The Empirical Likelihood MPEC Approach to Demand Estimation

Christopher T. Conlon*
Columbia University

First version: January 2009
September 25, 2013

Abstract

The family of Generalized Empirical Likelihood (GEL) estimators provide a number of potential advantages relative to Generalized Method of Moments (GMM) estimators. While it is well known these estimators share an asymptotic distribution, the GEL estimators may perform better in finite sample, particularly in the case of many weak instruments. A relatively new literature has documented that finite-sample bias in the demand estimation problem of Berry, Levinsohn, and Pakes (1995) is often large, especially in the absence of exogenous cost shifting instruments. This paper provides a formulation for a computationally tractable GEL estimator based on the MPEC method of Su and Judd (2012) and adapts it to the BLP problem. When compared to GMM, the GEL estimator performs substantially better, reducing the bias by as much as 90%. Furthermore, it is possible to use analytic bias correction to reduce the bias even more and obtain accurate estimates with relatively small numbers of markets.

*cconlon@columbia.edu. An early version of this paper was part of my doctoral dissertation at Yale University. The author would like to thank Steve Berry, Phil Haile, and Yuichi Kitamura. I have also benefited from discussions with: Bernard Salanie, Aviv Nevo, Che-Lin Su, Jeremy Fox, Tim Armstrong, and Ben Skrainka. The original formulation of the estimator was developed while attending the Institute on Computational Economics. The computations in the paper were performed on the Hotfoot cluster at Columbia University. Joonwoo Park provided excellent research assistance. Any remaining errors are my own.

1 Introduction

Economists often rely on structural models to estimate parameters from economic data and make inferences about equilibrium behavior under alternate conditions. Often, these models are estimated using the Generalized Method of Moments (GMM) approach of Hansen (1982). While powerful, a well understood drawback of the GMM approach is that it can exhibit bias in finite samples as the number of moment conditions increases (Bekker 1994). The bias can be substantial if there are a large number of moment conditions (Hansen and Singleton 1983), or those moment conditions become correlated (Abowd and Card 1989). This finite sample bias has been documented in Monte Carlo studies by Qin and Lawless (1994) and Altonji and Segal (1996).

There have been a number of attempts to improve upon the finite sample performance of GMM, including the Generalized Empirical Likelihood (GEL) family of estimators as described in Newey and Smith (2004). Some example of GEL estimators include the Empirical Likelihood (EL) estimator of Owen (1988), Qin and Lawless (1994); the exponential tilt (ET) estimator of Kitamura and Stutzer (1997), and Imbens, Spady, and Johnson (1998); and the continuously updating GMM estimator (CUE) of Hansen, Heaton, and Yaron (1996). These estimators possess the same asymptotic distribution as the GMM estimator, hence the same inference procedures apply, but they differ in higher-order terms, and finite-sample properties. These estimators are also understood to be partially robust to the problem of many weak instruments (Imbens 2002), (Newey and Windmeijer 2009).

Despite potential advantages, GEL estimators are not frequently used in applied work. One potential explanation is that GEL-type estimators are perceived to be computationally challenging, because they require estimating many additional parameters when compared to GMM estimators, and require optimization of a saddle-point problem. In cases where GEL estimators have been implemented on applied problems, it has generally been limited to either linear models or relatively simple nonlinear models.

This paper does two things, the first is to provide a computationally tractable version of GEL estimators based on the constrained optimization (MPEC) approach of Su and Judd (2012) and applied to the discrete choice demand estimation problem of Berry, Levinsohn, and Pakes (1995); the second is to compare the finite-sample performance of GEL and GMM under commonly used (and possibly weak) instruments.

One motivation for choosing the BLP problem is that it is an excellent example of an estimator that is widely used in the literature, but also known to be computationally challenging. There has been increased interest in the numerical properties of the BLP problem, since with the highly-cited work of Nevo (2000). For example, Knittel and Metaxoglou (2013) describe numerical challenges in obtaining the BLP estimator, and explore the potential of multiple maxima. Judd and Skrainka (2011) examine the effects of simulation error, and provide improved numerical integration methods for the BLP problem. Freyberger (2012) develops analytical bias corrections for the simulation error. Dubé, Fox, and Su (2012) provide a characterization of the BLP problem based on the MPEC method of Su and Judd (2012), and show that while it possesses the same econometric properties as the original BLP estimator, in many cases it may have improved numerical performance while also being easier to compute.

There has been a recent literature establishing the identification of a larger class of discrete choice models which includes the BLP problem as a special case. Recent work includes Fox and Gandhi (2012), Gandhi, Berry, and Haile (Forthcoming). Nonparametric identification has been demonstrated for a larger class of estimators that includes the micro-data BLP estimator of Berry, Levinsohn, and Pakes (2004) was established in Berry and Haile (2012b). Berry and Haile (2012a) establish nonparametric identification results in the case of aggregate (market) data, and examine the role that both exogenous cost shifting instruments, and characteristics of other products (BLP instruments) play in identification

There is also a literature on the econometric properties of the BLP problem, Berry, Linton, and Pakes (2004) study the asymptotic properties as the number of products $J \to \infty$. In other recent work, Armstrong (2013) relates those large market asymptotic results to the underlying economic primitives. That paper demonstrates that as J becomes large, the markup converges to a constant; instruments based on characteristics of competing products ("BLP-instruments") become weak and the problem may no longer be well identified. In a Monte Carlo analysis, Armstrong (2013) shows this leads to substantial finite-sample bias. In related work, Skrainka (2012) conducts an extremely large Monte Carlo study and shows that in many cases the finite sample bias of the two-step GMM BLP estimator may be so large that asymptotic standard errors no longer reasonably describe the performance of the estimator, and suggests that the bias may non-decreasing in the sample size.

One way to interpret these results is through the lens of the weak instruments literature (see Stock, Wright, and Yogo (2002) for a survey), where Stock and Wright (2000) describe the problem of weak identification in the GMM context. In the presence of weak instruments, the bias does not disappear as the sample size grows. Reynaert and Verboven (2012) consider optimal instruments in the form of Chamberlain (1987) for the BLP problem with an additional assumption of perfectively competitive supply, and show that under that assumption, finite sample bias found by Skrainka (2012) may be decreasing in sample size when optimal instruments are used. Standard practice in the demand estimation literature is to consider higher order functions of existing instruments, which act as a sieve approximation to the optimal instruments; this increases the number of instruments, but may also increase the correlation across moments. Again, GEL estimators are expected to outperform GMM estimators in the case of many weak instruments.

This paper shows that when (strong) cost-shifting instruments are employed, GMM and GEL estimators all perform quite well, even in relatively small samples. When only (potentially weak) "BLP Instruments" are employed, the bias can be quite large for small sample sizes (10 markets or less) but as the sample size increases (20 or more markets) the GEL estimators perform reasonably well, while the GMM estimator remains substantially biased. The Monte Carlo study indicates that the primary source of the bias is correlation between the BLP moment conditions and their Jacobian, which arises from the endogeneity of prices. This correlation is eliminated in the GEL estimators. An additional source of bias arises from the commonly chosen functional forms (higher powers of instruments) which can lead to an ill-conditioned weighting matrix. Thus, an additional advantage of the MPEC-GEL formulation is that it never explicitly requires computing the optimal weight matrix.

The organization of the rest of the paper is as follows. First, I present the BLP problem, explain the source of instruments, and relate it to the theoretical work on finite sample bias in Newey and Smith (2004). Then, I develop a computational technique for estimating GEL problems based on the MPEC method of Su and Judd (2012) that is not substantially more difficult to estimate than the GMM estimator. Next, I conduct a Monte Carlo analysis similar to the example in Armstrong (2013) and show that the GEL estimators reduce the bias by up to 90% when using only

"BLP Instruments". Finally, I construct confidence intervals by inverting the ELR test-statistc and show that when only "BLP Instruments" are available, identification may be weak in small samples, but generally well-identified in larger samples. In addition to Monte Carlo analysis, I demonstrate that the often cited pseudo-real cereal example of Nevo (2000) may suffer from weak identification, which provides an alternate interpretation for the results in Knittel and Metaxoglou (2013).

2 Preliminaries

2.1 Motivation and BLP Problem

We begin by reviewing the estimator proposed by Berry, Levinsohn, and Pakes (1995), and describing the potential sources of finite sample bias. Consider a consumer i choosing among J products $j = 1, \ldots, J$ and an outside good $j = 0$ over $t = 1, \ldots, T$ markets. Each product is described by characteristics x_{jt} and each consumer has tastes for characteristics $\beta_i \sim F(\beta|\theta)$. Consumers have a random idiosyncratic horizontal preference for each product $\varepsilon_{ij} \sim$ Type I EV, which allows aggregate shares s_{jt} to be obtained via integration.

$$
\begin{aligned}
u_{ijt} &= x_{jt}\beta_i + \xi_{jt} - \alpha_i p_{jt} + \varepsilon_{ij} \\
s_{jt}(\theta, \xi) &= \int \frac{x_{jt}\beta_i + \xi_{jt} - \alpha_i p_{jt}}{1 + \sum_k x_{kt}\beta_i + \xi_{kt} - \alpha_i p_{kt}} f(\beta_i|\theta) \qquad (1)
\end{aligned}
$$

In many common specifications, $f(\beta_i|\theta)$ is a multivariate normal distribution with a diagonal covariance matrix so that the integral in (1) can be computed with Monte Carlo integration or quadrature rules (Judd and Skrainka 2011) (Heiss and Winschel 2008) with weight w_i and nodes ν_{il}:

$$
\begin{aligned}
s_{jt}(\theta, \xi) &= \sum_{i=1}^{ns} w_i s_{ijt} = \sum_{i=1}^{ns} w_i \frac{\overbrace{x_{jt}\beta + \xi_{jt} - \alpha p_{jt}}^{\delta_{jt}} + \sum_l \nu_{il}\sigma_l x_{jt}^l}{1 + \sum_k x_{kt}\beta + \xi_{kt} - \alpha p_{kt} + \sum_l \nu_{il}\sigma_l x_{kt}^l} \qquad (2) \\
\xi_{jt} &= \delta_j(\mathbf{s_t}, \mathbf{x_t}, \mathbf{p_t}, \sigma) - \alpha p_{jt} - x_{jt}\beta \qquad (3)
\end{aligned}
$$

There is a one-to-one correspondence between the marketshares and the ξ_{jt} (unobservable product quality). This structural error is (potentially) correlated with prices, but conditionally independent of instruments so that $E[\xi_{jt}|z_{jt}] = 0$ or $E[\xi_{jt} f(z_{jt})] = 0$

where z_{jt} is a $1 \times M$ dimensional vector of instruments, and $f(\cdot)$ is an arbitrary function with four finite moments. To construct instruments for unobservable quality we consider the profit maximization problem of the firm.

