**Abstract**

In the estimation of linear and non-linear models, the generalized maximum entropy (GME) estimator has been found to outperform its traditional counterparts.

The generalized cross entropy (GCE) estimator is a generalization of GME that permits the incorporation of prior information. While informative priors can improve the performance of the GCE estimator relative to GME, prior information is often missing and achieving improved performance implies choosing among a possibly infinite number of priors. We thus develop a nested GCE estimator that utilizes the loss-minimizing or best-approximating reference distribution to improve upon the small sample performance of the GCE estimator.

**Introduction**

The generalized maximum entropy (GME) estimator seeks to recover the unknown probability distributions that characterize a given dataset by maximizing the Shannon (1948) entropy or information measure subject to stochastic moment conditions and proper probability constraints. In the estimation of both linear and non-linear models, the GME estimator has been found to outperform its traditional counterparts (e.g. ordinary least squares and maximum likelihood estimators), especially in the context of ill-posed, ill-conditioned, or noisy problems (Golan et al. 1996). As such, the GME estimator has witnessed widespread application in the social and natural sciences.

The generalized cross entropy (GCE) estimator instead uses the Kullback-Leibler (1951) divergence metric to minimize the discrepancy between the posterior probabilities and the researcher's chosen priors. For uniform priors, it can be shown that GME and GCE formalism yield identical solutions, and thus GCE is a generalization of the GME estimator (Golan et al. 1996). While more informative reference distributions can improve the performance of the GCE estimator relative to GME, such prior information is often missing and achieving improved performance implies choosing among a possibly infinite number of prior probability distributions. We thus identify a general basis for selecting among alternative reference distributions and develop a nested GCE (NGCE) estimator that improves upon the small sample performance of the GCE estimator.

In what follows, Section 2 discusses GCE formalism and Section 3 outlines our NGCE estimator. Section 4 details the sampling experiments through which we compare the performance of our estimation strategy to the leading alternative estimators. Section 5 elaborates upon the results of the experiments and Section 6 concludes.

**Nested Generalized Cross Entropy Estimator**

A central question in GCE estimation pertains to the choice of priors or reference distributions. Prior information is frequently incomplete or simply missing, and the researcher must choose among a possibly infinite number of reference distributions.

While in such situations priors are often specified as uniform (i.e.  and ), more informative reference distributions can enhance small sample performance, especially for ill-posed, ill-conditioned, or noisy problems. Accordingly, in this section we describe a general method for selecting among alternative reference distributions and outline an NGCE estimator in an effort to improve upon the small sample performance of the GCE estimator.

Consider the choice of prior in the context of GCE estimation of the simple linear model outlined above. Our NGCE estimator minimizes Eq. (3) subject to Eqs. (4)-(6) for (potentially many) alternative prior choices and then selects the model for which  is itself minimized. Intuitively,  can be interpreted as the quantity of information lost when  and  are used to approximate  and , respectively (Burnham and Anderson 2002). As such, among the alternative priors, the NGCE estimator identifies the loss-minimizing or best-approximating reference distribution, a property that can naturally be exploited to improve estimator performance. While computationally intensive, we demonstrate below that a relatively small number of priors is sufficient for the NGCE estimator to outperform its traditional counterparts …

**Results**

… Examining the above in greater detail, the first experiment in Table 3 displays baseline results for three alternative estimators: (1) our NGCE estimator; (2) the GME estimator; and (3) the OLS estimator. The NGCE and GME columns present the MSE() measures depicted in Figure 1 (i.e. for priors “6” and “7”, respectively). It is evident from Table 3 that GME outperforms OLS in terms of mean squared error loss, which is a well-known result (Golan et al. 1996). Further, we see that our NGCE estimator outperforms both GME and OLS, and the reduction in precision risk is non-negligible. For example, relative to GME, the NGCE estimator yields a 17 percent reduction in mean squared error loss. These relationships are maintained in experiments 2-5 and we see that the magnitude of the precision risk reductions decreases in *T*.

The largest reductions in precision risk are, however, associated with the noisy and ill-conditioned experiments. Rows 6-10 in Table 3 present the experiments with increased noise. Looking to experiment 6, which is the noisy case with *T*=10, we see that the GME estimator leads to a 72 percent reduction in precision risk relative to OLS. Further, relative to GME, the NGCE estimator reduces precision risk by approximately 48 percent. As is evident from experiments 7-10, we again see convergence in mean squared error as sample size increases. Similar results hold for the ill-conditioned cases presented in rows 11-15. For the *T*=10 case in row 11, we see that GME outperforms OLS (a 77 percent reduction in precision risk) and NGCE outperforms GME (a 33 percent reduction in precision risk). Relative to the noisy case, however, the precision risk reductions for the ill-conditioned case appear to dissipate faster as sample size increases.

We now turn to the relationship between the replication-average of  (i.e. the cross entropy of the signal) and MSE() for alternative choices of . For the baseline scenario, Figure 2 plots these two quantities for alternative prior choices and demonstrates a second core result: we find a strong positive relationship between the average of  and MSE(). Once again, the NGCE estimator (i.e. prior “6”) outperforms the GME estimator (i.e. prior “7”), and the first row of Table 4 implies a precision risk reduction of approximately 6 percent. Both estimators again outperform OLS and it is evident from experiments 2-5 that all estimators converge in mean squared error as the sample size increases.

Looking to the noisy cases presented in rows 6-10 in Table 4, we see once more that the mean squared error reductions are larger than the baseline case. Examining experiment 6, which is the noisy case with *T*=10, it is evident that GME leads to a precision risk reduction of 79 percent relative to OLS. Further, the NGCE estimator leads to precision risk reduction of approximately 22 percent relative to GME. These reductions again dissipate as sample size increases, as is evident from experiments 7-10. Finally, looking to the ill-conditioned cases in experiments 11-15, we see that GME continues to outperform OLS and NGCE generally outperforms GME. While in experiment 11 the performances of NGCE and GME are indeed identical, we do see some performance gains in other cases. For example, experiments 12 and 14 display 28 and 11 percent reductions in precision risk, respectively.

We conclude this section with a brief discussion of additional sampling experiments conducted. First, we increased the number of regressors from 3 to 7 and did not observe a qualitative change in our results in any case. Second, we incorporated alternative priors on  and found only modest precision risk reductions for the baseline and noisy cases, but non-negligible reductions in the ill-conditioned case (e.g. NGCE witnessed MSE()=7.85 for *T*=10). Finally, we doubled the parameter vector, which served to generate a situation where the parameter support was moderately ill-specified. While NGCE continued to outperform GME in this case, OLS outperformed both in select cases, most notably for the baseline specification. This final experiment underscores the importance of appropriately specifying the parameter support vector.