

# Estimation of Random Coefficient Demand Models: Challenges, Difficulties and Warnings

Christopher R. Knittel and Konstantinos Metaxoglou\*

Work in progress. Please do not cite without permission.

## Abstract

Empirical exercises in economics frequently involve estimation of highly nonlinear models. The criterion function may not be globally concave or convex and exhibit many local extrema. Choosing among these local extrema is rather challenging. First, the researcher may not uncover the global extremum. Second, the global extremum in small samples may not correspond to the asymptotic global extremum. Finally, the global extremum may not correspond to a consistent root. In this paper, we analyze the sensitivity of parameter estimates, and most importantly of economic variables of interest, to both starting values and the type of non-linear optimization algorithm employed. We focus on a class of demand models for differentiated products that have been used extensively in industrial organization. We find that convergence may occur at a number of local extrema, saddles and in regions of the objective function where the first-order conditions are not satisfied. In the interest of a routine market power analysis exercise, we find own- and cross-price elasticities that differ by a factor of over 100 depending on the set of candidate parameter estimates. In an attempt to evaluate the welfare effects of a change in an industry's structure, we undertake a hypothetical merger exercise. Our calculations indicate consumer welfare effects varying between negative one to negative forty million dollars depending on the set of parameter estimates used.

---

\**Knittel*: Department of Economics, University of California, Davis, CA and NBER. *Email*: crknittel@ucdavis.edu. *Metaxoglou*: Bates White LLC. *Email*: konstantinos.metaxoglou@bateswhite.com.

# 1 Introduction

Empirical research in economics often requires estimating highly nonlinear models, where the objective function may not be globally concave or convex. Obtaining parameter estimates in these cases requires a set of starting values, a nonlinear search algorithm and a set of convergence stopping rules. For a common class of demand models used in industrial organization, we find that convergence can occur at a number of local extrema, saddles and in regions of the objective function where the first-order conditions are not satisfied. Furthermore, parameter estimates and measures relevant for economic analysis lack robustness to both the starting values and the optimization algorithm used by the econometrician.

These results underline the importance of verifying both the first- and second- order conditions for a local optimum, as well as highlight that researchers will often have to choose among local extrema. Choosing among local extrema is not trivial for a variety of reasons. First, if the true parameter vector achieves the global extremum, then proofs of consistency require the researcher to find the global extremum associated with the sample. Unfortunately, there is no guarantee that one of the local extrema is the global one. Second, there are cases where the global extremum may *not* be consistent with little or no statistical guidance as to which of the local extrema is the consistent one (see Amemiya, 1985). Many consistency proofs are local in nature; they may say little about the behavior of the objective function outside of this neighborhood. Finally, even in cases where the global extremum is the consistent root and we are convinced we have uncovered it for our sample, this root may not remain the global extremum as the sample changes or grows. The possibility of such a “horse race” implies that reported standard errors may misrepresent the true uncertainty regarding parameter estimates, since typical published standard errors ignore issues of multiple extrema and the possibility of a horse race. Our results suggest that the *process* a researcher undertakes to estimate the model can have large effects on the economic consequences of the estimates.

Admittedly, our warnings are not novel; although, our empirical documentation of their importance may be. The difficulties of uncovering the global extremum of a nonlinear objective function are widely known and even discussed in first-year econometrics courses. However, to our knowledge, these difficulties have not been publicized to the extent that they should, at least in economics. As a profession, we rarely discuss the processes a researcher has undergone

to find the candidate set of parameters and how the conclusions of the paper change under alternative sets of parameters; gradients and Hessian eigenvalues are almost never reported. The purpose of this paper is to convey that an “exhaustive” optimization exercise and clearly documenting the estimation process are as important to the conclusion of an empirical study as the identification strategy adopted by the econometrician.

We draw our examples from a specific class of discrete-choice demand models for differentiated products which have been particularly popular following the seminal work by Berry, Levinsohn and Pakes (1995) (henceforth BLP). The BLP-type random coefficient logit model allows for more realistic substitution patterns across products compared to the simple logit or nested logit. Consumer heterogeneity exists not only through a mean-zero logit term, but also through variation in willingness to pay for particular attributes of the products under consideration. Faced with a price change in their most preferred good, consumers are more likely to switch to the ones with similar attributes.

Answers to a variety of economic questions use the BLP framework. Estimation of market power, merger analysis, examining the effects of international trade policies and new product valuation, to name a few, have extensively drawn their conclusions based on estimation of BLP-type demand systems.<sup>1</sup> More recently, the analysis of dynamic purchase decisions, such as those associated with durable goods or inventories, rely on BLP-type of models as their starting point.<sup>2</sup>

Estimation of a BLP-type model is a non-trivial optimization problem, even in a low-dimension parameter space. This is primarily due to the highly nonlinear nature of the structural error inferred by equating observed to estimated market shares of the products under consideration combined with the empirical approximation of the structural error used in estimation. To highlight these difficulties, we use two different data sets: the data from BLP and Nevo (2000).<sup>3</sup>

---

<sup>1</sup>New good valuation (Petrin [2004]), Trade Policies (BLP [1999]), Mergers (Nevo [2001a, 2001b]), Construction of price indices that account for quality changes and product introductions (Nevo [2003]).

