uniform distribution [a b], and $\mathcal{J} = 3$ alternatives. We consider the ratio $\delta^* = \theta_2/\theta_3$. The parameters β of the utility indicator U_n are set as follows: $\beta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)'$ and $\theta_i = 3$,for $i = 1, 2, 4, 5$ in all the experiment, whereas θ_3, the denominator of the ratio δ^* varies from 3 to 0.0001. The sample size T is set to $T = 1000, 5000$, and 10000. For the simulated choice probability method, we use $\mathcal{S} = 50$ draws to evaluate the simulator (3.26), while we use 12 integration points for the numerical integration method. We construct 95%-level confidence sets for δ^*, using the delta method and the Fieller-type method and compute empirical coverage rates based on 1000 replications. Table 2 reports the empirical coverage rates for both procedures.

4 Simultaneous Confidence Sets for Multiple Ratios of Parameters

Let us consider, in the context of model (2.1-2.2), $s \leq p - 1$ ratios of parameters ρ_i, $i = 1, ..., s$ with a common denominator $K'\theta$:

$$\rho_i = h_i(\theta) = L_i'\theta / K'\theta, \forall i = 1, ..., s, \tag{4.28}$$

where $\{L_1, L_2, ..., L_s, K\}$ is a linearly independent set of fixed (nonstochastic) $p \times 1$ vectors.[5] These s ratio functions have the same discontinuity set D_h defined by

$$D_h = \left\{\theta \in \Theta : K'\theta = 0\right\}. \tag{4.29}$$

Clearly, $D_h \neq \emptyset$ since $\theta = (0, ..., 0) \in D_h$.

[5] Observe that if $s \geq p$, then $\{L_1, L_2, ..., L_s, K\}$ are linearly dependent. Indeed, if $s > p$, then it is always possible to express at least $s - p$ elements of the set $\{L_1, L_2, ..., L_s\}$ as a linear combination of the others, and if $s = p$, then K is expressible as a linear combination of $L_1, ..., L_s$.

Table 2: Empirical coverage rates for the delta method- and the Fieller method- based confidence intervals for a parameter ratio in a kernel logit multinomial probit model.

T	1000		5000		10000	
θ_3	DCS	FCS	DCS	FCS	DCS	FCS
3	94.3	95.1	95.2	94.6		
2	94.8	95.4				
1	93.3	93.7	94.4	94.5		
0.5	90.5	93.9	95.1	94.1		
0.3	87.8	94.7				
0.2	82.2	94.1	90.6	95.1		
0.1	69.2	93.5				
10^{-2}	25.9	93.4	37.4	94.4		
10^{-3}	7.6	93.9	13.1	94.7		
10^{-4}	2.6	94.0				

Note: Numbers reported are empirical coverage rates for the confidence set based on the delta method [in the columns titled "DCS"] and for the one based on the Fieller's method [in the columns titled "FCS"]. θ_3 is the denominator of the ratio and T is the sample size. The nominal confidence level is 95%.

We aim to construct simultaneous confidence sets for the s ratios defined in (4.28) as well as for any linear combination of these ratios,

$$l_w(\rho_1, \rho_2, ...\rho_s) = \sum_{i=1}^{s} w_i \rho_i, \tag{4.30}$$

where $w = (w_1, w_2, ..., w_s)'$ is any known nonstochastic (fixed) $s \times 1$ vector. We provide explicit solutions for these simultaneous confidence limits.

Zerbe et al. (1982) construct simultaneous confidence limits for several ratios of linear combinations of parameters in a normal linear regression model using an analysis of variance method proposed in Scheffé (1959, 1953).[6] They show that for each ratio, these confidence limits are solutions to a quadratic equation and take the form (3.9). In addition, Zerbe et al. (1982) construct a joint confidence set for a finite number of ratios of linear combinations of regression coefficients with a common denominator and claimed that projections of this joint confidence region on the individual ratios' axes yield exactly the simulatneous confidence limits obtained through the Scheffé's method of analysis of variance.

In this section, we use results on *quadric* confidence sets (see Dufour and Taamouti (2003))

[6] In a normal linear regression model, this method gives simultaneous Wald type confidence sets for linear transfomations of a finite number of independent linear combinations of regression coefficients. The method applies to a ratio using a Fieller-type transformation, as in (3.7). It applies to the more general case where the ratios need not have the same denominator.

and derive simultaneous confidence limits for any linear combination of the s ratios defined in (4.28).

4.1 Joint confidence set for a finite number of parameters ratios

We define the following s linear combinations associated with the s ratios ρ_i, $i = 1, ..., s$, as in (3.7):

$$L_i'\theta - \rho_i K'\theta = 0, \qquad i = 1, ..., s. \tag{4.31}$$

Let

$$\begin{aligned} H &= \begin{bmatrix} L_1 & . & . & . & L_s & K \end{bmatrix}' \\ R &= \begin{bmatrix} & & I_s & & \\ -\rho_1 & . & . & . & -\rho_s \end{bmatrix}', \end{aligned} \tag{4.32}$$

where I_s is the s-dimensional identity matrix. The $s \times (s+1)$ matrix R has full row rank for any possible values for ρ_i, $i = 1, ..., s$; and since the set of $s+1$ vectors $\{L_1, L_2, ..., L_s, K\}$ is linearly independent, the matrix H has full row rank. Therefore, the s equations in (4.31) imply s non-redundant restrictions that we write in the form $RH\theta = 0$.

In order to obtain a joint confidence region for $\rho = (\rho_1, ..., \rho_s)'$ we propose, as in Zerbe et al. (1982) and Young et al. (1997), to invert a Wald test for the restrictions

$$H_0 : RH\theta = 0. \tag{4.33}$$

A $(1-\alpha)$ level confidence region for ρ, $\mathrm{CS}(\rho; 1-\alpha)$, is the set of all ρ such that the latter test is not significant at level α.

Let W_{RH} denote the Wald statistic defined to test (4.33):

$$W_{RH} = \left(RH\hat{\theta}\right)' \left(RH\hat{\Sigma}_{\hat{\theta}}H'R'\right)^{-1} \left(RH\hat{\theta}\right).$$

In our context, W_{RH} has an asymptotic $\chi^2(s)$ null distribution; let c_α be the $(1-\alpha)$ percentile point of the $\chi^2(s)$ distribution. Then, we define $\mathrm{CS}(\rho; 1-\alpha)$ as:

$$\mathrm{CS}(\rho; 1-\alpha) = \{\rho \in \mathbb{R}^s : W_{RH} \leq c_\alpha\}. \tag{4.34}$$

Zerbe et al. (1982) consider the following orthogonal decomposition that allows to characterize the analytical form of $\mathrm{CS}(\rho; 1-\alpha)$:

$$\begin{aligned} W_{RH} = &\left(H\hat{\theta}\right)' \left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} \left(H\hat{\theta}\right) - \\ &\left[(\rho', 1)\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} H\hat{\theta}\right]' \left[(\rho', 1)\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} (\rho', 1)'\right]^{-1} \left[(\rho', 1)\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} H\hat{\theta}\right]. \end{aligned} \tag{4.35}$$