$$\max_{p_j : j \in \mathcal{J}_f} \sum_{j \in \mathcal{J}_f} (p_j - c_j) s_j(\mathbf{p_t})$$

Under the assumption of constant marginal costs c_j the FOC for firm f and price p_j becomes:

$$s_j(\mathbf{p}) + \sum_{j \in \mathcal{J}_f} (p_k - c_k) \frac{\partial s_k(\mathbf{p_t})}{\partial p_j} = 0 \tag{4}$$

The intuition is clearer in the case of single product firms where the markup equation has the well-known form:

$$p_{jt} = c_{jt} + \frac{s_{jt}(\mathbf{p_t})}{\left| \frac{\partial s_{jt}(\mathbf{p_t})}{\partial p_{jt}} \right|} \tag{5}$$

This suggests two potential sources of instruments for prices; the first of which are variables that directly shift c_{jt}. The second are non-price characteristics that influence markups. The elasticity depends not only on own-product characteristics, but also on the characteristics of other products in the same market. The intuition from (4) is that characteristics of products owned by the same firm and characteristics of products owned by other firms affect markups differently.[1] This suggests (at least) three sets of markup shifting instruments or "BLP Instruments" (in addition to cost shifting instruments): $z_{jt}^{BLP} = [x_{jt}, \sum_{k \in J_f \setminus j} h(x_{kt}), \sum_{k : k \notin J_f} h(x_{kt})]$, where $h(\cdot)$ is some bounded continuous function with finite variance (commonly the sample average).

Armstrong (2013) derived three main results for these instruments as the number of products $J \to \infty$. The first is that for single product firms (5) converges to a constant markup as $J \to \infty$. The intuition comes from the plain logit model where (5) approaches a constant: $c_j + \frac{1}{|\alpha(1-s_{jt})|}$.[2] Obviously as the markup approaches a constant, constructing markup shifting instruments becomes impossible. His second result shows for a fixed number of firms F and an increasing number of products

[1] This intuition for the cost shifting and BLP instruments is formalized in Berry and Haile (2012a).
[2] The result itself applies to a more general class of mixtures over logits.

$J \to \infty$, markups converge to a firm-specific constant, and instruments of the form $\sum_{k \in J_f \setminus j} h(x_{kt})$ are able to exploit only variation between firms but not within firms. His third result shows that as the number of markets T grows sufficiently fast, with $N = T \cdot J$, $\frac{N}{J} \to \infty$, identification is possible but relies on differences in the number of products per market $\sum_{k \neq j} h(x_{kt})$, and $\frac{N}{J} \to c$ implies weak-instrument asymptotics.

The negative results presented by Armstrong (2013) relate the finite-sample bias to behavior of markups and the power of instruments as $J \to \infty$.[3] This study looks more closely at the choice of the instruments z_{jt}, especially in the absence of exogenous cost shifters (which are often unavailable in the empirical settings). Recall that the BLP moment conditions take the form $E[\xi_{jt}|z_{jt}] = 0$, but are generally estimated via GMM using the weaker orthogonality restriction $E[g_i(z;\theta_0)] = 0$, so that:

$$\overline{g}(z,\theta) = \frac{1}{N}\sum_{j,t} g(z_{jt};\theta) = \frac{1}{N}\sum_{j,t} z_{jt}\xi_{jt} = \frac{1}{N}\sum_{j,t} z_{jt}(\delta_{jt} - X'_{jt}\beta - \alpha p_{jt}) \tag{6}$$

One challenge, as described by Domínguez and Lobato (2004) is that a model identified by a conditional moment restriction may be non-identified under a non-trivial set of unconditional moments (including the optimal instruments). For GMM estimators, this creates a tradeoff between constructing a basis to approximate conditional moment restrictions (or optimal instruments) using higher powers, and bias that increases linearly in the number of moment conditions. The behavior of the bias is formally established in Newey and Smith (2004), and takes the form:

$$\begin{aligned} Bias(\hat{\theta}_{GMM}) &= B_G + B_\Omega + B_I \\ Bias(\hat{\theta}_{CUE}) &= B_\Omega + B_I \\ Bias(\hat{\theta}_{EL}) &= B_I \end{aligned} \tag{7}$$

[3] An omnipresent alternative is a potentially mis-specified model of supply.

When applied to the BLP problem these yield the following expressions:

$$
\begin{aligned}
B_G &= -\Sigma \left(\frac{1}{J \cdot T} \sum_{j=1}^{J} \sum_{t=1}^{T} \left[X_{jt} \; p_{jt} \; \frac{\partial \xi_{jt}}{\partial \sigma} \right]' z_{jt} P z_{jt} \xi_{jt} \right) \\
B_\Omega &= H \left(\frac{1}{J \cdot T} \sum_{j=1}^{J} \sum_{t=1}^{T} \xi_{jt}^3 z_{jt} z_{jt}' P z_{jt} \right) \\
B_I &= H \left(\frac{1}{J \cdot T} \sum_{j=1}^{J} \sum_{t=1}^{T} z_{jt} \left[X_{jt} \; p_{jt} \; \frac{\partial \xi_{jt}}{\partial \sigma} \right] H z_{jt} \xi_{jt} \right)
\end{aligned} \tag{8}
$$

B_G results from correlation between the moment conditions and their Jacobian, and arises because unobservable quality ξ_{jt} is correlated with prices p_{jt}. The Monte Carlo study in the later suggestion suggests that this is the primary source of the bias in the BLP estimator.[4]

A second challenge arises from the choice of the instruments z_{jt} and the behavior of the optimal weighting matrix. For the BLP problem, the (true) optimal GMM weighting matrix takes the form: $\Omega_0 = E[g_i(z_i, \theta) g_i(z_i, \theta)'] = \frac{1}{N} \sum_{j,t} z_{jt} \xi_{jt}^0 \xi_{jt}^{0\prime} z_{jt}'$. Most of the econometric literature (including (Hansen 1982) and (Newey and Smith 2004)) assumes that Ω_0 is full column rank, and that its minimum eigenvalue is bounded away from zero. For many formulations of the BLP instruments, including those used in Dubé, Fox, and Su (2012), Nevo (2000), Armstrong (2013) the condition number (ratio of largest to smallest eigenvalue) is quite large. The typical condition numbers for those examples are on the order of $(10^8, 10^8, 10^7)$, while condition numbers on the order of 10^4 or greater are considered ill-conditioned. This is problematic that small changes in the sample covariance of the moments may lead to large changes in its inverse. One advantage of the MPEC-GEL formulation presented later is that it avoids explicit inversion of the weighting matrix. More detail on conditioning (and some potential solutions) are provided in the Appendix.

2.2 Review of Empirical Likelihood

Empirical likelihood methods were first established by Owen (1988). Owen (1990) demonstrated that nonparametric maximum likelihood estimation (NPMLE) could

[4]Derivation of analytical expressions for the bias in the BLP problem can be found in the Appendix.

be adapted to moment condition models while retaining many of the attractive properties testing and efficiency properties of MLE. Kitamura (2006) provides a detailed review of the empirical likelihood literature, including derivations via NPMLE and Generalized Minimum Contrast (GMC), statistical tests, and extensions. Newey and Smith (2004) show that empirical likelihood estimators are often preferable to GMM estimators, particularly for models with many moment conditions, or correlated moment conditions where the asymptotic bias in GMM can be problematic.

What follows is a standard exposition for empirical likelihood in the case of moment condition models, for more details, consult (Kitamura 2006). We begin by assuming some data $\{z_i\}_{i=1}^n$ which are distributed IID according to unknown measure μ. Our econometric model gives us some moment conditions. That is $g(z, \theta) \in \mathbb{R}^m$ may be scalar valued ($m = 1$) or it may be vector valued ($m > 1$).

$$E[g(z_i, \theta)] = \int g(z, \theta) d\mu = 0, \quad \theta \in \Theta \in \mathbb{R}^k$$

We can assume (w.l.o.g.) the existence of a nonparametric (multinomial) (log) likelihood function where we search for a measure (set of probability weights) $(\pi_1, \ldots, \pi_n) \in \Delta$ the unit simplex, so that each π_i corresponds to the weight of associated with each z_i observed in the data.

$$\hat{\theta}_{EL} = \arg \max_{\theta, \pi_1, \ldots, \pi_n} \sum_{i=1}^n \log \pi_i \text{ s.t. } \quad \sum_{i=1}^n \pi_i g(z_i, \theta) = 0, \text{ and } (\pi_1, \ldots, \pi_n) \in \Delta \tag{9}$$

We search for a measure Π which is consistent with the model (satisfies the moment restrictions) and is as close as possible to the empirical measure μ. In the absence of any moment restrictions (that is when $E[g(z_i, \theta)] = 0 \quad \forall \theta$), the solution is trivial and we have that $\hat{\theta}_{GEL} = \frac{1}{n}$ or that the optimal weights are the observed empirical weights in the data.

The problem presented in (9) presents some estimation challenges. The first challenge is that the dimension of the problem is $N+K$, thus as the number of observations increases, so does the number of unknown parameters. The second challenge is that there are a number of constraints, including both the moment conditions and the π parameters. These are typically addressed by considering the dual formulation of the problem. We obtain this by writing down, and then differentiating the Lagrangian.

$$\begin{aligned}\mathcal{L} &= \sum_{i=1}^{n} \log \pi_i + \lambda(1 - \sum_{i=1}^{n} \pi_i) - n\gamma' \sum_{i=1}^{n} \pi_i g(z_i, \theta) \\ \hat{\gamma}(\theta) &= \arg\min_{\gamma \in \mathbb{R}^q} - \sum_{i=1}^{n} \log(1 + \gamma' g(z_i, \theta)) \\ \hat{\pi}_i(\theta) &= \frac{1}{n(1 + \hat{\gamma}(\theta)' g(z_i, \theta))}, \quad \hat{\lambda} = n\end{aligned}$$

After substituting the above equations, we obtain the following dual (saddle point) problem:

$$\hat{\theta}_{EL} = \arg\max_{\theta \in \Theta} l_{NP}(\theta) = \arg\max_{\theta \in \Theta} \min_{\gamma \in \mathbb{R}^q} - \sum_{i=1}^{n} \log(1 + \gamma' g(z_i, \theta)) \tag{10}$$

The dual formulation is an unconstrained problem with $m + K$ parameters (one γ for each moment condition, plus the model parameters). For this reason it is the formulation that most researchers prefer to work with.