<sup>2</sup>For dynamic extensions of the BLP model see, for example, Melnikov (2004), Hendel and Nevo (2006), Gowrinsankaran and Rysman (2006) and Hu and Knittel (2007).

<sup>3</sup>In reported results, we have also analyzed two simulated data sets, where we vary the number of random coefficients. We find similar results as those reported for the realworld data, although the issues are certainly less pronounced. The simulated data sets offer the comfort of a known data generating process; the non-simulated data sets are indicative of the real world challenges.

For each data set, we analyze ten different algorithms using 50 different starting values for each algorithm; that is for each data set we estimate the model 500 times. All algorithms are prone to uncover local minima, saddles and to converge at regions with non-zero gradients. The parameter estimates across the local minima and converged points vary greatly both within and across algorithms; more importantly, the variation in parameter values is also economically important. Elasticity estimates can vary by two orders of magnitude across the sets of estimates implied by different algorithm-starting values pairs. The interquartile ranges of a given product’s own-price elasticity often do not overlap across the different algorithms. Even when we focus on the set of starting values that yields the lowest objective function value within each algorithm across the 50 sets of starting values, the estimated own-price elasticities may vary by an order of magnitude. We find greater variation for estimated cross-price elasticities.

The variation across algorithms that exists even when we employ 50 starting values makes it clear that a “multi-start” approach to estimation (see, Andrews [1997]), where the researcher uses a single algorithm but many starting values, is not a sufficient estimation strategy, even when the researcher tries *fifty* starting values. For the two real-world data sets, we find that one algorithm does not find our “global” minimum across both data sets; algorithms that perform well for one data set, may perform very poorly in another.

To further highlight the importance of these issues, we analyze the welfare consequences of hypothetical mergers. The differences can have equally large effects on the estimated consumer welfare consequences of a merger; in short, we find anything is possible, from nearly no effect on consumers to large welfare consequences. Again, this is true even when we estimate the model 50 time for each algorithm and select the best across these 50 starting values.

We stress throughout the paper that the variation we uncover due to multiple extrema is not captured in typical standard error estimates. Reported standard errors are a function of the curvature of the objective function at the local extremum and do not account for the presence of multiple extrema. If consistent estimation requires finding the global minimum and multiple extrema exist, the small sample global minimum may change as the sample changes. While this is beyond the scope of the paper, we speculate that this may generate confidence intervals that are disjoint; at the very least they would be functions of these other local minima.

Part of the variation that we find is also due to the objective function being relative flat

in some regions of the parameter space. This leads to algorithms stopping at multiple places around a local extremum. Reported confidence intervals will account for some of this, but not all. For one, the weak identification literature suggests that “standard” standard errors may be biased. Second, if algorithms are prone to stop prior to reaching the true extremum, then reported confidence intervals will not be centered correctly. Therefore, while the confidence interval may be wide, it will not represent the true confidence interval given the sample.

This paper is related to a small literature analyzing nonlinearities in econometric objective functions. The spirit of our paper is closest to McCullough and Vinod (2003) which illustrates the potential problems associated with finding the solution to a maximum likelihood estimator. The authors have four recommendations for researchers: examine the gradient, the solution path, the eigensystem and the accuracy of the quadratic approximation. While we do not focus on standard error calculations, hence do not emphasize the fourth, we illustrate that verifying the first- and second- order conditions are paramount. McCullough and Vinod also call for a more thorough analysis of the eigensystem, focusing on the condition number, the ratio of the largest to smallest Hessian eigenvalue. A large condition number may indicate an ill-conditioned Hessian, suggesting inaccurate results. While Shachar and Nalebuff (2004) and Drukker and Wiggins (2004) ultimately showed that some of the claims regarding the specific empirical application from McCullough and Vinod were unwarranted, their general point remains; applied researchers should more thoroughly analyze their proposed solutions and report this analysis. In our view this advice has been largely ignored.

Unlike McCullough and Vinod, our goal is not to address the accuracy of specific published results; our paper is not a replication or validation exercise. Our hope is that this work will move the profession in the direction of discussing some of the issues and difficulties related to estimating nonlinear models. We focus on one class of models, but believe a number of our messages apply to nonlinear models in general.

The paper is also related to a literature chronicling some negative properties of GMM estimators; although we do not work out the econometric theory behind our empirical results. A number of papers have documented that GMM estimators in over-identified models may exhibit substantial biases.<sup>4</sup> In addition, when instruments are either weak or many, the criterion

---

<sup>4</sup>See, for example, Hansen, Heaton and Yaron (1996), Altonji and Segal (1996), Burnside and Eichenbaum (1996) and Pagan and Robertson (1997).

function may exhibit regions of plateaus or ridges. This is consistent with some of our results since, in practice, nonlinear search algorithms may “get stuck” at different points on these plateaus.<sup>5</sup> Finally, because over-identified models require an estimate of the moment weighting matrix, researchers often rely on a two-step estimator for the weighting matrix. The weighting matrix in the first step is some positive definite matrix, which yields consistent but inefficient parameter estimates. The second-step weighting matrix uses the consistent estimates from the first step to form the final weighting matrix. Given that the initial weighting matrix is somewhat arbitrary, results may differ across researchers.