Substituting (4.35) in (4.34) and rearranging terms then yields:

$$\Pr\{\rho_*' M \rho_* \leq 0\} = 1 - \alpha, \tag{4.36}$$

where

$$\rho_* = (\rho', 1)' \tag{4.37}$$

$$M = c\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} - \left[\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} H\hat{\theta}\right]\left[\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} H\hat{\theta}\right]'$$

$$c = \left(H\hat{\theta}\right)'\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1}\left(H\hat{\theta}\right) - c_\alpha.$$

Therefore, (4.34) is written as:

$$\mathrm{CS}\,(\rho; 1-\alpha) = \left\{\rho \in \mathbb{R}^s : \rho_*' M \rho_* \leq 0, \quad \rho_* = (\rho', 1)'\right\}. \tag{4.38}$$

The set $\mathrm{CS}\,(\rho; 1-\alpha)$ may take different forms, which depend on whether the common denominator of the ratios is statistically different from zero or not [Scheffé (1970), Zerbe et al. (1982), Young et al. (1997)]: the interior of an s-dimensional ellipsoid, or a hyperboloid, or the entire s-dimensional vector space $\mathbb{R}^s$. We use a theory of quadric confidence sets and derive explicit form for the *projection*-based simultaneous confidence sets for any scalar linear transformation of the s ratios with common denominator, $l_w(\rho) = w'\rho$, where w is a known fixed (nonstochastic) vector.

4.2 Explicit solutions for simultaneous confidence sets for linear transformations of ratios

We characterize simultaneous confidence sets for linear transformations of a finite number of ratios using the quadric confidence set theory, as developped in (Dufour and Taamouti (2003)). The set of points that satisfy an equation of the form $\rho'\Gamma\rho + \beta'\rho + \gamma = 0$, where Γ is a symmetric $s \times s$ matrix, β is a $s \times 1$ vector and γ is a scalar, constitutes a quadric surface. A confidence set for ρ of the form

$$C_\rho = \left\{\rho_0 : \rho_0'\Gamma\rho_0 + \beta'\rho_0 + \gamma \leq 0\right\}$$

is a quadric confidence set (Dufour and Taamouti (2003)). Depending on the values of Γ, β, and γ, it may take several forms, including ellipsoids, paraboloids and hyperboloids.

The confidence set $\mathrm{CS}\,(\rho; 1-\alpha)$ defined in (4.38) can be written in the form of a quadric conficence set. Indeed, since $\rho_* = (\rho', 1)'$, partition the matrix M accordingly in the form:

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \tag{4.39}$$

where M_{11} is an $s \times s$ matrix, M_{12} is a $s \times 1$ vector, $M_{21} = M_{12}'$, and M_{22} is a scalar. Let $S_1 = (I_s \sim \mathbf{0}_{s\times 1})$ be an $s \times (s+1)$ matrix and $S_2 = (0, ..., 0, 1)$ a $1 \times (s+1)$ vector. Then, the following relations hold:

$$\rho = S_1\rho_*, \qquad M_{11} = S_1 M S_1',$$
$$M_{22} = S_2 M S_2'.$$

The quadratic form $\rho_*' M \rho_*$ is equivalently expressed as:

$$\rho_*' M \rho_* = \rho' M_{11} \rho + 2M_{12}' \rho + M_{22}.$$

Thus, $\mathrm{CS}(\rho; 1-\alpha)$ is written as the following quadric confidence set:

$$\mathrm{CS}(\rho; 1-\alpha) = \left\{ \rho \in \mathbb{R}^s : \rho' M_{11} \rho + 2M_{12}' \rho + M_{22} \leq 0 \right\}. \tag{4.40}$$

The joint confidence region $\mathrm{CS}(\rho; 1-\alpha)$ is multidimensional and may be hard to interpret in practical applications. So, it is more convenient to derive confidence sets for individual ratios or for scalar linear transformations of them. We apply the projection technique to the quadric confidence set $\mathrm{CS}(\rho; 1-\alpha)$ (see Dufour and Taamouti (2003)) and obtain simultaneous confidence sets for scalar linear transformations of the s considered ratios.

The projection technique is based on the following elementary probability result: given a continuous fonction $g : \Theta \to \mathbb{R}^q$, $q \geq 1$, and any subset $E \subset \Theta$, we have

$$\forall x \in \Theta, \quad (x \in E) \Rightarrow (g(x) \in g(E)),$$

where

$$g(E) = \{ y \in \mathbb{R}^q : \exists\, x \in E, \quad g(x) = y \}.$$

This implies:

$$\forall x \in \Theta, \quad \Pr[x \in E] \leq \Pr[g(x) \in g(E)].$$

As a result,

$$(\Pr[\rho \in \mathrm{CS}(\rho; 1-\alpha)] \geq 1-\alpha\ ) \Longrightarrow (\Pr[g(\rho) \in g(\mathrm{CS}(\rho; 1-\alpha))] \geq 1-\alpha\ ).$$

This shows that $g(\mathrm{CS}(\rho; 1-\alpha))$ is a confidence set for $g(\rho)$ with level at least $(1-\alpha)$; so, it is a conservative confidence set for $g(\rho)$. More importantly, the projection-based confidence sets obtained for any number of transformations $g(\rho)$ of ρ are simultaneous, i.e. they satisfy the inequality in (2.5). In particular, if we consider scalar linear transformations of ρ, $l_w(\rho) = w'\rho$, $w \in \mathbb{R}^s$, then the sets

$$\mathrm{CS}\left(w'\rho; 1-\alpha\right) = \left\{ w'\rho_0 : \rho_0' M_{11} \rho_0 + 2M_{12}' \rho_0 + M_{22} \leq 0 \ \right\} \tag{4.41}$$

are simultaneous confidence sets for $w'\rho$, $w \in \mathbb{R}^s$. Special cases of linear combination include projections on the i-th component axe ρ_i of $\rho, i = 1, ...s$ and these correspond to $w = w_i = (\delta_{1i}, \delta_{2i}, ..., \delta_{si})'$, $i = 1, ...s$, where the *Kronecker delta* δ_{ji} is defined by

$$\delta_{ji} = \begin{cases} 1 & \text{when} \quad j = i \\ 0 & \text{when} \quad j \neq i \end{cases}.$$

We can now characterize the explicit form of the projection-based confidence sets for scalar linear transforms of ρ, as defined in (4.41). The following Lemma is needed for this result.

Lemma 2 *Let M_{11}, M_{12}, and M_{22} be defined by (4.37), (4.39). Then, M_{11} is nonsingular. In addition, let $d = M_{12}'M_{11}^{-1}M_{12} - M_{22}$. Then $d > 0$ if and only if M_{11} is a positive definite or a negative definite matrix.*

This result follows from the characterization provided by Zerbe et al. (1982, Appendix C) for the geometric form of the multivariate confidence region (4.38). For convenience, we give in the Appendix B the main steps of this characterization that are useful to establish Lemma 2.