The standard approach to solving the problem presented in (10) is to break it up into two nested steps. The first step is referred to as the *inner loop* which fixes a guess of θ and searches for the Lagrange multipliers which solve $\min_\gamma - \sum_{i=1}^n \log(1 + \gamma' g(z_i, \theta))$. The other step, referred to as the *outer loop*, searches for the $\max_\theta$ of the profiled likelihood $l_{NP}(\theta)$. The next section highlights some of the challenges associated with this approach.

2.3 Challenges of the Dual Formulation

There are a number of numerical and computational challenges associated with empirical likelihood estimators that can make them difficult to implement on real world data. One of the key challenges is what is known as the *convex hull problem.* This arises when, for a given value of the model parameters θ, the inner problem (linear programming) is infeasible. That is, there is no set of weights for which the moment conditions can be satisfied, or $\nexists(\pi_1, \ldots, \pi_n) \in \Delta$ such that $\sum_{i=1}^n \pi_i g(z_i, \theta) = 0$. In the language of the dual, this means the zero vector, $0 \notin co\{g(z_1, \theta) \ldots g(z_n, \theta)\}$ where $co(\cdot)$ denotes the convex hull of the vector space. When this occurs, the search for the Lagrange multipliers γ fails, no minimum exists, and both the likelihood and its derivatives are undefined.

The second challenge is that the problem posed by (10) is *domain restricted*. The distinction between *equilibrium constraints* and *domain restrictions* is that an equilibrium constraint only needs to be satisfied at an optima, whereas a domain restriction must be satisfied everywhere. When the domain restriction is violated, the objective function (and its derivatives) are undefined. When an equilibrium constraint is violated, the objective (and the constraints) and all the derivatives are well defined. The domain restrictions of the EL problem depend on the formulation. The primal problem has the restriction that $\pi_i \geq 0$. This is a fixed domain restriction, because when $\pi_i < 0$, then $\log \pi_i$ is undefined. This restriction is easy to handle by placing bounds on the parameter values. The second restriction is that for some values of (γ, z_i) the argument of $\log(\cdot)$ is negative or $\gamma' g(z_i, \theta) \leq -1$. This is more difficult to deal with, because the domain restriction varies with the choice of the parameter θ. We do not know in advance where the objective function will be defined or undefined. Any search procedure we use must be able to deal with an undefined objective function by backtracking or restarting.

The third challenge is that the problem posed by (10) is a *saddle-point problem*. There are very few off the shelf optimization routines developed for saddle-point problems. Most researchers, solve this problem using the inner-loop/outer-loop iterative procedure previously described rather than a single optimization problem. Notwithstanding the two challenges posed above, the inner loop is a M dimensional optimization problem, and is convex when $g(z_i, \theta)$ is convex (linear moment conditions), so quasi-Newton procedures can work quite well. The outer loop is a K dimensional nonlinear search, and is much more difficult for two reasons. The first is that the objective function is not explicitly defined, but is rather the implicit solution to an optimization problem, which makes obtaining derivatives difficult, especially when the moment conditions do not have analytic expressions. The second challenge is that the resulting problem (the max of a min) is non-convex. This combination of challenges leads researchers to employ custom routines (often using non-derivative based methods for the outer loop, and backtracking when the problem becomes "stuck" or infeasible) to estimate even fairly simple empirical likelihood models.

2.4 MPEC Formulation of GEL

If we work with the primal version of the empirical likelihood estimator we defined in (9), we notice that this problem can be expressed directly as an MPEC problem:[5]

$$\begin{aligned} \hat{\theta}_{EL} &= \arg\min_{\theta,\pi_1,\ldots,\pi_n} \sum_{i=1}^{n} f(\pi_i) \quad \text{s .t.} \\ & \sum_{i=1}^{n} \pi_i g(z_i,\theta) = 0 \\ & \sum_{i=1}^{n} \pi_i = 1 \quad -\pi_i \leq 0 \\ & \Psi(P,\theta) = 0 \end{aligned} \tag{11}$$

The formulation is (11) has not specified a choice for $f(\pi_i)$. If $f(\pi_i) = -\log(\pi_i)$ this produces the empirical likelihood estimator, if $f(\pi_i) = \pi_i \log(\pi_i)$ this corresponds to the exponential tilting (ET) estimator, and $f(\pi_i) = \pi_i^2$ corresponds to CUE. Together these are all different cases of Generalized Empirical Likelihood (GEL) estimators (Newey and Smith 2004).

When $f(\pi_i)$ is concave, this is a convex optimization problem with the exception of the moment condition constraint. When that constraint is affine, the entire optimization problem in (11) would be convex. For simple cases like instrumental variables $g(z_i,\theta)$ is quadratic, thus the resulting problem becomes a convex objective function with quadratic constraints.[6]

This primal problem can be solved using the large-scale constrained optimization software, employing the MPEC method of Su and Judd (2012). The optimization problem is now an $N + K$ dimensional search, whereas the dual was a M inside a K dimensional search. Many good commercial solvers, such as KNITRO or IPOPT, are able to solve convex objectives with quadratic constraints for very large problems (tens of thousands of parameters). Unlike the saddle point problem, which must

[5] It should be noted that this is definitely not the first presentation of the primal EL/GEL problem. Owen (1988) provides a statement of the primal problem (and its derivatives) on p.241 of his book, though it gets seldom used in practice. This presentation adds the (optional) additional equilibrium constraint $\Psi(P,\theta)$ as in Su and Judd (2012)

[6] The CUE estimator is of the form $f(\pi_i) = \pi_i^2$ and is an even simpler quadratically constrained quadratic programming problem (QCQP). This is substantially less complicated that the traditional formulation of CUE in Hansen, Heaton, and Yaron (1996) which is highly non-convex, and prone to multiple maxima.

be solved in a nested manner, the MPEC formulation can be directly solved by the optimizer. The second advantage is that the domain restriction has been eliminated, and the only remaining restrictions are the bounds restrictions on $\pi_i \geq 0 \; \forall i$.

It is important to notice that this estimator is not statistically different from the dual (saddle point) estimator. In both (11) and (10) the same $\hat{\theta}$ solves both optimization problems. Any differences between $\hat{\theta}$ obtained from the two approaches should be attributed to numerical error in the optimization routines (assuming both routines find the optimum) rather than statistical sampling. In fact, many software packages solve linear programming problems by re-formulating and solving the problem using duality. However, the MPEC formulation makes certain kinds of inference procedures easier to implement (especially restrictions that can be directly incorporated into the optimization problem). More detail is provided in Appendix A.1.

3 Estimation Algorithms

This section describes the two GMM-based approaches to estimating the BLP problem, as well as the new EL-MPEC approach. In this section, we review existing algorithms for obtaining the estimator of Berry, Levinsohn, and Pakes (1995) and then show how to adapt the BLP problem to the MPEC-EL framework.

GMM-Fixed Point (Berry, Levinsohn, and Pakes 1995)
The traditional (NFXP) estimator solves the following mathematical program, beginning with partitioning the parameter space between linear parameters and the variance of the random coefficients $\theta = [\underbrace{\alpha, \beta}_{\theta_1}, \theta_2]$:

$$\begin{aligned} \min_{\theta_2} \quad & \psi' \Omega^{-1} \psi \quad \text{s.t.} \\ \psi &= \xi(\theta_2)' Z \\ \xi_{jt}(\theta) &= \delta_j(\theta_2) - x_{jt}\beta - \alpha p_{jt} \\ \log(S_{jt}) &= \log(s_{jt}(\delta, \theta_2)) \end{aligned} \tag{12}$$

The solution algorithm uses a fixed point to reduce the parameter space. For each guess of θ_2 (nonlinear parameters) the $J \times T$ system of nonlinear equations $\log(S_{jt}) =$

$\log(s_{jt}(\delta, \theta_2))$ is solved for δ via the following iterative procedure:[7]

$$\delta_{jt}^{h+1} = \delta_{jt}^{h} + \log(S_{jt}) - \log(s_{jt}(\delta^h, \theta_2))$$

Once δ is obtained, ξ_{jt} is obtained as the residual of a two step IV-GMM regression of δ on the observables x_{jt}. This is then used to construct the GMM objective function, which is often optimized with a gradient free procedure such as Nelder-Mead, over θ_2 only. Dubé, Fox, and Su (2012) provide a detailed analysis of this algorithm as well as number of potential pitfalls and problems. When the dimension of θ_2 is small (such as a single random coefficient) this may still be the preferred estimation method.

GMM-MPEC (Dubé, Fox, and Su 2012)

The same statistical estimator can be rewritten as a constrained problem:

$$\begin{aligned} \min_{\theta_2, \alpha, \beta, \xi, \psi} \quad & \psi' \Omega^{-1} \psi \quad \text{s.t.} \\ \psi &= \xi' Z \\ \log(S_{jt}) &= \log(s_{jt}(\xi, \theta_2, \alpha, \beta)) \end{aligned} \tag{13}$$

The objective function is depends only on ψ, and none of the other parameters. This is a quadratic objective function with a series of $N = J \times T$ nonlinear constraints, and N linear constraints. The principal advantage of this approach is that we can use a derivative-based Quasi-Newton approach for optimization, and simultaneously solve the share equations while minimizing the objective function. The key to the MPEC approach is that modern optimization software it is able to exploit the sparsity of the problem. Here the sparsity arises because ξ_{jt} only enters the marketshares of products

[7] It is important to note that this part is entirely optional. The algorithm is unchanged regardless of how δ is obtained. Some have suggested using Newton's method, others have proposed modified fixed point iterations see (Reynaerts, Varadhan, and Nash 2012). It is also important to note that the contraction mapping induces noisy function evaluations unless it starts at the same δ^0 each time. That is evaluations of the objective function at the same θ will not necessarily be exactly the same. While this seems obvious, the Nevo (2000) code example employed by Knittel and Metaxoglou (2013) uses the δ vector from the previous iteration to reduce iterations of the contraction mapping. This is disastrous for finite difference derivatives, and is quite problematic for convergence overall. This perhaps provides some explanation why Knittel and Metaxoglou (2013) found SOLVOPT, a noisy function optimizer based on the subgradient method worked the best, or why Dubé, Fox, and Su (2012) recommend setting the contraction mapping tolerance close to machine precision $1e-14$.

in the same market t. Dubé, Fox, and Su (2012) show that this formulation is often easier to compute and that error in the share equations (either from optimization or integration) does not have as large an impact on parameter estimates as in the NFXP case. The computational advantage of the MPEC approach over NFXP is increasing in the number of nonlinear parameters.