## 2 The Model and Estimation

The starting point of the BLP-type demand model is an indirect utility function. A consumer  $i$  derives utility from a product  $j$  in market  $t$ :

$$u_{ijt} = x_{jt}\beta_i - \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt} \quad (1)$$

where  $p_{jt}$  is the product’s price,  $x_{jt}$  is a vector of product characteristics, other than price, and  $\xi_{jt}$  is the unobserved (to the econometrician) product characteristics of product  $j$ , (e.g., after-sale services or unobserved advertising, etc.). The  $\varepsilon_{ijt}$  term captures unobserved heterogeneity in consumers’ tastes.<sup>6</sup> Consumer heterogeneity is also captured by the individual-specific coefficients associated with the price and other observed characteristics, which may be written as:

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \Pi D_i + \Sigma v_i, \quad D_i \sim P_D^*(D), v_i \sim P_v^*(v), \quad (2)$$

where  $D_i$  is a vector of demographic variables,  $v_i$  is a random variable capturing other non-observable characteristics of the consumer and  $\Pi$  and  $\Sigma$  are matrices of parameters. The values of  $\Pi$  measure the importance of the demographic variables in shaping preferences, “observed variation”, while  $\Sigma$  captures “unobserved” variation in preferences. Demographics were introduced in Nevo’s work (e.g., Nevo [2001]); BLP captured heterogeneity only through the  $v_i$ .

---

<sup>5</sup>See, for example, Angrist and Kreuger (1991), Bekker (1994), Bound, Jaeger and Baker (1995), Staiger and Stock (1997), Stock and Wright (2000) and Stock, Wright and Yogo (2002).

<sup>6</sup>This specification ignores income effects, which may be important depending on the application.

Finally, to close the model, one typically includes an outside good with associated utility set to zero, a normalization that is without loss in generality because only relative utilities are identified.

Following Nevo (2000), we decompose the expression in (1) into a component that is common across all consumers and a component that highlights their heterogeneity:

$$u_{ijt} = \underbrace{x_{jt}\beta - \alpha p_{jt} + \xi_{jt}}_{\delta_{jt}(x_{jt}, \xi_{jt}; \theta_1)} + \underbrace{\sum_{k=1}^K D_{ik}\pi_k + \sum_{k=1}^K v_{ik}\sigma_k}_{\mu_{ijt}(x_{jt}, v_i; \theta_2)} + \varepsilon_{ijt} = V_{ijt} + \varepsilon_{ijt} \quad (3)$$

Here we also split the parameters into those associated with the linear portion of utility,  $\theta_1$  and those associated with the nonlinear portion of utility,  $\theta_2$ . Consumers purchase the product that yields the highest utility among the products available. The purchase decisions are limited to a single product. The set of consumers that purchase product  $j$  is given by:

$$A_{jt}(x_{jt}, \delta_{jt}; \theta_2) = \{(v_i, \varepsilon_{i0t}, \dots, \varepsilon_{iJt}) \mid u_{ijt} \geq u_{ilt}, \forall l \neq j\} \quad (4)$$

Therefore, the market share for product  $j$  is the probability that  $A_{jt}$  obtains, which under appropriate independence assumptions may be expressed as:

$$\begin{aligned} s_{jt}(x_{jt}, \delta_{jt}; \theta_2) &= \int_{A_{jt}} dP^*(D, v, \varepsilon) \\ &= \int_{A_{jt}} dP^*(\varepsilon \mid D, v) dP^*(v \mid D) dP_D^*(D) = \int_{A_{jt}} dP_\varepsilon^*(\varepsilon) dP_v^*(v) dP_D^*(D) \quad (5) \end{aligned}$$

For the purpose of estimation, distributional assumptions for  $\varepsilon$ ,  $v$  and  $D$  are necessary. The vast majority of the literature assume Extreme Value and standard normal distributions for  $\varepsilon$  and  $v$ , respectively. Nevo (2001) draws demographics from the empirical distribution of the Current Population Survey as opposed to a parametric distribution.

Estimation follows Berry (1994). For a given  $\theta_2$ , there is a vector of mean utilities,  $\delta$ , that equates implied market shares with actual market shares. We can then decompose this vector of  $\delta$ s into the mean level of observed quality,  $x\beta - \alpha p$ , and unobserved quality,  $\xi$ .