Let

$$f = -M_{11}^{-1}M_{12}, \qquad d = M_{12}'M_{11}^{-1}M_{12} - M_{22}, \tag{4.42}$$

and let

$$f_i = -\left(M_{11}^{-1}\right)_{i.} M_{12}$$

be the i-th element of f,

$$\left(M_{11}^{-1}\right)_{i.}$$

be the i-th row of M_{11}^{-1}, and

$$\left(M_{11}^{-1}\right)_{ii}$$

be the i-th element of the main diagonal of M_{11}^{-1}. We can now state our main result in the following theorem.

Theorem 3 ***Projection-based confidence sets for scalar linear transformations of ρ.***

Let M_{11}, M_{12}, M_{22}, f and d be defined by (4.37), (4.39) and (4.42). Let the joint $(1-\alpha)$ level confidence set for ρ, $\mathrm{CS}\,(\rho; 1-\alpha)$, be defined as in (4.38)-(4.40). Let $w \in \mathbb{R}^s \backslash \{\mathbf{0}\}$ and $W_{11} = w'M_{11}^{-1}w$.

1. *If all the eigenvalues of M_{11} are positive, then the projection-based confidence set for $w'\rho$ defined by (4.41) corresponds to the bounded set*

$$\mathrm{CS}\left(w'\rho; 1-\alpha\right) = \left[\ w'f - (dW_{11})^{1/2}\,,\quad w'f + (dW_{11})^{1/2}\ \right].$$

2. *If M_{11} has at least two negative eigenvalues, then $\mathrm{CS}\,(w'\rho; 1-\alpha) = \mathbb{R}$.*

3. *If M_{11} has exactly one negative eigenvalue, then:*

   (a) *If $w'M_{11}^{-1}w < 0$, then the projection-based confidence set for $w'\rho$ is a union of two unbounded sets:*

$$\mathrm{CS}\left(w'\rho; 1-\alpha\right) = \left]\ -\infty,\quad w'f - (dW_{11})^{1/2}\ \right] \cup \left[\ w'f + (dW_{11})^{1/2}\,,\quad +\infty\ \right[;$$

   (b) *If $w'M_{11}^{-1}w > 0$, then $\mathrm{CS}\,(w'\rho; 1-\alpha) = \mathbb{R}$.*

   (c) *If $w'M_{11}^{-1}w = 0$, then $\mathrm{CS}\,(w'\rho; 1-\alpha) = \mathbb{R}\backslash\{w'f\}$.*

Proof. Since M_{11} is a real symmetric matrix, we have

$$M_{11} = G' D_{11} G$$

where G is an orthogonal matrix and D_{11} is a diagonal matrix whose elements are the eigenvalues of M_{11}. Let $\lambda_1, \lambda_2, ..., \lambda_s$ denote the s eigenvalues of M_{11}. Using the transformation

$$z = G(\rho - f),$$

the inequality $\rho_0' M_{11} \rho_0 + 2M_{12}' \rho_0 + M_{22} \leq 0$ is equivalent to

$$\lambda_1 z_1^2 + \lambda_2 z_2^2 + ... + \lambda_s z_s^2 \leq d.$$

We may then write $\mathrm{CS}(w'\rho; 1-\alpha)$ as

$$\mathrm{CS}\left(w'\rho; 1-\alpha\right) = \left\{w'\rho_0 : \lambda_1 z_1^2 + \lambda_2 z_2^2 + ... + \lambda_s z_s^2 \leq d, \qquad z = G(\rho_0 - f)\right\}.$$

Since $G'G = I_s$, we have

$$\begin{aligned} w'\rho &= w'G'G\rho \\ &= w'G'G(\rho - f) + w'G'Gf \\ &= w'G'\left[G(\rho - f)\right] + w'f \\ &= v'z + w'f \end{aligned}$$

where $v = Gw$. Define

$$\mathrm{CS}\left(v'z; 1-\alpha\right) = \left\{v'z_0 : \lambda_1 z_1^2 + \lambda_2 z_2^2 + ... + \lambda_s z_s^2 \leq d\right\}. \tag{4.43}$$

Then, for $x \in \mathbb{R}$,

$$\left[x \in \mathrm{CS}\left(w'\rho; 1-\alpha\right)\right] \Leftrightarrow \left[x - w'f \in \mathrm{CS}\left(v'z; 1-\alpha\right)\right].$$

The problem is then reduced to characterize $\mathrm{CS}(v'z; 1-\alpha)$. Clearly, the explicit form of $\mathrm{CS}(v'z; 1-\alpha)$ depends on the number of negative eigenvalues of M_{11}. Then, the characterization results given in the Theorem obtain from Lemma 2 and Theorems 5.1-5.2 in Dufour and Taamouti (2003). ■

Theorem 3 characterizes the possible explicit forms for simultaneous projection-based confidence limits for scalar linear transformations of ρ from the joint confidence set defined in (4.38)-(4.40). The explicit form depends on the number of negative eigenvalues of M_{11}. From the proof of Lemma 2 given in Appendix B, we see that the eigenvalues of M_{11} have the same sign as c and a, where

$$c = \left(H\hat{\theta}\right)' \left(H\hat{\Sigma}_{\hat{\theta}} H'\right)^{-1} \left(H\hat{\theta}\right) - c_\alpha$$

and

$$a = c - \left[P'S_1\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1}H\hat{\theta}\right]'\left[P'S_1\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1}H\hat{\theta}\right]$$

where c has multiplicity $s-1$ and a has multiplicity one.

Further, using Zerbe et al. (1982, Appendices C, D), it can be shown that

$$a = \left(K'\hat{\theta}\right)'\left(K'\hat{\Sigma}_{\hat{\theta}}K\right)^{-1}\left(K'\hat{\theta}\right) - c_{\alpha}.$$

If

$$\left(K'\hat{\theta}\right)'\left(K'\hat{\Sigma}_{\hat{\theta}}K\right)^{-1}\left(K'\hat{\theta}\right) < c_{1,\alpha} \tag{4.44}$$

where $c_{1,\alpha}$ denote the $(1-\alpha)$ percentile point of the $\chi^2(1)$ distribution, then the Wald test of $H_0 : K'\theta = 0$ is not significant at level α; the common denominator of the considered ratios may be arbitrarily close to zero. In this case, all the ratios $\rho_i = h_i(\theta) = L_i'\theta/K'\theta, \forall i = 1, ..., s$ are near their discontinuity region defined in (4.29).

Further,

$$\left[\left(K'\hat{\theta}\right)'\left(K'\hat{\Sigma}_{\hat{\theta}}K\right)^{-1}\left(K'\hat{\theta}\right) < c_{1,\alpha}\right] \Rightarrow \left[\left(K'\hat{\theta}\right)'\left(K'\hat{\Sigma}_{\hat{\theta}}K\right)^{-1}\left(K'\hat{\theta}\right) < c_{\alpha}\right].$$

Therefore, if

$$\left(K'\hat{\theta}\right)'\left(K'\hat{\Sigma}_{\hat{\theta}}K\right)^{-1}\left(K'\hat{\theta}\right) < c_{1,\alpha}$$

and

$$\left(H\hat{\theta}\right)'\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1}\left(H\hat{\theta}\right) > c_{\alpha}$$

then $a < 0$ and $c > 0$, and as a consequence M_{11} has exactly one negative eigenvalue. The individual ratios and any linear combination of them is almost unidentified, and this may cause a level correct confidence set to be either unbounded or the entire real line.