GEL-MPEC

The MPEC version of the Generalized Empirical Likelihood Estimator looks slightly different:

$$\begin{aligned}
\min_{\theta_2,\alpha,\beta,\xi,\pi} & \quad \sum_{j,t} f(\pi_{jt}) \quad \text{s.t.} \\
0 &= \sum_{j,t} \pi_{jt}\xi_{jt}z_{jt}^q \quad \forall q \\
& \quad \sum_{i=1}^{n} \pi_i = 1 \quad -\pi_i \leq 0 \\
\log(S_{jt}) &= \log(s_{jt}(\xi,\theta_2,\alpha,\beta))
\end{aligned} \tag{14}$$

This problem is somewhat more challenging than the GMM problem above. The objective function is still convex, but no longer quadratic (except in the case of CUE). Instead of N linear moment constraints, the introduction of the EL weights π results in N quadratic constraints. In practice, because the difficulty usually lies in solving the N non-linear potentially non-convex share equations, this problem is not appreciably more difficult to solve than the GMM-MPEC approach.

Concentrated GEL-Primal

In some cases the dimension of the nonlinear parameters is small, and it may be easier to reformulate the Empirical Likelihood estimator as a hybrid between the full MPEC approach above and something more similar to the traditional NFXP approach employed for GMM estimators.

$$\begin{aligned}
\arg\min_{\theta_2} GELP(\theta_2) &= \min_{\alpha,\beta,\xi,\pi} \sum_{j,t} f(\pi_{jt}) \quad \text{s.t.} \\
0 &= \sum_{j,t} \pi_{jt}\xi_{jt} z_{jt}^q \\
\xi_{jt} &= \delta_j(\theta_2) - x_{jt}\beta - \alpha p_{jt} \\
\log(S_{jt}) &= \log(s_{jt}(\delta,\theta_2))
\end{aligned} \tag{15}$$

In this approach, for a guess of the nonlinear parameters θ_2 we solve the system of share equations for $\delta(\theta_2)$ (perhaps using the contraction mapping of Berry, Levinsohn, and Pakes (1995)). Given $\delta(\theta_2)$ the resulting primal formulation of the empirical likelihood problem is a globally convex optimization problem with quadratic constraints, which is trivial to solve for $(\alpha, \beta, \xi, \pi)$ and produces the profiled GEL function $GELP(\theta_2)$. As long as the dimension of θ_2 is small, it is possible to minimize the profiled likelihood over θ_2.[8]

For purposes of completeness, a (computationally infeasible) dual formulation of the EL problem is presented in the Appendix.

4 Empirical Examples

4.1 Monte Carlo Study

In a Monte Carlo study, I replicate the data generating process in Armstrong (2013). One of the nice features of this setting is that it explicitly considers the differences between exogenous cost shifting instruments, and BLP-type markup shifting instruments (including average characteristics of other products). Another nice feature not present in many of the other Monte Carlo studies of the BLP problem (with the exception of Skrainka (2012)) is that prices are not randomly generated via a reduced form, but endogenously determined as the solution to the firms' optimization problem. This is crucial because it makes explicit the relationship between markups and characteristics of other products, which is the justification of the "BLP Instruments".

Armstrong's setting is very simple in that it considers only a single exogenous product characteristic x_{jt} and a single cost shifter z_{jt} as $U[0, 1]$ random variables.

[8] I recommend this approach for one or two nonlinear parameters.

Three more standard uniform random variables u_{1j}, u_{2j}, u_{3j} are constructed where $\xi_j = 0.9 \cdot u_{1j} + 0.1 \cdot u_{3j} - 1$ and $\eta_j = 0.9 \cdot u_{1j} + 0.1 \cdot u_{2j} - 1$. Marginal costs are constructed as $MC_j = (x_j', z_j')\gamma + \eta_j$ with parameters $\gamma' = (2, 1, 1)$, and utilities are constructed with parameters $(1, 3, 6)$ corresponding to price, the constant term, and x_{jt}.[9] Consumers vary in their taste for x_{jt} as determined by a normal distribution with mean zero and $\sigma = 3$.[10] After generating marginal costs and consumer utilities, I solve for endogenous prices and shares following the modified fixed-point formula of Morrow and Skerlos (2011).[11]

I consider several different market sizes $T = \{3, 10, 25, 50, 100\}$ and numbers of products per market $J = \{20, 60, 100\}$ and $F = 10$ firms per market. For each setting I generate 100 fake datasets. Where this study departs from Armstrong (2013), is that he considered a just-identified GMM setting. I consider an over-identified setting with three different sets of instruments.

The full set of instruments includes: a constant, the regressor x_{jt}, the exogenous cost shifter z_{jt} and its square z_{jt}^2, and two BLP instruments: the average x_{jt} of all other products in market t, $\overline{x}_{-j,t}$, and the average x_{jt} of products owned by other firms. I then consider a reduced set of instruments that does not include cost shifters z_{jt}, z_{jt}^2, but does include quadratic interactions of the other instruments; I label this "BLP Only" instruments. For the third set of instruments, I consider a regularized (lower dimensional) version of the BLP Only instruments. More specifically, I construct the principal components of the Z-scores for the set of BLP instruments. I follow Carrasco (2012) and select principal components corresponding to the largest eigenvalues (Spectral Cutoff) and label them "PC Instruments". The hope is that these solve two potential problems. One is that bias in GMM grows in the number of instruments, so reducing the number of instruments should alleviate the bias. The

[9] At these parameters, the outside good share is very small, and the Lipschitz constant is close to one. As shown in Dubé, Fox, and Su (2012) this implies that the rate of convergence of the traditional BLP contraction mapping may be very slow. MPEC or NFP and Newton's method are unaffected by the Lipschitz constant.

[10] Because there is only a single random coefficient, I use quadrature rules that are exact to 16 degrees of polynomial accuracy. This effectively eliminates simulation error in this problem. For a discussion of quadrature rules in mixed logit demand settings, refer to Heiss and Winschel (2008) or Judd and Skrainka (2011).

[11] Both Armstrong (2013) and Skrainka (2012) mention that in some cases Newton's method fails to find an equilibrium vector of prices and quantities. This was not my experience with the ζ-fixed point approach of Morrow and Skerlos (2011), which is both faster and more robust than Newton's method for solving Bertrand-Nash pricing games. The modified fixed-point proposed in that work is similar in spirit to Skrainka's suggestion of normalizing by the marketshares.

second is that the principal components provide an orthogonal basis which guarantees the weighting matrix is well-conditioned. This relationship is not explicitly considered in Carrasco (2012) but is similar in spirit to Knight and Fu (2000), Caner and Yildiz (2012), or Caner (2008). The three sets of instruments are of dimension $\{6, 7, 5\}$ respectively.

In Table 1, I report the condition number (ratio of largest to smallest eigenvalues) for the optimal GMM weighting matrix evaluated at the true ξ_0. The condition number appears to be (slowly) decreasing in the number of markets, but the largest difference seems to be across instrument sets; with the Cost Instruments and Principal Components BLP instruments having condition numbers on the order of 10^3 or less (well-conditioned) and the original BLP instruments having a condition number on the order of 10^6 or more (ill-conditioned).

Table 1: Median Condition Number of Optimal Weight Matrix

J	T	Cost Inst	BLP Inst	PC Inst
20	3	3.33E+03	4.04E+06	6.23E+02
60	3	1.01E+04	8.33E+07	1.08E+03
100	3	1.37E+04	2.12E+08	1.47E+03
100	10	8.50E+03	1.57E+07	8.87E+02
20	20	1.52E+03	3.50E+05	4.05E+02
60	20	4.36E+03	3.01E+06	4.49E+02
100	20	7.36E+03	1.02E+07	7.55E+02
100	50	7.01E+03	6.95E+06	7.15E+02
20	100	1.35E+03	2.25E+05	3.77E+02
60	100	4.07E+03	2.25E+06	4.17E+02
100	100	6.98E+03	6.21E+06	7.12E+02

For each set of instruments I compute both the 2-step optimal GMM estimator and an EL estimator.[12] I begin by considering the case of exogenous cost shifting (strong) instruments. Table 2 reports the median bias of the price parameter α and the average own price elasticity.[13] When markets are large, the bias is negligible for

[12] I use the 2SLS weight matrix $\left(\frac{1}{J \cdot T} \sum_{j,t} Z_{jt} Z'_{jt}\right)^{-1}$ for the first step of the optimal GMM estimator.

[13] Following the suggestion in Berry and Pakes (2001), focusing on the elasticity rather than in-sample fit or parameter values provides more information about counterfactual predictions of the model. For the EL estimators, the mean own price elasticity is computed over the empirical likelihood measure π rather than the empirical measure $\frac{1}{n}$.(Brown and Newey 1998)

all of the estimators; in small markets $T = 3, 10$ the GEL estimators appear to exhibit bias approximately one half as large as GMM.

Table 2: Bias in α and average own-price elasticity with cost shifting instruments

		Bias in α			Average Own Price Elasticity		
J	T	GMM	CUE	EL	GMM	CUE	EL
20	3	0.0399	0.0271	0.0090	0.0840	-0.0096	-0.0296
60	3	0.0087	-0.0025	-0.0030	0.0852	0.0413	0.0352
100	3	0.0143	0.0095	0.0090	0.0567	0.0391	0.0394
100	10	0.0038	-0.0001	-0.0003	0.0018	-0.0089	-0.0091
20	20	0.0028	-0.0003	-0.0006	0.0155	0.0068	0.0069
60	20	0.0048	0.0034	0.0032	0.0305	0.0219	0.0222
100	20	0.0002	-0.0011	-0.0011	0.0028	-0.0020	-0.0017
100	50	-0.0001	-0.0007	-0.0007	0.0074	0.0059	0.0058
20	100	0.0019	0.0017	0.0016	0.0039	0.0020	0.0018
60	100	0.0003	0.0002	0.0002	0.0108	0.0096	0.0096
100	100	-0.0007	-0.0010	-0.0010	-0.0004	-0.0017	-0.0016

A more interesting case to applied researchers is what happens when only BLP instruments are available (and exogenous cost shifters are not). I report results for the price parameter α, and the average own price elasticity in Tables 3 and 4 respectively. Because most policies such as merger evaluation are more closely linked to the elasticities, I focus primarily on those. Figures 1 shows how the median bias declines in the sample size for the average own-price elasticity. When we examine the bias in the mean elasticity we see that the baseline GMM-BLP estimator shows substantial bias that does not appear to die out very quickly with the sample size. Even at $T = 100$ markets, larger than in many empirical studies, the bias is about 10% of the average own price elasticity. In contrast, the GMM estimator using the principal components instruments shows similar bias in small markets, but the bias declines more quickly in the sample size so that for $T = 100$ markets it is approximately one half as large. The two GEL estimators (CUE and EL) exhibit similar bias to one another and outperform the GMM estimator in nearly all cases. The GEL estimators with $T = 20$ markets exhibit less bias than the GMM estimator does with $T = 100$ markets, and at $T = 100$ markets the GEL estimators reduce the bias by around 90%. Economically, this means that with $J = 100$ products and $T = 20$ markets, the median bias represents almost 20% of the true elasticity with GMM whereas for CUE it represents less than 2%. Additionally, the GEL estimators do not appear to benefit from the use of the principal components instruments. This makes intuitive sense because they do not require directly calculating the optimal weighting matrix (or its inverse), as GEL weights π are chosen so that all moment conditions must be exactly satisfied at the optimum.[14] Table 5, reports the MAD for the three estimators. Despite it's improved bias, there seems to be no efficiency advantage for the GEL estimators when compared to GMM. Similarly, the regularized instruments do not appear to improve the efficiency of the estimator.[15] Rejection rates for the ELR test with 5% significance are reported in Table 6, and indicate that BLP-Only instruments have little power for sample sizes $T \leq 20$, but look reasonable for $T \geq 50$.