Unobserved quality becomes a ‘‘GMM-like’’ structural error term and the econometrician can readily handle endogenous characteristics. Endogeneity of at least a subset of characteristics is an issue because unobserved quality is known to the firms and consumers. Therefore, we would expect any product characteristic that is easily adapted, such as price, to be correlated with our structural error term.<sup>7</sup>

Formally, if we define  $\theta = (\theta_1, \theta_2)$ , the structural error term is a function of  $\theta$ . Given an appropriate set of instruments, parameter estimation is a GMM problem, with the parameter vector  $\theta$  solving:

$$\hat{\theta} = \arg \min_{\theta} \xi(\theta)' Z \Phi^{-1} Z' \xi(\theta), \quad (6)$$

where  $\Phi$  is a consistent estimate of  $E[Z' \xi \xi' Z]$ . Given the presence of the logit term, estimates is aided by explicitly integrated out the  $\varepsilon$ , yielding:

$$s_{jt}(x_{jt}, \delta_{jt}, P_{ns}; \theta_2) = \frac{1}{ns} \sum_{i=1}^{ns} s_{ijt} \quad (7)$$

$$s_{ijt}(x_{jt}, \delta_{jt}, P_{ns}; \theta_2) = \frac{\exp \left[ \delta_{jt} + \sum_{k=1}^K x_{jt}^k (\sigma_k v_i^k + \pi_{k1} D_{i1} + \dots + \pi_{kd} D_{id}) \right]}{1 + \sum_{m=1}^J \exp \left[ \delta_{mt} + \sum_{k=1}^K x_{mt}^k (\sigma_k v_i^k + \pi_{k1} D_{i1} + \dots + \pi_{kd} D_{id}) \right]}$$

The individual market shares  $s_{ijt}$  are functions of  $\delta_{jt}$  and  $\theta_2$  only. More precisely, for a given  $\theta_2$ , there is a  $\delta_{jt}$  that equates observed market shares with predicted market shares for all  $j$  and  $t$ . The difference between the  $\delta_{jt}$  and  $x_{jt}\beta$  defines the structural error term:

$$\xi_{jt} = \delta_{jt}(S_t; \theta_2) - x_{jt}\beta \quad (8)$$

Ideally, we would obtain analytical expressions for  $\delta_{jt}$ ; however, the highly nonlinear nature of the market share function makes this infeasible. BLP’s estimation strategy is to numerically solve for the vector  $\delta$  that equates observed and predicted market shares. BLP shows that these can be found by a contraction mapping, for which the  $h$ th iterate is given by:

$$\delta_{.t}^{(h+1)} = \delta_{.t}^{(h)} + \ln S_t - \ln S(x_{.t}, \delta_{.t}, P_{ns}; \theta_2) \quad (9)$$

---

<sup>7</sup>In principle, if the  $\xi$  terms are serially correlated we might worry about other characteristics to also be endogenous. For example, if Ford vehicles have a history of being unreliable, Ford may endogenously change their engine characteristics.



The empirical solution to this fixed point yields an empirical approximation of the nonlinear market share function used for estimation.

Estimation requires a nonlinear search to find  $\theta_1$  and  $\theta_2$  that minimizes the GMM criterion function in (6). Because the structural error term is linear in  $\theta_1$ , estimation is aided by “conditioning out”  $\theta_1$  and searching only over  $\delta$  and  $\theta_2$ . In more detail: a new iterate of  $\theta_2$  gives rise to a new vector of mean utilities  $\delta$ , which implies a new iterate for  $\theta_1$ . The new iterate for  $\theta_1$  is derived by means of a linear instrumental-variable approach.

### 3 Computer Code and Search Algorithms

For our optimization exercises, we adapted the code used by Nevo (2000a), written in the Matlab matrix language developed by Mathworks.<sup>8</sup> For a given set of starting values, a few lines of the main body of the code used by Nevo had to be altered to accommodate the setup of the search algorithms used. The bulk of the changes automate loops through 50 starting values and ten algorithms. Given an algorithm, we required a set of starting values and stopping rules. Convergence to a stationary point was declared on the basis of gradients and Hessian eigenvalues, once the algorithm under consideration was terminated according to the stopping rules.

The starting values for the mean utility vector  $\delta$  are the fitted values of a simple logit after adding draws from a zero-mean normal distribution with a standard deviation equal to the standard error of the logit regression, thereby representing the regression error plausibly obtained by a given researcher. For the vector of coefficients  $\theta_2$  of variables entering the nonlinear part of the utility function in (3), we use draws from a standard normal distribution; this represents the fact that little is known about the magnitude of  $\theta_2$  *a priori*.

In unreported results, we have increased the variance associated with the mean utility starting values; our conclusions are magnified. We restrict ourselves to a small region of the parameter space because our goal here is not to uncover all of the local extrema, but rather to analyze the range of possible estimates different researchers may obtain, given reasonable search methods, using the same data and the same model.

---

<sup>8</sup> Available at <http://www.econ.ucdavis.edu/faculty/knittel/>

We employ ten search algorithms that are either derivative-based or direct-search methods. The former utilize some information about the steepness and the curvature of the objective function. The latter are based on function evaluations and are divided into *deterministic* and *stochastic* depending on whether or not they include a random component in their searches. All the algorithms are coded in Matlab. The codes for five of the algorithms are part of the Mathworks Optimization and Genetic Algorithm and Direct Search (GADS) toolboxes. The codes for the remaining algorithms are publicly available from their authors.