On the other hand, if $a > 0$ and $c > 0$, which implies that all the eigenvalues of M_{11} are positive, then

$$\left(K'\hat{\theta}\right)'\left(K'\hat{\Sigma}_{\hat{\theta}}K\right)^{-1}\left(K'\hat{\theta}\right) > c_{1,\alpha}.$$

As a result, all the ratios and any linear combination of them are well-identified; this case corresponds to bounded projection-based confidence set in Theorem 3.

The main point in Theorem 3 is that it gives easy-to-compute expressions for the confidence limits for any linear transformation $w'\rho$ of the considered ratios, and not only for the individual ratios. It can be checked that for the individual ratios $\rho_i, i = 1, 2, ..., s$, the simultaneous confidence limits given in Theorem 3 are numerically identical to the ones obtained by Zerbe et al. (1982).[7] Specifically, for any individual ratio $\rho_i, i = 1, 2, ..., s$, the confidence limits are solutions to the following quadratic inequality:

$$A_i\delta_{i0}^2 + 2B_i\delta_{i0} + C_i \leq 0,$$

[7] This follows from the uniqueness of the projection on any individual ρ_i-axe.

where

$$\begin{cases} A_i &= (K'\theta)^2 - c_\alpha \left(K'\hat{\Sigma}_{\hat{\theta}} K\right) \\ B_i &= c_\alpha \left(L_i'\hat{\Sigma}_{\hat{\theta}} K\right) - (L_i'\theta)(K'\theta) \\ C_i &= (L_i'\theta)^2 - c_\alpha \left(L_i'\hat{\Sigma}_{\hat{\theta}} L_i\right). \end{cases}$$

The following corollary shows that the projection-based confidence set for any linear transformation $w'\rho$ cannot be empty; as a special case, the simultaneous confidence sets for the individual ratios are non-empty sets.

Corollary 4 *The simultaneous projection-based confidence sets defined by (4.41) for any number of scalar linear transformations of ratios with common denominator are non-empty sets. In particular, the simultaneous projection-based confidence sets for the individual ratios $\rho_i, i = 1, ...s$ are non-empty.*

Proof. From Dufour and Taamouti (2003), the only case where the projection-based confidence set for $w'\rho$ is an empty set corresponds to: (i) M_{11} is positive definite and (ii) $d < 0$. Using Lemma 2, this is impossible, and the result follows. ■

5 Empirical applications

In this section we illustrate the simultaneous confidence sets procedure discussed in the previous section through three empirical applications. The models we analyze are related to important issues in transportation and energy economics. The first one considers the trinomial logit model of travel demand discussed by Ben-Akiva and Lerman (1985)[8] in order to construct simultaneous confidence sets for the values of in-vehicle and out-of vehicle travel times. The second one concerns inference for three values of travel time in multinomial probit models that combine maximum simulated likelihood (SML) estimator with Geweke-Hajivassiliou-Keane (GHK) choice probability simulator, in which cases standard asymptotics performs poorly. The third illustration is related to simultaneous inference for price- and income-elasticities in sector total energy demand models for the Province of Québec.

5.1 Value of time in trinomial logit model of travel demand

Estimating value of time is one important application of travel demand models with linear random utility. Although various theories of time allocation reveal that the value of travel time can be perceived in different ways, most of empirical studies refer to the value of travel time as the amont of money the traveler agrees to pay in order to save one unit of the total duration of his travel [Ashton (1947), De Vany (1974), Truong and Hensher (1985), Bates (1987), Ben-Akiva, Bolduc and Bradley (1993)]. In a discrete choice framework, when the traveler's utility fonction is specified as a linear function of travel cost, travel time and other variables, his evaluation of

[8] We consider this model in section (3.2.1) for an illustration of confidence set procedures for one ratio of parameters.

Table 3: Simultaneous confidence sets for values of total travel time and of out-of-vehicle time from Ben-Akiva and Lerman(1985)'s trinomial logit model, 95% nominal level.

Type of travel time	Delta method		Fieller method		
Total travel time (δ_{tot})	[-2.655,	30.253]	$]-\infty,$	$-40.843] \cup [3.918,$	$+\infty[$
Out-of-vehicle time (δ_{out})	[-4.055,	46.639]	$]-\infty,$	$-66.877] \cup [6.760,$	$+\infty[$

Notes: _ The delta method-based confidence intervals are not simultaneous.

the value of travel time is, up to a scalar constant, equal to the ratio of the coefficient of the time variable over the coefficient of the cost variable [Truong and Hensher (1985), Bates (1987)].

We consider the trinomial logit model of travel demand from Ben-Akiva and Lerman (1985, Chapters 3, 5 and 7.). We have described this model in section 3.2.1. Here, we construct simultaneous confidence sets for the value of total travel time (δ_{tot}) and the value of out-of-vehicle time (δ_{out}) defined in equations (3.14) and (3.15) respectively. Using the notation of section 3.2.1, each of these values of travel time is a linear combination of the ratios functions $h_1(\theta) = \theta_3/\theta_5$ and $h_2(\theta) = \theta_4/\theta_5$ defined in model (3.12)-(3.13); so we can obtain Fieller-type projection-based simultaneous confidence sets. We have also computed the delta method based confidence sets for (δ_{tot}) and (δ_{out}), which are not simultaneous. We use sample average values for annual household income (equal to 12900\$/year) and for one-way distance (equal to 810 centimiles). Our results are reported in Table 3.

5.2 Value of time in multinomial probit models in transportation

The second application we consider is, as the previous one, related to the economic value of time in discrete choice models. We consider results from the multinomial probit (MNP) model with correlated utilities estimated on a data bank on the choice of transportation modes for the morning peak journey to work in the central business district of Santiago; for details on the model specification and the data see Bolduc (1999). Three specific uses of travel time are considered: in vehicle time, walking time and waiting time. The utility that a worker derives from his journey to work is assumed to be a linear function of transportation modes'specific dummies and of other variables including cost/income, walking time, in vehicle time, waiting time, a sex dummy and a dummy for no cars/no permit holders. This leads to three specific values of time that are expressed as ratios of parameters with common denominator which is given by the coefficient of the variable cost/income. The estimation method combines the simulated maximum likelihood (SML) and the Geweke-Hajivassiliou-Keane (GHK) choice probability simulator based on analytically computed scores. From the estimation results, Bolduc (1999) provided point estimates for value of time ratios and standard delta method-based asymptotic t-statistics. Apply our characterization results, we obtain simultaneous projection-based confidence sets for the three value of time ratios. Tables **??** and **??** report the results along with the delta method-based confidence intervals.

Table 4: Simultaneous confidence sets for values of time as percentage of net personal income, 95% nominal level.