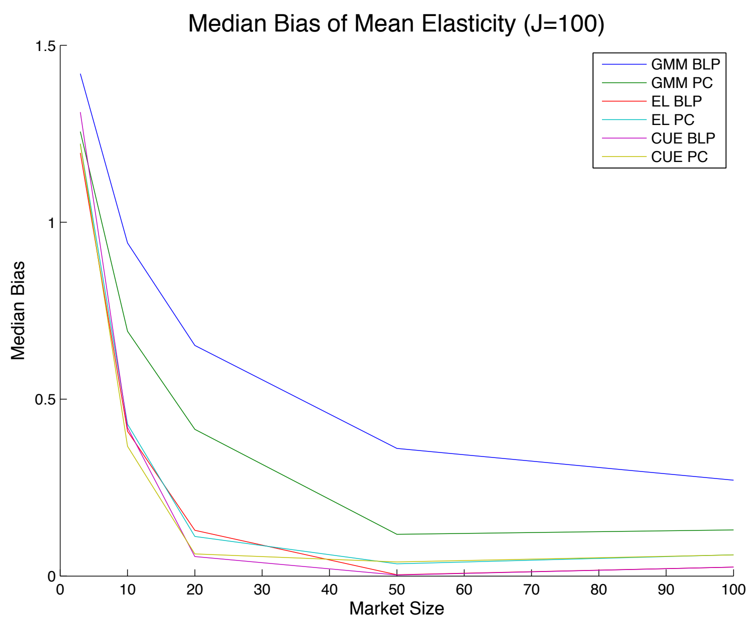

Figure 1: Median Bias of Average Own Price Elasticity

To understand the source of the bias, it is possible to evaluate the analytic expressions for the bias given in Newey and Smith (2004) for the BLP estimator (8)

[14]This is even true for CUE. Though CUE is often presented as a continuously updating GMM estimator where the weight matrix is updated for each evaluation of θ as in Hansen, Heaton, and Yaron (1996), the GEL-MPEC formulation means that this weighting matrix and its inverse are never explicitly required during parameter estimation.

[15]RMSE calculations are also available upon request, but omitted for space considerations. RMSE is more sensitive to outliers than MAD, so trimming needs to be employed in order to make meaningful comparisons.

Table 3: Bias in α with BLP and PC Instruments

		BLP Inst.			PC Inst.		
J	T	GMM	EL	CUE	GMM	EL	CUE
20	3	0.3308	0.2886	0.2663	0.3660	0.3097	0.3641
20	20	0.2487	0.0942	0.0901	0.2034	0.0820	0.0830
20	100	0.1158	0.0240	0.0232	0.0667	0.0280	0.0281
60	3	0.3717	0.3092	0.3188	0.3992	0.3946	0.3990
60	20	0.2570	0.1445	0.1605	0.1597	0.0959	0.0914
60	100	0.0669	0.0104	0.0077	0.0331	0.0137	0.0121
100	3	0.4155	0.3542	0.3880	0.3627	0.3526	0.3523
100	10	0.2788	0.1225	0.1237	0.2123	0.1285	0.1119
100	20	0.1937	0.0398	0.0132	0.1233	0.0315	0.0196
100	50	0.1059	-0.0034	-0.0014	0.0333	-0.0112	-0.0147
100	100	0.0802	0.0081	0.0082	0.0386	0.0186	0.0186

Table 4: Median Bias in Average Own-Price Elasticity

		BLP Inst.			PC Inst.		
J	T	GMM	CUE	EL	GMM	CUE	EL
20	3	1.0133	0.8226	0.8257	1.1334	1.0998	0.9161
60	3	1.2638	1.1083	1.0691	1.3756	1.3811	1.3557
100	3	1.4196	1.3110	1.1958	1.2559	1.2208	1.2219
100	10	0.9414	0.4195	0.4093	0.6918	0.3667	0.4291
20	20	0.7893	0.2892	0.3212	0.6668	0.2369	0.2374
60	20	0.8717	0.5465	0.4856	0.5381	0.3113	0.3156
100	20	0.6516	0.0551	0.1292	0.4147	0.0625	0.1120
100	50	0.3609	0.0031	-0.0036	0.1180	-0.0401	-0.0347
20	100	0.3632	0.0817	0.0841	0.2026	0.0876	0.0878
60	100	0.2276	0.0297	0.0435	0.1156	0.0494	0.0561
100	100	0.2711	0.0254	0.0254	0.1304	0.0597	0.0602

Table 5: Median Absolute Deviation of Average Own-Price Elasticity

		Cost Inst	BLP inst			PC inst		
J	T	GMM	GMM	CUE	EL	GMM	CUE	EL
20	3	0.2550	1.0764	2.1072	2.1109	1.3984	2.1255	2.1637
60	3	0.1485	1.3144	2.1349	2.0812	1.4654	1.9346	2.0793
100	3	0.1468	1.4856	2.1053	2.0290	1.5633	1.9699	1.9768
100	10	0.0766	1.0136	1.2920	1.2812	0.9945	1.2203	1.2187
20	20	0.0863	0.8353	1.0404	1.0165	0.7938	0.9053	0.8947
60	20	0.0584	0.9051	0.9678	0.9545	0.7576	0.9079	0.9071
100	20	0.0483	0.7564	0.9348	0.9099	0.7370	0.8376	0.8362
100	50	0.0335	0.4104	0.4999	0.4876	0.4362	0.4684	0.4543
20	100	0.0403	0.4251	0.4431	0.4423	0.4586	0.4852	0.4865
60	100	0.0218	0.3664	0.4012	0.3860	0.3611	0.3609	0.3603
100	100	0.0184	0.4139	0.3596	0.3604	0.3680	0.3602	0.3600

Table 6: 5% ELR Test Rejection Rates for α (Cost Inst, BLP Inst, PC Inst)

J	T	Truth	Zero
100	3	3,14,3	100,18,6
100	10	3,10,1	100,43,16
100	20	3,8,2	100,68,58
100	50	4,8,1	100,94,89
100	100	6,8,4	100,100,100

under the three different sets of instruments. These results are reported in Table 7, and correspond to the observed bias in the Monte Carlo study. As one might expect, all three bias components are substantially smaller when the exogenous cost shifting instruments are available. All of the bias components also appears to be declining in both J and T.[16] In all cases, the largest bias term is B_G which comes from the correlation of ξ_{jt} with p_{jt}. The use of the PC instruments increases the magnitude of the residual bias B_I and decreases the magnitude of the endogeneity bias B_G. This makes sense because the principal components are chosen so as to approximate but not fully span the space of instruments, while the orthogonal basis reduces some of the correlation in B_G, this may partially explain why GMM benefits from the regularized instruments and the GEL estimators do not. Also, the bias in B_Ω is generally quite small which suggests similar performance for CUE and EL. However, B_Ω does not account for possible ill-conditioning of the optimal weighting matrix.

Table 7: Analytic Bias Expressions

		Cost Inst			BLP Inst			PC Inst		
J	T	B_I	B_G	B_Ω	B_I	B_G	B_Ω	B_I	B_G	B_Ω
20	3	0.0130	0.0342	0.0034	0.1180	0.3443	0.0519	0.1881	0.2606	0.0180
60	3	0.0044	0.0123	0.0004	0.0917	0.3817	0.0273	0.1578	0.3164	0.0057
100	3	0.0026	0.0079	0.0002	0.0996	0.3439	0.0145	0.1663	0.2540	0.0090
100	10	0.0008	0.0024	0.0000	0.0738	0.2800	0.0049	0.1240	0.2574	0.0067
20	20	0.0018	0.0056	0.0001	0.0785	0.2452	0.0192	0.1244	0.2463	0.0135
60	20	0.0007	0.0020	0.0000	0.0572	0.2101	0.0037	0.0927	0.2008	0.0024
100	20	0.0004	0.0012	0.0000	0.0543	0.2150	0.0027	0.0891	0.1775	0.0025
100	50	0.0002	0.0005	0.0000	0.0309	0.1198	0.0020	0.0478	0.0896	0.0003
20	100	0.0004	0.0011	0.0000	0.0235	0.0749	0.0052	0.0464	0.0727	0.0025
60	100	0.0001	0.0004	0.0000	0.0224	0.0786	0.0011	0.0317	0.0599	0.0004
100	100	0.0001	0.0002	0.0000	0.0195	0.0742	0.0009	0.0298	0.0582	0.0001

As a final example, it is possible to use the analytical expressions for bias given in Newey and Smith (2004) to remove all of the asymptotic bias from the EL-BLP estimator. A well-known property of empirical likelihood estimators is that once they are bias-corrected they are higher order efficient. Moreover, Brown and Newey (1998) showed that the higher order efficiency property is inherited by functions of the form $E[a(z,\theta)] = \sum_{i=1}^{N} \hat{\pi}_i^{GEL} a(z,\theta)$, which includes the BLP average own price

[16] Armstrong (2013)'s results suggest that instruments become weaker as J increases.

elasticity. The bias and MAD of the bias-corrected EL estimator is reported in Tables 8 and (9). Here the results are quite striking. The analytic bias correction is able to remove nearly all of the bias of the EL-BLP estimator, so that even with a small number of markets, and without exogenous cost shifters, bias corrected EL is able to obtain accurate estimates of the price elasticity. Also, though the behavior of the EL estimator is largely unaffected by the principal components regularization, the bias-corrected EL estimator performs much better under the regularized PC instruments. Part of this may come from the fact that while the GEL estimator does not require inverting the variance matrix, the analytic bias correction does. Again, because the regularized PC instruments result in a better conditioned inverse, this may yield a more accurate expression for the bias in finite-samples.