Two of our derivative-based algorithms are quasi-Newton, while a third is a conjugate gradient. Judd (1998) and Miranda and Fackler (2002) discuss extensively quasi-Newton methods. Judd and Venkataraman (2002) outline the ingredients of a conjugate gradient algorithm in a very informative way. The codes for the quasi-Newton algorithms are available in the Mathworks Optimization Toolbox and on the website maintained by Hans Bruun Nielsen, respectively.<sup>9</sup> The code for the conjugate-gradient algorithm is also available on the website maintained by Hans Bruun Nielsen. Alexei Kuntsevich and Franz Kappel provide code for our last derivative-based routine based on Shor's r-algorithm that comes from the constrained optimization literature taking into account constraints with the method of exact penalization.<sup>10</sup> Burke et al. (2007) provide a compact self-contained discussion of Shor's r-algorithm (see Kappel and Kuntsevich [2000] for additional details).

Our four deterministic direct-search algorithms are all part of the Mathworks Optimization and GADS toolbox. They include an application of the Nelder-Mead simplex, the Generalized Pattern Search (GPS), and the Mesh Adaptive Direct Search (MADS). We refer to Lagarias et al. (1998) for the mechanics of the Nelder-Mead simplex. Torczon (1997) provides a detailed description of the GPS. Material related to MADS, a generalization of the GPS algorithm, is available in Audet and Dennis (2006).

Our stochastic direct-search routines include genetic algorithms and a simulated annealing algorithm. The codes for the genetic algorithms are provided in the GADS toolbox and on the website maintained by Michael Gordy.<sup>11</sup> Our simulated annealing code is our translation of the code originally developed for the Gauss matrix language by E.G. Tsionas available at the

---

<sup>9</sup> Available at <http://www2.imm.dtu.dk/~hbn/Software/>

<sup>10</sup> Available at: <http://www.uni-graz.at/imawww/>

<sup>11</sup> Available at: <https://www.federalreserve.gov/research/staff/gordymichaelx.htm>

Gauss archive of the American University.<sup>12</sup> We refer to Dorsey and Meyer (1995) and Goffe et al. (1994) for a compact discussion of the genetic and simulated annealing algorithms in the context of econometrics, respectively.

In the results section of the paper, we refer to the ten algorithm we used following the sequence that they were described above as: Quasi-Newton 1, Quasi-Newton 2, Conjugate Gradient, Simplex, GPS, MADS, GA Matlab, GA JBES, Simulated Annealing and SolvOpt.

We experimented with a number of stopping rules for the various optimization algorithms we employed. For the majority of the algorithms, termination was dictated by the change in the objective function and the parameter vector (in some norm) between two consecutive iterations of an algorithm on the basis of a specified tolerance, and/or the number of function evaluations. We used a tolerance of 1E-3 and an upper bound on function evaluations equal to 4,000. Once again, we distinguish between the notions of termination as dictated by the stopping rules and convergence due to meeting the tolerance.

The variation across algorithms in the use of convergence criteria in forming stopping rules is something that we would like to stress; how these rules are used varies across algorithms. As an example, it took us several rounds of e-mail exchanges with Mathworks before we find out how the tolerance settings for the objective function and the parameter vector are used to determine the termination of its quasi-Newton and the simplex optimization routines. The alternative route to retrieve this information was through cross references in numerous highly convoluted Matlab routines. *Future versions of the paper will include a more thorough description of each algorithm.*

Imposing an upper bound on the number of function evaluations was largely dictated by the use of direct-search algorithms, notably the Nelder-Mead simplex algorithm, which repeatedly appeared to “stall”: it continued to move in a small neighborhood of the parameter space, without appreciable changes in the objective function. As an example, for the first set of starting values used with the BLP data set the simplex algorithm did not improve upon the value of the objective function by more than 1E-3 after the first 60 iterations that required 157 function evaluations for the next 7,843 function evaluations.<sup>13</sup> However, because the diameter

---

<sup>12</sup>Available at: <http://www.american.edu/academic.depts/cas/econ/gaussres/optimize/optimize.htm>. The Matlab code is available from the authors upon request.

<sup>13</sup>We capped the number of function evaluations at 8,000 in this example.

of the simplex was never less than 1E-3, convergence is never achieved.

In what follows, we assume a simplex algorithm has converged if the objective function remains unchanged to the third decimal point for at least 200 iterations (which typically correspond to over 800 function evaluations). The bulk of our reported results are broken up by algorithm, so the reader can see that this assumption does not change the conclusions.<sup>14</sup> For the remaining algorithms, we focus only on those sets of starting values that meet our convergence criteria and omit those results that are bound by the function evaluation constraint.

## 4 Results using Real-world Data

In this section, we summarize our demand estimation results using the ten optimization algorithms described above for two real-world data sets. The first consists of the data on automobile sales used in BLP. The second is the cereal data used in Nevo (2000a). Much of our motivation for the use of these data was due to the fact that they are publicly available. *We are also in the process of analyzing simulated data.*

In the case of the cereal data, our specification of the demand equation is the one used by Nevo. In the case of the automobile data, our specification of the demand equation is slightly different from the one used by BLP; for example, we have no interaction between price and income and a smaller number of automobile characteristics with random coefficients. Recall that we do not attempt any sort of replication or validation of the optimization approaches followed by any of the authors.