Type of travel time	Confidence set	MNP i.i.d.	SML MNP R = 50 homoscedastic	SML MNP R = 50 unconstrained
In vehicle time	delta	[117.90, 285.52]	[122.07, 300.46]	[102.69, 265.37]
	Fieller	[95.77, 352.52]	[101.77, 382.24]	[61.88, 411.03]
Walking time	delta	[240.79, 450.66]	[239.39, 468.58]	[178.65, 397.82]
	Fieller	[222.08, 548.30]	[223.13, 590.17]	[141.48, 631.17]
Waiting time	delta	[453, 1093.37]	[507.58, 1201.34]	[286.99, 830.65]
	Fieller	[370.12, 1350.40]	[437.36, 1533.36]	[178.70, 1373.47]

Note: _ The delta method-based confidence intervals are not simultaneous.

Table 5: Simultaneous confidence sets for values of time as percentage of net personal income, 95% nominal level (continued).

Type of travel time	Confidence set	SML MNP R = 250 unconstrained	SML MNP R = 250 constrained
In vehicle time	delta	[110.32, 286.98]	[121.00, 307.96]
	Fieller	[65.89, 457.18]	[76.41, 489.12]
Walking time	delta	[182.09, 413.40]	[192.22, 439.45]
	Fieller	[143.11, 678.24]	[151.55, 719.44]
Waiting time	delta	[300.35, 879.82]	[310.43, 902.24]
	Fieller	[189.56, 1512.05]	[205.82, 1555.02]

Note: _ The delta method-based confidence intervals are not simultaneous.

5.3 Price- and Income-Elasticities in total energy demand models

We consider a partial adjustment model of total energy demand for four sectors of energy use (industrial, commercial, residential and manufactured) in the Province of Québec. For each sector, total energy demand depends on sector-specific explanatory variables and the model is estimated with annual data set from 1962 to 2002. The demand equations are specified as follows.

- For the residential sector:

$$\begin{aligned} \ln(\mathrm{HTE}_t) &= a_0 + a_1 \ln(\mathrm{HTE}_{t-1}) + a_2 \ln(\mathrm{PriceE}_t) + a_3 \ln(\mathrm{HINC}_t) + \\ &\quad a_4 \ln(\mathrm{DDH}_t) - a_1 a_4 \ln(\mathrm{DDH}_{t-1}) + u_{1t} \end{aligned} \tag{5.45}$$

  where for year t, HTE_t is average annual total energy demand per household, HINC_t is average annual disposable income per household, PriceE_t is aggregate real price of energy, DDH_t is heating degree days.

- For the commercial sector:

$$\begin{aligned} \ln(\mathrm{TEC}_t) &= b_0 + b_1 \ln(\mathrm{TEC}_{t-1}) + b_2 \ln(\mathrm{PriceE}_t) + b_3 \ln(\mathrm{GDPC}_t) + \\ &\quad b_4 \ln(\mathrm{DDH}_t) - b_1 b_4 \ln(\mathrm{DDH}_{t-1}) + u_{2t} \end{aligned} \tag{5.46}$$

  where for year t, TEC_t is total energy demand by the commercial sector, GDPC_t is real gross domestic product in the commercial sector.

- For the manufactured sector:

$$\ln(\mathrm{TEM}_t) = c_0 + c_1 \ln(\mathrm{TEM}_{t-1}) + c_2 \ln(\mathrm{PriceE}_t) + c_3 \ln(\mathrm{GDPM}_t) + u_{3t} \tag{5.47}$$

  where for year t, TEM_t is total energy demand by the manufactured sector, GDPM_t is real gross domestic product in the manufactured sector.

- For the industrial sector:

$$\ln(\mathrm{TEI}_t) = d_0 + d_1 \ln(\mathrm{TEI}_{t-1}) + d_2 \ln(\mathrm{PriceE}_t) + d_3 \ln(\mathrm{GDPI}_t) + u_{4t} \tag{5.48}$$

  where for year t, TEI_t is total energy demand by the industrial sector, GDPI_t is real gross domestic product in the industrial sector.

Convergence of the adjustment process for each sector requires $0 < a_1 < 1$, $0 < b_1 < 1$, $0 < c_1 < 1$, $0 < d_1 < 1$. The error terms $u_{1t}, u_{2t}, u_{3t}, u_{4t}$ are assumed i.i.d. normal. The parameters of each equation are estimated using maximum likelihood estimator. The dynamic specification of energy demand in (5.45)-(5.48) allows one to compute the long-run price and income elasticities of sector energy demand. As an example, in the residential sector the long-run price elasticity $\mathcal{E}_{rp}$ is given by

$$\mathcal{E}_{rp} = \frac{a_2}{1 - a_1} \tag{5.49}$$

Table 6: Simultaneous confidence sets for long run price- and income- elasticities of total energy demand, nominal level: 95%.

Sector	Elasticities	Delta method		Fieller method			
Residential	Price	[-0.3682,	0.0911]	$]-\infty,$	0.1112] $\cup$ [0.9542,	$+\infty[$
	Income	[-4.4576,	5.0129]	$]-\infty,$	1.3333] $\cup$ [7.0938,	$+\infty[$
Commercial	Price	[-0.7034,	-0.3935]	[-0.7419,	-0.3162]	
	Income	[0.7857,	1.1298]	[0.6950,	1.1371]	
Industrial	Price	[-0.6531,	0.1257]	[-1.6534,	0.1304]	
	Income	[0.6445,	1.4643]	[0.4449,	2.1119]	
Manufacture	Price	[-0.2277,	-0.0420]	[-0.2766,	-0.0230]	
	Income	[0.6042,	0.9088]	[0.6211,	0.8917]	

Note: _ The delta method-based confidence intervals are not simultaneous.

and the long-run income elasticity $\mathcal{E}_{rinc}$ is

$$\mathcal{E}_{rinc} = \frac{a_3}{1 - a_1}. \tag{5.50}$$

So, $\mathcal{E}_{rp}$ and $\mathcal{E}_{rinc}$ are ratios of parameters with a common denominator and we can obtain simultaneous confidence sets for $\mathcal{E}_{rp}$ and $\mathcal{E}_{rinc}$. Table 6 reports simultaneous confidence sets for long run price-elasticities and income-elasticities for four sectors of energy use.

6 Conclusion

To be completed.

A Appendix: Characterization of the solutions to the Fieller-type confidence set for one parameters ratio

In this appendix, we characterize the Fieller-type confidence set for one parameter ratio. In the context of exact Fieller confidence sets, Scheffé (1970) gives such a characterization for the ratio of two means of normals and Zerbe et al. (1982) provides an extension to parameter ratio in normal linear regressions. Here, we extend these results to parameter ratio when the normal distribution of the estimator is only asymptotically justified.

Proposition 5 *Let A, B, C be defined as in (3.11) and let $\Delta = B^2 - AC$. Then, the $(1-\alpha)$-level Fieller-type confidence set* $\mathrm{FCS}(\delta; 1-\alpha)$ *for the ratio $\delta = \theta_1/\theta_2$, defined in (3.9), is characterized as follows:*

1. *If $\Delta > 0$, then*
   - *(a) if $A > 0$, then* $\mathrm{FCS}(\delta; 1-\alpha) = \left[\frac{-B-\sqrt{\Delta}}{A}, \quad \frac{-B+\sqrt{\Delta}}{A}\right]$,
   - *(b) else, if $A < 0$, then* $\mathrm{FCS}(\delta; 1-\alpha) = \left]-\infty, \quad \frac{-B-\sqrt{\Delta}}{A}\right] \cup \left[\frac{-B-\sqrt{\Delta}}{A}, \quad +\infty\right[$.
2. *If $\Delta < 0$, then $A < 0$ and* $\mathrm{FCS}(\delta; 1-\alpha) = \mathbb{R}$.