Table 8: Finite Sample Bias in Bias-Corrected GEL Average Elasticity

		BLP inst		PC inst	
J	T	CUE	EL	CUE	EL
20	3	0.3178	0.5605	-0.0100	0.1213
60	3	0.8487	0.8723	0.7730	0.9014
100	3	1.1539	1.0983	0.6837	0.6873
100	10	0.1311	0.0973	-0.1117	0.0208
20	20	-0.0671	0.0858	-0.2812	-0.1798
60	20	0.2457	0.3061	-0.0825	-0.1024

Table 9: Relative Performance of Bias-Corrected EL

		Price Parameter α				Mean Own-Price Elasticity			
		Relative Bias		Relative MAD		Relative Bias		Relative MAD	
J	T	BLP	PC	BLP	PC	BLP	PC	BLP	PC
20	3	0.67	0.22	0.93	0.93	0.68	0.13	0.93	0.94
60	3	0.84	0.65	0.97	0.89	0.82	0.66	0.96	0.88
100	3	0.92	0.58	1.03	0.95	0.92	0.56	1.03	0.95
100	10	0.24	0.09	0.98	1.01	0.24	0.05	0.98	1.01
20	20	0.24	-0.64	1.12	0.96	0.27	-0.76	1.12	0.95
60	20	0.60	-0.35	0.98	1.11	0.63	-0.32	0.95	1.09

4.2 Replication of Nevo(2000)

The fake-data example of Nevo (2000) has received much attention in the new literature examining the properties of the BLP estimator, including but not limited to recent work by: Knittel and Metaxoglou (2013), Dubé, Fox, and Su (2012), and Lee and Seo (2013). The dataset examines 94 markets and 24 brands, with 20 simulation draws per market, designed to mimic the study of ready-to-eat cereal in Nevo (2001). In the literature that has subsequently examined the Nevo (2000) example, the 20 simulation draws have been fixed and treated as if the true population is a 20-point discrete distribution (an assumption I maintain) for the 9 nonlinear parameters. Also, previous analysis of the Nevo (2000) example such as Knittel and Metaxoglou (2013) or Dubé, Fox, and Su (2012) have employed one-step GMM using the TSLS weighting matrix $W = (Z'Z)^{-1}$. This matrix is ill-conditioned with a condition number 1.9×10^8, which makes two-step optimal GMM estimates (and asymptotic standard errors) difficult to obtain.

I report three sets of the results in Table 4.2: the results of the original nested-fixed point estimator of Nevo (2000), the results of the (TSLS) GMM-MPEC estimator of Dubé, Fox, and Su (2012), and the results of the EL-MPEC estimator.[17] The GMM-MPEC and EL-MPEC approaches produce very similar point-estimates. I also report the EL-MPEC asymptotic standard errors in brackets[18], which produce reasonable confidence intervals in most of the (non-price) parameters.

As a second exercise, I construct a confidence interval for the price parameter α, by inverting the ELR test statistic. This is similar to the idea of Anderson and Rubin (1949), which provided a weak-instrument robust test of the hypothesis that $\hat{\alpha} = \alpha_0$, and was formalized in relation to generalized empirical likelihood in in the GELR statistic of Guggenberger and Smith (2005). More detail on how the ELR statistic is computed in this setting is provided in the Appendix. The ELR robust confidence interval consists of all of the values α for which the ELR test does not reject the null hypothesis $\{\alpha : ELR(\alpha) - \max_{\alpha'} ELR(\alpha') \leq \chi^2_1\}$.

Figure 4.2 plots the maximized profile ELR statistic: $ELR(\alpha) = \max_\theta EL(\alpha, \theta) - N \log N$. As the figure indicates, despite relatively small asymptotic standard errors,

[17] The NFP estimator can be adjusted as suggested in Dubé, Fox, and Su (2012) by tightening the tolerance of the inner loop (contraction mapping) at which point it produces estimates nearly identical to the GMM-MPEC case.

[18] The GMM-MPEC standard errors (and point estimates) are nearly identical.

Table 10: Replication of Nevo Example

	Nevo	GMM-MPEC	EL-MPEC	[S.E.]
Price	-28.189	-62.726	-61.433	[14.999]
σ_p	0.330	0.558	0.527	[0.116]
σ_{const}	2.453	3.313	3.143	[1.270]
σ_{sugar}	0.016	-0.006	0.000	[0.013]
σ_{mushy}	0.244	0.093	0.085	[0.181]
$\pi_{p,inc}$	15.894	588.206	564.262	[275.220]
$\pi_{p,inc2}$	-1.200	-30.185	-28.930	[14.349]
$\pi_{p,kid}$	2.634	11.058	11.700	[4.115]
$\pi_{c,inc}$	5.482	2.291	2.246	[1.197]
$\pi_{c,age}$	0.204	1.284	1.379	[0.644]
GMM	29.3611	4.564		
ELR			5.275	

the ELR test fails to reject at 5% for a wide range of parameter values $[-161, -40]$. The ELR test statistic also fails to reject $\alpha_0 = 0$ at 2.5% and nearly anything at 1%. This provides some indication that the fake-data example of Nevo (2000) suffers from weak identification, and suggests an alternate hypothesis for the negative results found by Knittel and Metaxoglou (2013), namely that the failure was not exclusively of the BLP method or the optimization *per se* but rather of the instruments, which results in a mostly flat objective function.[19]

5 Conclusion

Empirical likelihood and GEL estimators represent a major development in the econometric literature over the past two decades. These estimators provide advantages over GMM estimators, especially as it pertains to finite-sample bias, and higher-order efficiency. Additionally, EL estimators are known to be partially robust to the problem of many weak instruments, which often arises when constructing moment conditions from conditional moment restrictions. Despite having desirable properties, EL estimators are rarely found in applied work. The principal drawback is that computing

[19]Using higher-order functions of instruments to form moments, or using smoothed conditional moments (Kitamura, Tripathi, and Ahn 2004) does not seem to aid identification in this particular example.

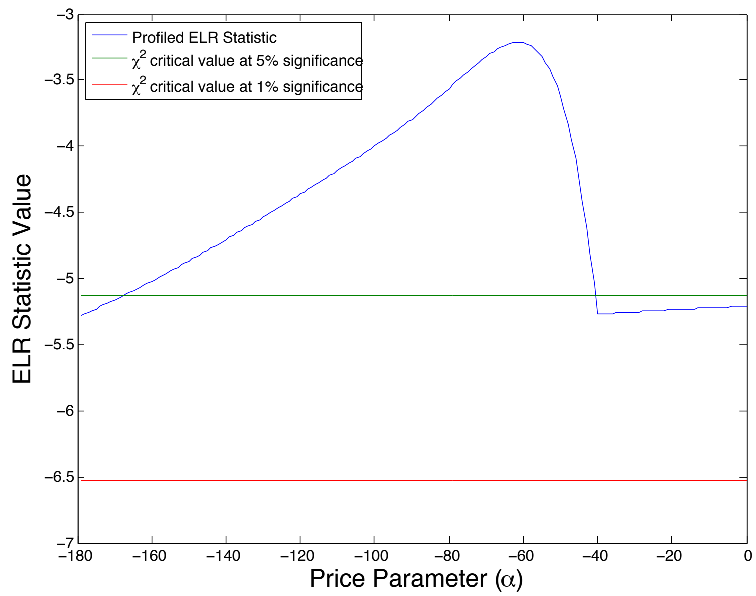

Figure 2: Profile Empirical Likelihood of Price Coefficient for Nevo Dataset

the estimator can be very difficult, and that past studies have found only modest improvements when using EL instead of GMM.

The GEL-MPEC formulation provided in this paper provides a computationally attractive method to obtain GEL estimators, including the computationally challenging simulated nonlinear problem of Berry, Levinsohn, and Pakes (1995). When both exogenous cost-shifting instruments and characteristics of other products (BLP instruments) are available, GMM and GEL estimators both perform very well, even in small samples. However, in many empirical settings, economists do not have access to high quality data on costs or cost shocks. In those cases, the GMM estimator tends to exhibit substantial finite-sample bias, so that even as the number of markets becomes large, the behavior of the estimator is not well described by its asymptotic distribution. The GEL estimators generally perform better, so that the bias is modest, even in small numbers of markets. Moreover, it is possible to use the analytic bias correction of Newey and Smith (2004) and remove all asymptotic bias from the EL estimator. After bias correction, the EL estimator demonstrates negligible bias even for small samples. This could prove extremely valuable for applied researchers, for example, in the case of $J = 100$ products, and $T = 10$ markets, the median bias in the average own price elasticity of the GMM-BLP estimator is approximately 0.94, and for bias-corrected EL with regularized instruments it is only 0.02. Moreover, the MPEC method makes the GEL-BLP estimator not appreciably more difficult to compute than the GMM estimator.

This suggests some strategies for empirical researchers estimating BLP-type problems. The first is that the EL-MPEC method provides substantial econometric advantages over GMM. The second, is that the choice of the weighting matrix (and it's condition number) play an important role in finite-sample performance of GMM-BLP estimators, and for bias correction of GEL estimators. The third recommendation is that profiling the empirical likelihood function in the spirit of Anderson and Rubin (1949) or Guggenberger and Smith (2005), can provide diagnostic information about the degree of identification provided by the instruments, and is especially important for the BLP price parameter.

References

Abowd, J. M., and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2), 411–445.

Altonji, J. G., and L. M. Segal (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business and Economic Statistics*, 14(3), 353–366.

Anderson, T., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20(1), 46–63.

Armstrong, T. (2013): "Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply," Working Paper.

Bekker, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62(3), pp. 657–681.

Berry, S., and P. Haile (2012a): "Identification in Differentiated Products Markets Using Market Level Data," Working Paper.

——— (2012b): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," Working Paper.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

——— (2004): "Automobile Prices in Market Equilibrium," *Journal of Political Economy*, 112(1), 68–105.

Berry, S., O. B. Linton, and A. Pakes (2004): "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems," *Review of Economic Studies*, 71, 613–654.

Berry, S., and A. Pakes (2001): "Comments on Alternative models of demand for automobiles by Charlotte Wojcik," *Economics Letters*, 74(1), 43–51.

Bickel, P. J., and E. Levina (2008): "Regularized Estimation of Large Covariance Matrices," *The Annals of Statistics*, 36(1), 199–227.

Brown, B. W., and W. K. Newey (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica*, 66(2), 453–464.

Caner, M. (2008): "Nearly-singular design in GMM and generalized empirical likelihood estimators," *Journal of Econometrics*, 144(2), 511 – 523.

Caner, M., and N. Yildiz (2012): "CUE with many weak instruments and nearly singular design," *Journal of Econometrics*, 170(2), 422 – 441, Thirtieth Anniversary of Generalized Method of Moments.