We present results with respect to the value of the GMM objective function, parameter estimates, own- and cross-price elasticities. We also check how the variation in parameter estimates affects our conclusions regarding the welfare effects of two hypothetical merger exercises: one for each data set. We focus only on the sets of starting values for which the various algorithms converged. Within each of the ten search algorithm, we define the “best” set of parameters as the set that minimizes the GMM objective function across the 50 sets of starting values. We

---

<sup>14</sup>This is why some have recommended that researchers begin with a Simplex method and then switch to a quasi-Newton method when the simplex method has “stalled.” If we were to adopt this strategy, we are confident that each of the reported Simplex results would then converge after switching to the Quasi-Newton algorithm.

define, the “best of the best” set of parameters is the set that minimizes the GMM objective function across all 500 combinations of starting values and optimization algorithms. Our definition of best is admittedly somewhat arbitrary because the consistency proofs may be local in nature.

We start by analyzing the range of the GMM objective function values implied by the sets of starting values for the parameters that allowed the algorithms considered to converge; we spend little time on these since they are difficult to interpret. Figures 1 and 2 are whisker plots of the GMM values across parameter starting values and algorithms.<sup>15</sup> We truncate the GMM values at their 90th and 75pth percentiles for the automobile and the cereal data, respectively. The GMM values fluctuate substantially across starting values even within an algorithm. Such a finding is not surprising. In fact, many researchers often try more than one starting value. However, the GMM values also vary widely *across* algorithms, even when we focus only on those GMM values implied by the best set of parameters.

For the automobile data, the lowest value of the GMM objective function for each of the ten algorithms lies between 99.90 and over 35,000 for those starting values of the parameters that allowed the algorithms to converge. The number of parameter starting values that led to convergence is 350. The range of GMM values implied by the best sets of parameters is between 99.90 and 230.23. Only the MADS algorithm gave rise to a GMM value equal to 99.90 even after using 50 starting values for each algorithm. For the cereal data, the GMM values lie between 4.56 and 530.12, if we focus only on the best set of parameters for each algorithm. Once we consider the entire set of parameter starting values that implied convergence, the range of the GMM values is between 4.56 and 2,241.51. Convergence was achieved for 301 of the sets of parameter starting values.

One interesting finding in our extensive optimization exercise is that different algorithms uncover the best of the best set of results across the two data sets. For the cereal data, SolvOpt finds the global minimum across all 50 starting values, which we believe to be above the typical number of parameter starting values employed in the majority of the empirical exercises in

---

<sup>15</sup>The box represents the 25th and 75th percentiles with a median line. Whiskers extend from the box to the upper and lower adjacent values and are capped with an adjacent line. The upper adjacent value is the largest data value that is less than or equal to the third quartile plus  $1.5 \times \text{IQR}$  and the lower adjacent value is the smallest data value that is greater than or equal to the first quartile minus  $1.5 \times \text{IQR}$ . Dots represent values outside these “adjacent values”.

economics. For the automobile data, SolvOpt is dominated by MADS. Therefore, it is not enough for researchers to use multiple starting values and one algorithm, even when they use as many as 50 starting values. *For an exhaustive nonlinear search, researchers will need to use multiple starting values, at least 50, and multiple algorithms.* As we will show below, even 20 different sets of starting values and ten algorithms may not be enough.

Focusing only on the value of the GMM objective function may be misleading. If the objective function is steep around the true parameter values, a local minimum may yield parameter values that are close to the true values, but have an objective function value that is very different. Therefore, we focus on the economic meaning of the variation in parameter estimates.

Tables 1 and 2 report the best set of parameter value across algorithms for the automobile and cereal, respectively. The variation in parameter values across the different algorithms suggests economically meaningful differences. For example, for the automobile data, the absolute value of the coefficient associated with the log of price ranges between 0.57 and 0.90; sometimes the mean marginal utility for horsepower is positive, while other times it is negative. In the cereal data, the coefficient associated with price lies between 17.77 and 62.72 (in absolute value), although the presence of interaction terms may make this variation somewhat misleading. The parameter values associated with the interaction term of price and income vary between -0.07 and 588.56.

Furthermore, all of the parameter values seem “reasonable.” This is important because if only the parameter values associated with the best of the sets of results seemed reasonable, the researcher may continue to search until he or she found this minimum. For example, if all but one of the sets of parameters yielded upward sloping demand curves, this would provide an economic justification for choosing among the candidate set of results.

Taken alone, even the parameter values can be difficult to interpret because monotonic transformations of a particular set of parameters may yield similar behavior and price/demographic interactions make interpretation difficult.<sup>16</sup> To gauge the economic significance of the different parameter values, we construct a variety of often-used functions of these parameters. We focus on three measures: own- and cross-price elasticities, as well as welfare calculations from

---

<sup>16</sup>While fixing the variance of the logit error term allows us to identify the parameters, proportional increases in the remaining parameters are likely to yield similar substitution patterns.

hypothetical mergers. Analyzing implied price-cost margins we drew largely similar conclusions.

