Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

European Journal of Operational Research 185 (2008) 195–203

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/ejor

Stochastics and Statistics

Comparative statics of the generalized maximum entropy estimator of the general linear model

Michael R. Caputo [a,*], Quirino Paris [b]

[a] *Department of Economics, University of Central Florida, Orlando, FL 32816-1400, USA*
[b] *Department of Agricultural and Resource Economics, University of California, One Shields Avenue, Davis, CA 95616-8512, USA*

Received 12 December 2005; accepted 16 December 2006
Available online 17 January 2007

Abstract

The generalized maximum entropy method of information recovery requires that an analyst provides prior information in the form of finite bounds on the permissible values of the regression coefficients and error values for its implementation. Using a new development in the method of comparative statics, the sensitivity of the resulting coefficient and error estimates to the prior information is investigated. A negative semidefinite matrix reminiscent of the Slutsky-matrix of neoclassical microeconomic theory is shown to characterize the said sensitivity, and an upper bound for the rank of the matrix is derived.

Keywords: Comparative statics; Sensitivity analysis; Generalized maximum entropy; General linear model; Regression

1. Introduction

Alternative methods for the recovery of economic information show up routinely in the econometrics and statistics literature. Recently, a method of information recovery based on the maximum entropy formalism (Shannon, 1948; Jaynes, 1957) has been developed by Golan et al. (1996). Golan et al. (1996) developed what they call the generalized maximum entropy (GME) approach to information recovery, and related GME to ordinary least squares and other estimators using real data and Monte Carlo experiments. The basic idea underlying the GME approach is to reparameterize the general linear model $Y_t = \sum_{k=1}^{K} X_{tk}\beta_k + \varepsilon_t, t = 1, 2, \ldots, T$, so that it can be accommodated within the classical maximum entropy framework. This is accomplished by treating each regression coefficient β_k as a discrete random variable with a compact *support interval* consisting of $2 \leqslant M < +\infty$ possible outcomes, where a support interval is defined as a closed and bounded interval of the real line in which each β_k is restricted to lie. Taking the case of $M = 2$ for pedagogical reasons, and letting z_{k1} and z_{k2} be the two possible outcomes and thus the finite endpoints of the support interval, β_k is expressed

* Corresponding author. Tel.: +1 407 823 1405; fax: +1 407 823 3269.
E-mail address: mcaputo@bus.ucf.edu (M.R. Caputo).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2006.12.031

as a convex combination of z_{k1} and z_{k2}, namely, $\beta_k = p_1^k z_{k1} + p_2^k z_{k2}$, where $p_1^k \geqslant 0, p_2^k \geqslant 0$, and $p_1^k + p_2^k = 1$. Similarly, each error ε_t is treated as a finite and discrete random variable with a compact support interval consisting of $2 \leqslant N < +\infty$ possible outcomes. Assuming $N = 2$, each ε_t can also be written as a convex combination of the two finite support values v_{t1} and v_{t2}, to wit, $\varepsilon_t = q_1^t v_{t1} + q_2^t v_{t2}$, where $q_1^t \geqslant 0$, $q_2^t \geqslant 0$, and $q_1^t + q_2^t = 1$.

As a result of this reparameterization, the entropy objective function consists of the coefficient entropy plus the error entropy, with the reparameterized linear model and adding up conditions on the probabilities serving as constraints. Given the support intervals specified by the analyst, the solution of the GME problem recovers estimates of the probabilities $(\hat{p}_1^k, \hat{p}_2^k)$ and $(\hat{q}_1^t, \hat{q}_2^t)$ that maximize entropy and thus are the most uniform or uncertain. Accordingly, the GME solution recovers estimates of the probability distribution of each regression coefficient and error. Point estimates of the regression coefficients and errors are then obtained by using the formulas described above, videlicet, $\hat{\beta}_k = \hat{p}_1^k z_{k1} + \hat{p}_2^k z_{k2}$ and $\hat{\varepsilon}_t = \hat{q}_1^t v_{t1} + \hat{q}_2^t v_{t2}$.

Arguably, the most controversial aspect of the GME reparameterization is that the researcher specifies the support intervals of the regression coefficients and errors as priors. Given this feature, it is natural to ask the ensuing question—How sensitive are the estimated regression coefficients and errors to the prior information provided by the analyst? The central result of this paper is a precise analytical and qualitative answer to this question. In particular, we show that the change in the GME estimates of the regression coefficients and errors of the general linear statistical model with respect to a change in their support intervals—known as the comparative statics, or equivalently, sensitivity results, of the GME problem—are contained in a negative semidefinite matrix, and provide an upper bound to the rank of the matrix. Moreover, this matrix takes a form that is familiar to students of neoclassical microeconomic theory, scilicet, that of the Slutsky-matrix of the theory of the consumer, and explain why this is so.

In deriving the above results, we take the position of the ‘‘least informed researcher’’, or equivalently, what we refer to as the ‘‘worst case scenario’’, terms we shall rigorously define in Section 3. For the moment it is sufficient to say that we take the position of a researcher who has little or no a priori knowledge about the true values of the coefficients and errors to be estimated. Not only is this the archetypal situation in applied work, especially when working with flexible functional forms for which the individual coefficients typically have no economic meaning in and of themselves, but it is also the instance when specification of the support intervals is most crucial. We first show that in general, a primal comparative statics analysis of the GME problem does not yield refutable implications for the effects of changes in the coefficient and error support bounds on the individual coefficient and error estimates, and explain why this is the case.

In spite of the aforementioned lack of refutable sensitivity results in the GME problem, it is not the case that refutable qualitative implications are not forthcoming in the GME problem. It turns out that a primal view of the GME problem leads one down a path of logic that prevents the discovery of its intrinsic and refutable comparative statics properties. The new comparative statics formalism of Partovi and Caputo (2006) is thus employed to uncover the intrinsic semidefinite matrix that characterizes the refutable comparative statics properties of the GME problem.

2. Literature review

The Monte Carlo evidence presented by Golan et al. (1996) is quite compelling and typically shows the ‘‘superiority’’ of the GME coefficient estimates over that of OLS and other estimators in the case of ill-conditioned design matrices. This evidence, however, is not entirely convincing in our view seeing as the true values of the coefficients are known in Monte Carlo experiments, thus permitting the researcher to select the support intervals so that the true values of the coefficients are always contained in the support interval, and thereby resulting in GME coefficient estimates that are often ‘‘superior’’ to that of the other estimators. Applied economists, on the other hand, never know the true values of the coefficients they are attempting to estimate and therefore do not have such additional a priori information when specifying the support intervals for the coefficients and errors. Given that this is the universal situation in applied work, it is manifestly important to come to some general understanding regarding the sensitivity of the GME coefficient and error estimates to the specification of their support intervals. As has already been noted, this is precisely the issue we address in the paper.

Golan et al. (1997) presented Monte Carlo evidence on the sensitivity of a mean squared error (MSE) loss function with respect to changes in the coefficient and error supports for a Tobit specification. Based on the