Carrasco, M. (2012): "A Regularization Approach to the Many Instruments Problem," *Journal of Econometrics*, 170(2), 383 – 398, Thirtieth Anniversary of Generalized Method of Moments.

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18(1), 5–46.

——— (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34(3), 305–334.

Domínguez, M. A., and I. N. Lobato (2004): "Consistent Estimation of Models Defined by Conditional Moment Restrictions," *Econometrica*, 72(5), 1601–1615.

Donald, S. G., and W. K. Newey (2001): “Choosing the Number of Instruments,” *Econometrica*, 69(5), pp. 1161–1191.

Dubé, J.-P. H., J. T. Fox, and C.-L. Su (2012): “Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation,” *Econometrica*, 80(5), 2231–2267.

Fox, J., and A. Gandhi (2012): “Nonparametric Identification and Estimation of Random Coefficients in Nonlinear Economic Models,” Working Paper.

Freyberger, J. (2012): “Asymptotic theory for differentiated products demand models with many markets,” CeMMAP working papers CWP19/12, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Gandhi, A., S. Berry, and P. Haile (Forthcoming): “Connected Substitutes and Invertibility of Demand,” *Econometrica*.

Grant, N. (2012): “Overcoming The Many Weak Instrument Problem Using Normalized Principal Components,” Working Paper.

Guggenberger, P., and R. J. Smith (2005): “Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak, and Strong Identification,” *Econometric Theory*, null, 667–709.

Hansen, L. P. (1982): “Large Sample Propertes of Generalized Method of Moments Estimators,” *Econometrica*, 50, 1029–1054.

Hansen, L. P., J. Heaton, and A. Yaron (1996): “Finite-sample properties of some alternative GMM estimators,” *Journal of Business and Economic Statistics*, 14(3), 262–280.

Hansen, L. P., and K. J. Singleton (1983): “Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns,” *The Journal of Political Economy*, 91(2), 249–265.

Heiss, F., and V. Winschel (2008): “Likelihood approximation by numerical integration on sparse grids,” *Journal of Econometrics*, 144(1), 62 – 80.

Imbens, G. (2002): “Generalized Method of Moments and Empirical Likelihood,” *Journal of Business and Economic Statistics*, 20(4), 493–506.

Imbens, G., R. Spady, and P. Johnson (1998): “Information Theoretic Approaches to Inference in Moment Condition Models,” *Econometrica*, 66(2), 333–357.

Judd, K. L., and B. Skrainka (2011): “High performance quadrature rules: how numerical integration affects a popular model of product differentiation,” CeMMAP working papers CWP03/11, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Kitamura, Y. (2006): “Empirical Likelihood Methods in Econometrics: Theory and Practice,” Cowles Foundation Discussion Paper 1569.

Kitamura, Y., and M. Stutzer (1997): “An Information-Theoretic Alternative to Generalized Method of Moments Estimation,” *Econometrica*, 65(4), 861–874.

Kitamura, Y., G. Tripathi, and H. Ahn (2004): “Empirical Likelihood-Based Inference in Conditional Moment Restriction Models,” *Econometrica*, 72(6), 1667–1714.

Knight, K., and W. Fu (2000): “Asymptotics for Lasso-Type Estimators,” *The Annals of Statistics*, 28(5), 1356–1378.

Knittel, C. R., and K. Metaxoglou (2013): “Estimation of Random Coefficient Demand Models: Two Empiricists’ Perspective,” *Review of Economics and Statistics*, Working Paper.

LEE, J., AND K. SEO (2013): "A computationally efficient estimator for aggregate random coefficients logit demand models," Working Paper.

MORROW, W. R., AND S. J. SKERLOS (2011): "Fixed-Point Approaches to Computing Bertrand-Nash Equilibrium Prices Under Mixed-Logit Demand," *Operations Research*, 59(2), 328–345.

NEVO, A. (2000): "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand (including Appendix)," *Journal of Economics and Management Strategy*, 9(4), 513–548.

——— (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342.

NEWEY, W. K., AND R. J. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72(1), 219–255.

NEWEY, W. K., AND F. WINDMEIJER (2009): "Generalized Method of Moments With Many Weak Moment Conditions," *Econometrica*, 77(3), 687–719.

OWEN, A. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional,," *Biometrika*, 75(2), 237–249.

——— (1990): "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, 18(1), 1725–1747.

——— (2001): *Empirical Likelihood*. Chapman and Hall, 1st edn.

QIN, J., AND J. LAWLESS (1994): "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22(1), 300–325.

QUARTERONI, A., F. SALERI, AND P. GERVASIO (2010): *Scientific Computing with MATLAB and Octave*, Texts in computational science and engineering. Springer.

REYNAERT, M., AND F. VERBOVEN (2012): "Improving the Performance of Random Coefficients Demand Models: the Role of Optimal Instruments," Working Paper.

REYNAERTS, J., R. VARADHAN, AND J. C. NASH (2012): "Enhancing the Convergence Properties of the BLP (1995) Contraction Mapping," Vives discussion paper series 35, Katholieke Universiteit Leuven, Faculteit Economie en Bedrijfswetenschappen, Vives, http://ideas.repec.org/p/ete/vivwps/35.html.

SKRAINKA, B. S. (2012): "A Large Scale Study of the Small Sample Performance of Random Coefficient Models of Demand," Working Paper.

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68(5), 1055–1096.

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20(4), 518–529.

SU, C.-L., AND K. L. JUDD (2012): "Constrained optimization approaches to estimation of structural models," *Econometrica*, 80(5), 2213–2230.

Appendix

A.1 EL-Dual

For completeness, consider how the EL estimator might be obtained using the dual:

$$\hat{\theta}_{EL} = \arg\max_{\theta\in\Theta}\min_{\gamma} -\sum_{i=1}^{n}\log(1+\gamma' g(z_i,\theta))$$

In the case of demand estimation $g(z_i,\theta) = \xi_i z_{i*}$ is $1\times Q$. There are two options, the first would be to solve the EL optimization problem subject to the marketshare constraint or the "constrained-dual":

$$\begin{aligned}
\hat{\theta}_{EL} &= \arg\max_{\theta,\xi,\alpha,\beta}\min_{\gamma} -\sum_{i=1}^{n}\log(1+\xi_{jt}\gamma' z_{jt}) \quad \text{s.t.}\\
s_{jt}(\theta) &= \int \frac{\exp[x_{jt}\beta_i + \xi_{jt} - \alpha_i p_{jt}]}{1+\sum_k \exp[x_{kt}\beta_i + \xi_{kt} - \alpha_i p_{kt}]} f(\beta_i|\theta)\\
\log(S_{jt}) &= \log(s_{jt}(\delta,\theta_2))
\end{aligned}$$

The "constrained-dual" approach has some drawbacks. The objective function is a globally convex minimax (saddle-point) optimization problem, but it is now subjected to N nonlinear equality constraints for the marketshares. It's not entirely clear how to go about solving this sort of constrained minimax problem. There may be methods to solve these problems, but they would be"cutting-edge" in the optimization literature, and are generally unavailable in standard software packages.

The second would be to solve a doubly nested problem, which we obtain by combining the fixed point approach in Berry, Levinsohn, and Pakes (1995) and the saddle point approach in the empirical likelihood literature and put them together:

$$\begin{aligned}
\hat{\theta}_{EL} &= \arg\max_{\theta}\min_{\gamma} -\sum_{i=1}^{n}\log(1+\xi_i(\theta)\gamma' z_i)\\
\xi_{jt}(\theta) &= \delta_j(\theta_2) - x_{jt}\beta_i - \alpha_i p_{jt}\\
\log(S_{jt}) &= \log(s_{jt}(\delta,\theta_2))
\end{aligned}$$

The doubly-nested fixed point approach involves the following for each guess of θ. First solve the share equations via the contraction mapping for the δ vector. Then obtain the ξ as the residuals of regression of δ on observables. Then, plug those ξ values into the saddle-point, and solve that via Newton-steps for γ. Using (γ,ξ) we can then evaluate the objective function for θ. Even though the objective function is convex in ξ, it is not necessarily convex in θ, and differentiating through the maximum of a fixed point points is numerically unstable, so that a evaluation only method such as Nelder-Mead would be used for the outer loop. For large problems, this approach can be computationally infeasible.

A.2 EL-MPEC Inference

Qin and Lawless (1994) show that the asymptotic distribution of the empirical likelihood estimator attains the semiparametric efficiency bound of Chamberlain (1982) and has the well known form:

$$\begin{aligned}
D &= E_{\pi}[\nabla_{\theta} g(z,\theta_0)]\\
S &= E_{\pi}[g(z,\theta_0)g(z,\theta_0)']\\
\sqrt{n}(\hat{\theta}_{EL}-\theta_0) &\to^{d} N(0,(D'SD)^{-1})
\end{aligned}$$

The MPEC formulation of the EL estimator does not change the asymptotic behavior. It should be noted that the expectations above are not evaluated with $\frac{1}{n}$ weights for each z_i but rather the π_i weights obtained via the empirical likelihood estimator. The only difference is that D requires computing the Jacobian of the moments not in the unrestricted way that MPEC does, but along the manifold of the constraint $\Psi(P, \theta)$ (in the case of BLP along the constraint $S_{jt} = s_{jt}(\delta, \theta)$.

The MPEC formulation of EL estimators also allows for testing within the optimization problem. We are often concerned with testing an s dimensional restriction on our parameters such that $R(\theta_0) = 0$. We can simply incorporate this as an additional equilibrium constraint and re-estimate:

$$\begin{aligned} \hat{\theta}_0 &= \arg \min_{\theta, \pi_1, \ldots, \pi_n} - \sum_{i=1}^{n} \log \pi_i \quad \text{s .t.} \quad R(\theta) = 0 \\ & \sum_{i=1}^{n} \pi_i g(z_i, \theta) = 0 \quad \sum_{i=1}^{n} \pi_i = 1 \quad - \pi_i \leq 0 \end{aligned} \tag{16}$$

Following Kitamura (2006) we can construct the empirical likelihood ratio (ELR) statistic, where $s = \dim(\Theta) - \dim(R)$:

$$r = -2 \left(l(\hat{\theta}_0) - l(\hat{\theta}_{EL}) \right) \sim \chi^2_s \tag{17}$$

In many economic models the quantity of interest is often some other prediction of the model (such as an average price elasticity in the BLP case), that is not merely a function of the parameters, but also of the data. For example, many statistics of interest are of the form $E[a(z, \theta)] = \int a(z, \theta) d\mu$. It is well known in the empirical likelihood literature (Brown and Newey 1998), that we can obtain more efficient estimates using the empirical likelihood mean, than we could under the $\frac{1}{n}$ sample mean:

$$E[a(z, \theta)] = \sum_i \hat{\pi}_{i,EL} a(z_i, \hat{\theta}_{EL}) \tag{18}$$

We might also consider testing the hypothesis that $E[a(z, \theta)]] = b$. This can be viewed as an overidentifying restriction. We augment the moment conditions and obtain the constrained likelihood value $l(\theta_0)$ (Kitamura 2006). The empirical log likelihood ratio function tests overidentifying restrictions and and can be constructed following (Owen 2001):

$$\begin{aligned} elr(\theta) &= -2[l(\theta) + n \log n] \\ r &= elr(\theta_0) - elr(\hat{\theta}_{EL}) \end{aligned}$$

Under the null this should also have χ^2_s distribution where s is the dimension of the restriction. This is a special case (for empirical likelihood) of the robust confidence interval construction procedure of Guggenberger and Smith (2005) that involves inverting the test statistic.