## 4.1 Own-Price Elasticities

The number of elasticities that we estimate is rather immense. Each data set has over 2,000 product and market combinations. Every combination of parameter starting values and optimization algorithm yields an elasticity matrix for each market. In the case of the cereal data there are 94 market with 24 products in each market leading to 2,256 market and product combinations. Each of the 500 pairs of parameter starting values and algorithms implies an elasticity matrix of dimension  $24 \times 24$  for each of the 94 markets. In the case of the automobile data, there are 20 markets, which may have as many as 150 products, for a total of 2,217 product and market combinations.

To keep the discussion of the own-price elasticity estimates manageable, we begin by focusing on four products for each of the data sets. These four products have a market share that corresponds to the first, second, third and fourth (maximum) quartiles of distribution of market shares. For the purpose of the discussion below, we refer to these products, as products 1 (1st quartile), 2, 3 and 4 (max), respectively. Subsequently, we provide kernel density plots of the own-price elasticity estimates for all products.

When we discuss elasticity estimates for the four products, we present results for all starting values that allowed convergence, as well as for the best set of parameter values for each algorithm. When we present elasticity estimates for all products we focus only on the best set of parameter values for each algorithm. Recall that we define as best the set of parameter that give rise to the minimum GMM values for the algorithm under consideration. Holding the algorithm fixed, but looking across starting values, illustrates the importance of trying multiple starting values. Variation across the best set of results for each algorithm stresses the importance of trying multiple starting values *in combination* with multiple algorithms.

In the presence of a globally convex objective function the elasticity estimates for a specific product would be the same across all sets of starting parameter values that implied convergence. Furthermore, if the nonlinear search problems were mild enough such that trying many starting values would suffice to overcome them, the distribution of the estimated elasticities would

be identical across the ten best sets of estimates. The amount of variation in the estimated elasticities gauges our concerns for the optimization method employed in our demand estimation exercise.

#### 4.1.1 Automobile Data

Table 3 lists the own-price elasticity estimates for the products with market shares that correspond to products 1 through 4, as defined in the previous section, across all algorithms. For each algorithm, we report the range of own elasticities associated with those starting values that permitted convergence. The table also includes elasticities implied by the best set of parameters for each algorithm. Recall that MADS achieved the lowest GMM value. Because the JBES genetic algorithm never converges, we omit discussion of the implied elasticities.

The variation in the estimates of the utility parameters affects the own-price elasticity estimates for each of the four products significantly. If we focus on parameter starting values that implied convergence, the own-price elasticity for product 1 varies between -10.11 and 0.12. If we believed a researcher would not trust the positive own-price elasticity, the range is between -10.11 and -1.78, a difference of over 500 percent. The other three products exhibit even larger variation. The elasticity estimates for product 2 lie between -20.46 and -0.26. Moving to products 3 and 4, the corresponding ranges are -12.67 to -0.19, and -16.89 to -0.38, respectively.

To get a feel for whether these extremes are outliers, Figure 3 plots the histogram of elasticities for the 350 parameter starting values that permitted convergence. The figure also reports the “true” elasticity, where the truth is defined as the elasticity associated with those parameter values that yield the lowest value for the GMM objective function. For all four products, the truth lies outside the modal bin, and can be very far from the mode. Furthermore, in each case, a significant amount of the distribution falls outside what would appear to be reasonable variation in the estimates. We admit that given these differences do not represent sampling variation, one could argue that reasonable variation is no variation.

Another way to summarize our findings is to calculate the standard deviation of the own-price elasticity for each product-market combination for each of the 350 sets of parameter starting values that allowed convergence. Ideally the distribution of these standard deviations



should be degenerate with a mass at zero. The mean of this standard deviation is 4.25, its median is 3.07, while the 25th and 75th percentiles are 2.52 and 3.64, respectively. To put these numbers in context, the mean elasticity is -3.66.

Next, we focus on the best set of parameter starting values for each algorithm. This mimics cases where ten researchers opt for a particular algorithm and use 50 different starting values. The variation remains substantial. For product 1, elasticities vary by more than 100 percent, from -6.38 to -2.54. The other three products exhibit smaller, but still economically meaningful, variation: -5.45 to -3.20 (product 2), -7.14 to -4.37 (product 3), -4.64 to -3.95 (product 4). The mean within-product-market standard deviation among the best set of parameter starting values is 0.84 representing 23% of the mean elasticity.

Finally, we turn to summaries of the entire set of own-price elasticities. Figure 4 plots kernel density estimates for the best set of parameters for each of the ten algorithms. Again, absent nonlinear search problems, we would expect the lines associated with the kernel densities to lie on top of each other; they do not. We should also note that these densities are likely to mask meaningful variation in the elasticities across algorithms because it is possible to change the elasticity of specific products without changing the distribution. The densities exhibit large fluctuations, with the true density appearing on one extreme. The JBES genetic algorithm never converges, so we do not place much weight on this algorithm, but the remaining nine algorithms that do converge, do so at different points in the parameter space.