Proof. We solve the equation

$$A\delta_0^2 + 2B\delta_0 + C = 0$$

where

$$\begin{cases} A &= \hat{\theta}_2^2 - z_{\alpha/2}^2 \hat{v}_2 \\ B &= -\hat{\theta}_1\hat{\theta}_2 + z_{\alpha/2}^2 \hat{v}_{12} \\ C &= \hat{\theta}_1^2 - z_{\alpha/2}^2 \hat{v}_1 \end{cases}$$

for real solutions δ_0. Except for a set of measure zero, $A \neq 0$, so we have a quadratic equation. Similarly, except for a set of measure zero, $\Delta \neq 0$; so we discuss the two cases where $\Delta > 0$ and $\Delta < 0$. Real solutions exist if and only if

$$\Delta > 0.$$

If $\Delta > 0$, then we have two distinct real solutions δ_{01} and δ_{02} given by

$$\delta_{01} = \frac{-B - \sqrt{\Delta}}{A}$$

$$\delta_{02} = \frac{-B + \sqrt{\Delta}}{A}.$$

Therefore,

$$\mathrm{FCS}\left(\delta;1-\alpha\right)=\begin{cases}\left[\frac{-B-\sqrt{\Delta}}{A},\quad\frac{-B+\sqrt{\Delta}}{A}\right] & if\quad A>0\\ \left]-\infty,\quad\frac{-B-\sqrt{\Delta}}{A}\right]\cup\left]\frac{-B-\sqrt{\Delta}}{A},\quad+\infty\right] & if\quad A<0\end{cases}.$$

To complete the proof, let us show that if $\Delta<0$, then $A<0$. First, let us write Δ as

$$\Delta=\left(\hat{v}_{12}^2-\hat{v}_1\hat{v}_2\right)z_{\alpha/2}^4+\left(\hat{\theta}_1^2\hat{v}_2+\hat{\theta}_2^2\hat{v}_1-2\hat{\theta}_1\hat{\theta}_2\hat{v}_{12}\right)z_{\alpha/2}^2.$$

Since $\hat{v}_{12}^2-\hat{v}_1\hat{v}_2<0$ (Cauchy-Schwartz inequality), Δ is negative if and only if

$$z^*=\frac{\hat{\theta}_1^2\hat{v}_2+\hat{\theta}_2^2\hat{v}_1-2\hat{\theta}_1\hat{\theta}_2\hat{v}_{12}}{\hat{v}_1\hat{v}_2-\hat{v}_{12}^2}<z_{\alpha/2}^2. \tag{A.51}$$

From

$$\hat{\theta}_2^2/v_2-z^*=-\frac{\left(\hat{\theta}_2\hat{v}_{12}-\hat{\theta}_1\hat{v}_2\right)^2}{\hat{v}_2\left(\hat{v}_1\hat{v}_2-\hat{v}_{12}^2\right)},$$

we get

$$\hat{\theta}_2^2/\hat{v}_2-z^*<0. \tag{A.52}$$

Then, from (A.51) and (A.52), we have

$$\Delta<0\quad\Rightarrow\hat{\theta}_2^2/\hat{v}_2<z^*<z_{\alpha/2}^2,$$

which establishes that

$$\Delta<0\Rightarrow A<0.$$

Clearly, this implies

$$\Delta<0\Rightarrow\left(\forall\delta_0\in\mathbb{R},\ A\delta_0^2+2B\delta_0+C<0\right).$$

■

B Appendix: Proof of lemma 2

To prove Lemma 2, we need the following result known as the Sylvester's law of inertia.

Lemma 6 *(Sylvester's law of inertia) Let Π_1 and Π_2 be any $p \times p$ symmetric matrices of the same rank $r \leq p$. If $\Pi_1 = N\Pi_2 N'$ for some matrix N, then Π_1 and Π_2 have the same number of positive eigenvalues.*

Proof. (Lemma 6) Let us recall that the eigenvalues of any symmetric matrix are real numbers. In addition, for any symmetric matrix Π, there exists an orthogonal matrix R (i.e. $R' = R^{-1}$) such that $R\Pi R' = D$, where D is a diagonal matrix with the eigenvalues of Π on its main diagonal.

Let $\lambda_1^{(1)}, \lambda_2^{(1)}, ..., \lambda_p^{(1)}$ denote the p not necessarily distinct eigenvalues of Π_1; similarly let $\lambda_1^{(2)}, \lambda_2^{(2)}, ..., \lambda_p^{(2)}$ denote the eigenvalues of Π_2, and let

$$\begin{aligned} D^{(1)} &= \operatorname{diag}\left(\lambda_1^{(1)}, \lambda_2^{(1)}, ..., \lambda_p^{(1)}\right) \\ D^{(2)} &= \operatorname{diag}\left(\lambda_1^{(2)}, \lambda_2^{(2)}, ..., \lambda_p^{(2)}\right), \end{aligned}$$

where $\operatorname{diag}(a_1, a_2, ..., a_p)$ denote the diagonal matrix with $a_1, a_2, ..., a_p$ as its main diagonal elements. Then, there exist two orthogonal matrices $R^{(1)}$ and $R^{(2)}$ such that

$$R^{(1)}\Pi_1 R^{(1)\prime} = D^{(1)} \quad \text{(B.53)}$$

$$R^{(2)}\Pi_2 R^{(2)\prime} = D^{(2)}. \quad \text{(B.54)}$$

Let l_1 and l_2 be the number of positive eigenvalues of Π_1 and Π_2 respectively. Order the eigenvalues of Π_1 so that the first l_1 scalars on the main diagonal of $D^{(1)}$ are positive and the next $r - l_1$ are negative; and do so for Π_2. This implies the following:

$$\begin{aligned} \lambda_1^{(1)} &> 0, \lambda_2^{(1)} > 0, ..., \lambda_{l_1}^{(1)} > 0, \\ \lambda_{l_1+1}^{(1)} &< 0, \lambda_{l_1+2}^{(1)} < 0, ..., \lambda_r^{(1)} < 0, \\ \lambda_{r+1}^{(1)} &= 0, \lambda_{r+2}^{(1)} = 0, ..., \lambda_p^{(1)} = 0; \end{aligned}$$

and

$$\begin{aligned} \lambda_1^{(2)} &> 0, \lambda_2^{(2)} > 0, ..., \lambda_{l_2}^{(2)} > 0, \\ \lambda_{l_2+1}^{(2)} &< 0, \lambda_{l_2+2}^{(2)} < 0, ..., \lambda_r^{(2)} < 0, \\ \lambda_{r+1}^{(2)} &= 0, \lambda_{r+2}^{(2)} = 0, ..., \lambda_p^{(2)} = 0. \end{aligned}$$