A.3 Ill-Conditioned Covariance Matrices

We begin by considering the BLP moment conditions $E[g(z_{jt}, \theta)] = E[(\delta - x_{jt}\beta)Z_{jt}] = 0 \in \mathbb{R}^m$. This implies that the covariance of the moment conditions is $\Omega_0 = E[g(z_{jt}, \theta_0)' g(z_{jt}, \theta_0)]$ with sample analogue $\widehat{\Omega} = \frac{1}{n} \sum_{j,t} [g(z_{jt}, \theta)' g(z_{jt}, \theta)]$. Under IID sampling and four finite moments, we know that $\widehat{\Omega} \to^p \Omega_0$ at rate $\frac{1}{\sqrt{n}}$ so long as $\frac{m}{n} \to 0$.

Much of the literature on GMM (Hansen 1982) or GEL (Newey and Smith 2004) assumes that Ω is full rank and hence invertible. This is reasonable for asymptotic results because over the $m \times m$ field of real matrices, the Lebesgue measure of the set of non-invertible (zero determinant matrices) is zero. For real square symmetric matrices this implies that that $det(\Omega) = \prod_{i=1}^{m} \lambda_i \neq 0$ or that all eigenvalues are nonzero. Like all covariance matrices, Ω_0 is positive semi-definite, which implies the even stronger condition that $\lambda_i > 0 \; \forall i$.

However, in finite sample $[\widehat{\Omega}]^{-1}$ may be a poor approximation of Ω_0^{-1} if the matrix is ill-conditioned. The ill-conditioning results when the minimum eigenvalue $\lambda^{min}(\Omega_0)$ is very small or vanishing. This implies that small changes in $\hat{\Omega}$ result in large changes in $\hat{\Omega}^{-1}$.

Consider $\hat{\Omega} = \Omega_0 + o_p(\sqrt{n})$. There is a well-known result in numerical linear algebra which considers a perturbation to the system $Ax = b$ so that $\hat{x} = (A + \Delta A)^{-1}(b + \Delta b)$ and considers the effect on the value of the solution x on the logarithmic scale (Quarteroni, Saleri, and Gervasio 2010):

$$\frac{\|x - \hat{x}\|}{\|x\|} = \frac{K(A)}{1 - \lambda^{max}(\Delta A)/\lambda^{min}(A)} \left(\frac{\lambda^{max}(\Delta A)}{\lambda^{max}(A)} + \frac{\|\Delta b\|}{\|b\|} \right) \leq K(A)$$

For a normal matrix the condition number is the ratio of the extreme eigenvalues $K(A) = \frac{\lambda^{max}(A)}{\lambda^{min}(A)}$. Even though $(\hat{\Omega} - \Omega) \rightarrow^p 0$ as the sample covariance matrix converges to the population covariance matrix, in finite sample the inverse may be badly behaved as $K(\Omega)$ is routinely on the order of 10^6 or more.

One special matrix known to be ill-conditioned is the Vandermonde Matrix. That matrix has the form:

$$V = \begin{bmatrix} 1 & z_1 & z_1^2 & \cdots & z_1^k \\ 1 & z_2 & z_2^2 & \cdots & z_2^k \\ \vdots & \vdots & \vdots & & \\ 1 & z_2 & z_n^2 & \cdots & z_n^k \end{bmatrix}$$

This is a similar to instrument matrices commonly employed in econometrics when considering conditional moment restrictions of the form $E[\xi | z_i] = 0$ or to approximate optimal instruments in a nonlinear setting $E[\xi' f(z_i)] = 0$. This approach is prevalent throughout the BLP literature, for example Dubé, Fox, and Su (2012) construct 42 instruments from higher-order interactions of regressors x_{jt} and instruments z_{jt}. Interactions of instruments are also employed in the approximation of the optimal instruments in Berry, Levinsohn, and Pakes (1995).

There has been a recent literature on regularization approaches to estimating covariance matrices and the ill-conditioning problem including Knight and Fu (2000), Bickel and Levina (2008), Caner (2008), Caner and Yildiz (2012). One of the simplest suggestions (Carrasco 2012) is to consider the first k principal components of the instruments Z. The Singular Value Decomposition (SVD) of $Z'Z$ allows for representation $Z = UDV'$ where D is a diagonal matrix corresponding to the square roots of the eigenvalues of Z, and the columns of U correspond to eigenvectors of ZZ' and the columns of V correspond to eigenvectors of $Z'Z$. The idea is to choose an approximation on the orthonormal basis of eigenvectors that nearly spans the vector space $Z'Z$, which by sorting the eigenvalues (and the corresponding eigenvectors) by descending magnitude and choosing only eigenvalues above some threshold. Ideally, this threshold is chosen to minimize the MSE criteria in the many instrument selection problem presented in Donald and Newey (2001) or Newey and Windmeijer (2009). In the Monte Carlo data, the MSE minimizing number of instruments is usually 4, which I fix across Monte Carlo trials.

In an alternative specification, rather than selecting components in descending order of eigenvalues, I follow Grant (2012) and select the normalized principal components corresponding to the largest coefficients in the regression on price (i.e.: the strongest instruments) $p_{jt} = \gamma z_{jt}^{PCA} + \eta_{jt}$. There, I select the three principal components corresponding to the largest values of γ^2 and label them "PC2 Instruments".[20] Because the performance is generally worse than the spectral cutoff

[20]The rationale given in Grant (2012) is that the principal components are selected to explain

strategy of Carrasco (2012) described above, I do not present those results.

A.4 Analytic Bias Expressions

Here I write the expressions for analytic bias in GMM as found in Newey and Smith (2004) and plug in for the sample analogues of the BLP problem:

$$\begin{aligned}
g_i &= g_i(\beta_0) = z_{jt}\xi_{jt} \\
G_i &= \frac{\partial \xi_{jt}}{\partial \theta} = z_{jt} \times [X_{jt}\ p_{jt}\ \frac{\partial \xi_{jt}}{\partial \sigma}] \\
\Omega &= E[g_i g_i'] = \left(\frac{1}{J \cdot T} \sum_{j=1}^{J} \sum_{t=1}^{T} \xi_{jt}^2 z_{jt} z_{jt}' \right) \\
G &= E[G_i] = \frac{1}{J \cdot T} \sum_{j=1}^{J} \sum_{t=1}^{T} z_{jt} \times [X_{jt}\ p_{jt}\ \frac{\partial \xi_{jt}}{\partial \sigma}] \\
\Sigma &= (G'\Omega^{-1}G)^{-1}, \quad H = \Sigma G' \Omega^{-1}, \quad P = \Omega^{-1} - \Omega^{-1} G \Sigma G' \Omega^{-1}
\end{aligned}$$

Which gives the analytic bias expressions:

$$\begin{aligned}
B_G &= -\Sigma \left(\frac{1}{J \cdot T} \sum_{j=1}^{J} \sum_{t=1}^{T} [X_{jt}\ p_{jt}\ \frac{\partial \xi_{jt}}{\partial \sigma}]' z_{jt} P z_{jt} \xi_{jt} \right) \\
B_\Omega &= H \left(\frac{1}{J \cdot T} \sum_{j=1}^{J} \sum_{t=1}^{T} \xi_{jt}^3 z_{jt} z_{jt}' P z_{jt} \right) \\
B_I &= H \left(\frac{1}{J \cdot T} \sum_{j=1}^{J} \sum_{t=1}^{T} z_{jt} [X_{jt}\ p_{jt}\ \frac{\partial \xi_{jt}}{\partial \sigma}] H z_{jt} \xi_{jt} \right)
\end{aligned}$$

The Jacobian term $\frac{\partial \xi_{jt}}{\partial \theta}$ depends on the nonlinear parameters σ, where $\frac{\partial \xi_{jt}}{\partial \sigma} = \underbrace{\frac{\partial \xi_{jt}}{\partial \delta}}_{\mathbf{I}_{jt}} \cdot \frac{\partial \delta_{jt}}{\partial \sigma} = \frac{\partial \delta_{jt}}{\partial \sigma}$ We can compute this for a single market t using the implicit function theorem (following the Appendix of Nevo (2000)):

$$\frac{\partial \delta_{jt}}{\partial \sigma} = - \begin{pmatrix} \frac{\partial s_{1t}}{\partial \delta_{1t}} & \cdots & \frac{\partial s_{1t}}{\partial \delta_{Jt}} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_{Jt}}{\partial \delta_{1t}} & \cdots & \frac{\partial s_{Jt}}{\partial \delta_{Jt}} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial s_{1t}}{\partial \sigma_1} & \cdots & \frac{\partial s_{1t}}{\partial \sigma_K} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_{Jt}}{\partial \sigma_1} & \cdots & \frac{\partial s_{Jt}}{\partial \sigma_K} \end{pmatrix}$$

---

the variation in the instrument matrix Z but not necessarily the endogenous regressor p_{jt}. It might be that the component corresponding to the dominant eigenvalue of Z is uncorrelated with p_{jt} or the component corresponding to the smallest eigenvalue is the only component correlated with p_{jt}. This is resolved by selecting components based on the fraction of variance in p_{jt} that they explain, which is a monotone function of the regression coefficient γ^2 once the Z has been normalized so that all columns have equal variance. In practice, I have tried it both ways, and the B_I term is generally substantially larger under the PC2 instruments than under the PC instruments, though the B_G term is slightly smaller on average.

The components of the two matrices are just the partial derivatives with respect to (δ, σ) which are computed in the MPEC estimator as the Jacobian of the share constraint.

The inclusion of the $\frac{\partial \xi_{jt}}{\partial \theta}$ term has a negligible effect on the bias (and generally the bias of σ is small). The primary source of the bias generally arises from the correlation of p_{jt} with z_{jt} and ξ_{jt} in B_G.