Let

$$\begin{aligned} D_1^{(1)} &= \operatorname{diag}\left(\sqrt{\lambda_1^{(1)}}, \sqrt{\lambda_2^{(1)}}, ..., \sqrt{\lambda_{l_1}^{(1)}}, -\sqrt{-\lambda_{l_1+1}^{(1)}}, ..., -\sqrt{-\lambda_r^{(1)}}, 0, ..., 0\right) \\ D_1^{(2)} &= \operatorname{diag}\left(\sqrt{\lambda_1^{(2)}}, \sqrt{\lambda_2^{(2)}}, ..., \sqrt{\lambda_{l_2}^{(2)}}, -\sqrt{-\lambda_{l_2+1}^{(2)}}, ..., -\sqrt{-\lambda_r^{(2)}}, 0, ..., 0\right). \end{aligned}$$

Define $D_0^{(1)}$ and $D_0^{(2)}$ by:

$$D_0^{(1)} = \left[\begin{array}{ccc} I_{l_1} & 0 & 0 \\ 0 & -I_{r-l_1} & 0 \\ 0 & 0 & 0_{p-r} \end{array}\right]$$

and

$$D_0^{(2)} = \left[\begin{array}{ccc} I_{l_2} & 0 & 0 \\ 0 & -I_{r-l_2} & 0 \\ 0 & 0 & 0_{p-r} \end{array}\right]$$

Then we have:

$$\begin{aligned} D^{(1)} &= U^{(1)} D_1^{(1)} D_0^{(1)} D_1^{(1)} U^{(1)\prime} && \text{(B.55)} \\ D^{(2)} &= U^{(2)} D_1^{(2)} D_0^{(2)} D_1^{(2)} U^{(2)\prime} && \text{(B.56)} \end{aligned}$$

where $U^{(1)}$ and $U^{(2)}$ are permutation matrices and so they are orthogonal. Hence, substituting (B.55) and (B.56) into (B.53) and (B.54) respectively, we obtain:

$$\begin{aligned} \Pi_1 &= R^{(1)\prime} U^{(1)} D_1^{(1)} D_0^{(1)} D_1^{(1)} U^{(1)\prime} R^{(1)} \\ \Pi_2 &= R^{(2)\prime} U^{(2)} D_1^{(2)} D_0^{(2)} D_1^{(2)} U^{(2)\prime} R^{(2)}. \end{aligned}$$

Then, we can write $D_0^{(1)}$ and $D_0^{(2)}$ in the form:

$$\begin{aligned} D_0^{(1)} &= P^{(1)} \Pi_1 P^{(1)\prime} && \text{(B.57)} \\ D_0^{(2)} &= P^{(2)} \Pi_2 P^{(2)\prime} && \text{(B.58)} \end{aligned}$$

where $P^{(1)} = \left[R^{(1)\prime} U^{(1)} D_1^{(1)}\right]^{-1}$ and $P^{(2)} = \left[R^{(2)\prime} U^{(2)} D_1^{(2)}\right]^{-1}$.

To prove the lemma, it suffices to show that $l_1 = l_2$. Since $\Pi_1 = N \Pi_2 N'$, using (B.57) and (B.58) we get

$$D_0^{(1)} = P^{(1)} \Pi_1 P^{(1)\prime} = P^{(1)} N \left(\left(P^{(2)}\right)^{-1} D_0^{(2)} \left(P^{(2)\prime}\right)^{-1} \right) N' P^{(1)\prime}$$

Let $\digamma = P^{(1)} N \left(P^{(2)}\right)^{-1}$. Then

$$D_0^{(1)} = \digamma D_0^{(2)} \digamma'. \tag{B.59}$$

Suppose $l_2 < l_1$. Let $Y = \left(Y_1', \mathbf{0}_{1\times(p-l_1)}\right)'$ where $Y_1 = (y_1, y_2, ..., y_{l_1})'$ and $Y_1 \neq 0$. We have

$$Y' D_0^{(1)} Y = \sum_{i=1}^{l_1} y_i^2 > 0. \tag{B.60}$$

Now, partition $\digamma'$ in the form:

$$\digamma' = \left[\begin{array}{cc} \digamma_1 & \digamma_2 \\ \digamma_3 & \digamma_4 \end{array}\right]$$

where $\digamma_1$ is $l_2 \times l_1$. Since $l_2 < l_1$, the null space of $\digamma_1$ is not reduced to the null vector space and we can choose $Y_1 \neq 0$ so that $\digamma_1 Y_1 = \mathbf{0}_{l_2 \times 1}$. Define $z = \digamma_3 Y_1$, and write $z = (z_1, z_2, ..., z_{p-l_2})'$. Then we have $\digamma' Y = (\mathbf{0}_{1\times l_2}, z')'$. Using (B.59), we obtain

$$\begin{aligned} Y' D_0^{(1)} Y &= \left(\digamma' Y\right)' D_0^{(2)} \left(\digamma' Y\right) \\ &= -\sum_{j=1}^{r-l_2} z_j^2 \leq 0. \end{aligned} \tag{B.61}$$

Clearly, (B.60) and (B.61) are in contradiction; as a result, $l_2 < l_1$ is impossible.

Similarly, interchanging the roles of $D_0^{(1)}$ and $D_0^{(2)}$, we can see that $l_1 < l_2$ is also impossible. Hence, $l_1 = l_2$, and Lemma 6 is proved. ■

Proof. (Lemma 2) Since we assume that $\det\left(\hat{\Sigma}_{\hat{\theta}}\right) \neq 0$, the covariance matrix $\hat{\Sigma}_{\hat{\theta}}$ is symmetric and positive definite. Then, since H has full row rank, it follows that $\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1}$ is symmetric and positive definite. Similarly, since S_1 has full row rank, $Q = S_1\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} S_1'$ is a symmetric and positive definite matrix. Then, there exists a nonsingular matrix P such that $P'QP = I_s$. Using Lemma 6, the two matrices $P'M_{11}P$ and M_{11} have the same number of positive eigenvalues and the same number of negative eignenvalues. In addition, we have:

$$\begin{aligned} P'M_{11}P &= P'\left(S_1 M S_1'\right)P \\ &= P'\left[S_1\left(c\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} - \left[\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} H\hat{\theta}\right]\left[\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} H\hat{\theta}\right]'\right)S_1'\right]P \\ &= cP'S_1\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} S_1'P - P'S_1\left[\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} H\hat{\theta}\right]\left[\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} H\hat{\theta}\right]' S_1'P \\ &= cI_s - \left[P'S_1\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} H\hat{\theta}\right]\left[P'S_1\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} H\hat{\theta}\right]'. \end{aligned}$$

The last expression shows that $P'M_{11}P$ is a patterned matrix of the type discussed in Graybill (1983, p. 206). Thus, $P'M_{11}P$ has $(s-1)$ eigenvalues equal to c and one eigenvalue equal to

$$a = c - \left[P'S_1\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} H\hat{\theta}\right]'\left[P'S_1\left(H\hat{\Sigma}_{\hat{\theta}}H'\right)^{-1} H\hat{\theta}\right].$$

Except for a set of values for $\hat{\theta}$ of measure zero, $c \neq 0$ and $a \neq 0$; so, zero is not an eigenvalue of M_{11} and M_{11} is nonsingular. The sign of $\det(M_{11})$ is the same as the sign of $\det(P'M_{11}P) = ac^{s-1}$.

Similarly, using the same arguments as above for the matrix M, the sign of $\det(M)$ is the same as that of $-c_\alpha c^s$. In addition, using block matrix inversion formula, and since $M_{22} - M_{21}M_{11}^{-1}M_{12}$ is a scalar, we get:

$$\begin{aligned} \det(M) &= \det(M_{11})\det\left(M_{22} - M_{21}M_{11}^{-1}M_{12}\right) \\ &= -\det(M_{11})\left(M_{21}M_{11}^{-1}M_{12} - M_{22}\right) \\ &= -\det(M_{11})\, d, \end{aligned}$$

Then, $d = -\det(M) / \det(M_{11})$, which implies that d has the same sign as $c_\alpha c/a$. So, we have the following results:

- If $a > 0$, then $c > 0$: all the eigenvalues of M_{11} are positive, and M_{11} is positive definite. If $c < 0$, then $a < 0$: all the eigenvalues of M_{11} are negative, and M_{11} is negative definite. Clearly, we have $d > 0$ in these two cases.
- On the other hand, if ($c > 0$ and $a < 0$), we have $d < 0$; then M_{11} has at least one positive eigenvalue and at least one negative eigenvalue; thus, M_{11} is neither positive definite nor negative definite. Lemma 2 is then proved.

■

References

Abdelkhalek, T. and Dufour, J.-M. (1998), 'Statistical inference for computable general equilibrium models, with application to a model of the Moroccan economy', *Review of Economics and Statistics* **LXXX**, 520–534.

Ashton, H. (1947), 'The time element in transportation', *American Economic Review* **37(2)**, 423–440.

Banks, J., Blundell, R. and Lewbel, A. (1997), 'Quadratic engel curves and consumer demand', *The Review of Economics and Statistics* **79(4)**, 527–539.

Bates, J. J. (1987), 'Measuring travel time values with a discrete choice model: A note', *The Economic Journal* **97**(386), 493–498.

Ben-Akiva, M., Bolduc, D. and Bradley, M. (1993), 'Estimation of travel choice models with randomly distributed values of time', *Transportation Research Records* **1413**, 88–97.

Ben-Akiva, M., Bolduc, D. and Walker, J. (2001), Specification, identification, and estimation of the logit kernel (or Continuous Mixed Logit) model, Technical report, MIT working paper.

Ben-Akiva, M. and Lerman, S. R. (1985), *Discrete Choice Analysis: Theory And Application to Travel Demand*, The MIT Press, Cambridge, MA.

Bolduc, D. (1992), 'Generalized autoregressive errors in the multinomial probit model', *Transportation Research Part B* **26B**(2), 155–170.

Bolduc, D. (1999), 'A practical technique to estimate multinomial probit models in transportation', *Transportation Research Part B* **33**, 63–79.

Bucephala, J. P. and Gatsonis, C. A. (1988), 'Bayesian inference for ratios of coefficients in a linear model', *Biometrics* **44(1)**, 87–101.

Buonaccorsi, J. P. (1985), 'Ratios of linear combinations in the general linear model', *Communications in Statistics, Theory and Methods* **14**, 635–650.

Darby, S. C. (1980), 'A bayesian approach to parallel-line bioassay', *Biometrika* **3**, 607–612.

Davidson, R. and MacKinnon, J. G. (1999*a*), 'Bootstrap testing in nonlinear models', *International Economic Review* **40**, 487–508.

Davidson, R. and MacKinnon, J. G. (1999*b*), 'The size distortion of bootstrap tests', *Econometric Theory* **15**, 361–376.

Davidson, R. and MacKinnon, J. G. (2000), 'Bootstrap tests: How many bootstraps', *Econometric Reviews* **19**, 55–68.

De Vany, A. (1974), 'The revealed value of time in air travel', *Review of Economics and Statistics* **56(1)**, 77–82.

Deaton, A. S. and Muellbauer, J. (1980), 'An almost ideal demand system', *American Economic Review* **70**, 312–326.

Dufour, J.-M. (1989), 'Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions', *Econometrica* **57**, 335–355.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**, 1365–1389.

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics', *Canadian Journal of Economics* **36(4)**, 767–808.

Dufour, J.-M. and Taamouti, M. (2003), Projection-based statistical inference in linear structural models with possibly weak instruments, Technical report, C.R.D.E., Université de Montréal.

Fieller, E. C. (1940), 'The biological standardization of insulin', *Journal of the Royal Statistical Society (Supplement)* **7(1)**, 1–64.

Graybill, Franklin, A. (1983), *Matrices with applications in statistics*, Belmont, Calif. : Wadsworth International Group, Belmont.

McFadden, D. (1989), 'A method of simulated moments for estimation of discrete response models without numerical integration', *Econometrica* **57**, 995–1026.

Miller, Jr., R. G. (1981), *Simultaneous Statistical Inference*, second edn, Springer-Verlag, New York.

Savin, N. E. and Würtz, A. H. (1998), The effect of nuisance parameters on critical values and power: Lagrange multiplier tests in logit models, Technical report, Department of Economics, University of Iowa and Department of Economics, University of Aarhus.

Scheffé, H. (1959), *The Analysis of Variance*, first edn, John Wiley & Sons, New York.

Scheffé, H. (1970), 'Multiple testing versus multiple estimation in proper confidence sets, estimation of directions and ratios', *Annals of Mathematical Statistics* **41**, 1–29.

Selwyn, M. R. and Hall, N. R. (1984), 'On bayesian methods in bioequivalence', *Biometrics* **40**, 1103–1108.

Stock, J. H. and Wright, J. H. (2000), 'GMM with weak identification', *Econometrica* **68**, 1097–1126.

Stock, J. H., Wright, J. H. and Yogo, M. (2002), 'A survey of weak instruments and weak identification in generalized method of moments', *Journal of Business and Economic Statistics* **20**(4), 518–529.

Stock, J. H. and Yogo, M. (2002), Testing for weak instruments in linear IV regression, Technical Report 284, N.B.E.R., Harvard University, Cambridge, Massachusetts.

Truong, P. T. and Hensher, D. A. (1985), 'Measurement of travel time values and opportunities cost from a discrete coice model', *The Economic Journal* **95(378)**, 439–451.

Walker, J. (2001), Extended discrete choice models: Integrated framework, flexible error structures and latent variables, Technical report, Ph.D thesis, Massachusetts Institute of Technology.

Young, D. A., Zerbe, G. O. and Hay, W. W. (1997), 'Fieller's theorem, Scheffé simultaneous confidence intervals, and ratios of parameters of linear and nonlinear mixed-effects models', *Biometrics* **53(3)**, 838–847.

Zerbe, G. O. (1978), 'On fieller's theorem and the general linear model', *The American Statistician* **32(3)**, 103–105.

Zerbe, G. O., Laska, E., Meisner, M. and Kushner, H. B. (1982), 'On multivariate confidence regions and simultaneous confidence limits for ratios', *Communications in Statistics, Theory and Methods* **11(21)**, 2401–2425.