Much of our discussion above is consistent with a hypothetical setup of ten researchers using 50 different sets of starting values, but a single algorithm, to solve the same optimization problem. We speculate that many researchers do not try 50 distinct sets of starting values most of the times. We have often heard researchers arguing that their estimation process can take weeks to converge. An estimation exercise that requires computation time worth of a week, using a single computer would imply almost an entire year to try 50 different sets of starting values, in the absence of parallel processing. Furthermore, it seems to be the case that the more complex the model, the larger the number of starting values that the researcher should employ. This, of course, is unfortunate because computation time and complexity are positively correlated, implying the more complex the model the more starting values a research should try.

To understand the importance of trying many starting values, we replicate the own-elasticity density plots for all products assuming a researcher using only the first 20 of our starting values. These results are plotted in Figure 5. *The “truth” is never found.* Ironically, the variation across algorithms, with the exception of the simulated annealing, is smaller using only 20 starting values. This smaller variation is somewhat misleading because the GMM objective function can only be improved by trying additional starting values, suggesting fewer starting values may give researchers a false sense of security.

#### 4.1.2 Cereal Data

In general, our findings regarding own-price elasticities for the cereal data are more robust, although we would argue the variation is still economically significant. Table 4 reports the range of possible own-price elasticity estimates for four products. We maintain the nomenclature for product identification that we developed in the previous sections discussing results associated for the automobile data set. For product 1, the range of elasticities across all sets of parameter associated with convergence is between -5.77 and -0.42. For product 2, focusing only on the negative values of elasticities, the range is -17.97 to -1.71. The range for products 3 and 4 is -7.51 to -0.01, and -2.46 to -0.01, respectively. Assuming a researcher tries 50 starting values, a single algorithm and reports the results from the best set of parameters, the range is still large, range from -4.43 to 2.93, -13.08 to -2.62, -2.92 to -1.09 and -1.99 to -0.77 for products 1, 2, 3 and 4, respectively.

The mean standard deviation of the elasticity estimates for a given product-market combination, is 0.82 among estimates associated with convergence, while the mean elasticity is -3.58. Using a normal approximation, this suggests that 34 percent of the time the own-price elasticity will vary by over 20 percent. Interestingly, although the four specific products exhibit less variation compared to the automobile data, their within product-market combination standard deviations are similar.

Histograms of own-price elasticities are provided in Figure 6. Largely because of two algorithms, the truth always lies in the modal bin. As noted above, SolvOpt finds the truth for each of the 50 sets of starting values and quasi-Newton 2 does for 32 of 50 sets of starting values. Because we are unaware how often these algorithms are used, we also provide histograms using

only the remaining six algorithms that exhibited convergence. Omitting SolvOpt and Quasi-Newton 2 changes the shape of the histograms considerably. The remaining algorithms rarely find elasticities economically near the “true” elasticity.

The kernel densities of own-price elasticities across all products for each of the best set of parameters are plotted in Figure 8. We see large shifts in the distribution of elasticities across algorithms, although not as dramatic as with the automobile data set, especially among those algorithms that converged (simulated annealing and the genetic algorithms did not converge). Given that the results from the individual product estimates and the standard deviations suggest that for a given product the variation can be considerable, these densities are likely hiding meaningful variation. This is corroborated by the within product-market standard deviations discussed above.

## 4.2 Cross-Price Elasticities

In some respects, analyzing cross-price elasticities is more important than own-price elasticities because the random coefficient logit is designed to provide more appealing substitution patterns than the simple and nested logit models. The number of cross elasticities is even more immense when compared to the number of own elasticities. There are over 230,000 cross elasticities in the automobile data and over 50,000 in the cereal data. As with our analysis of own-price elasticities, we first focus on four specific products and then include densities for a large number of elasticities. The product-specific analysis requires choosing a cross-product. We choose the product that is the closest substitute using the parameter values that achieve the lowest value of the GMM objective function.

In principle our density plots can include all of the estimated cross elasticities; however, we have found that the density plots of the entire set of cross-price elasticities hide meaningful variation in the estimated elasticities, as the following example illustrates. Suppose we are interested in the cross elasticities of products  $a, b, c$  and  $d$  and we use  $\eta_{ij}$  to denote the price cross price elasticity of good  $i$  with respect to good  $j$ . Furthermore, assume that one combination of starting values and optimization algorithms implies  $\eta_{ab} = 0.05$  and  $\eta_{cd} = 0.15$ , while a second combination implies  $\eta_{cd} = 0.05$  and  $\eta_{ab} = 0.15$ . The density plots will not reveal the meaningful variation in these cross elasticities due to the different combinations of starting values and

algorithms. To alleviate similar problems, we focus on a subset of the cross elasticities, chosen independent of the results, keeping this subset the same across the the pairs of starting values and optimization algorithms. For the cereal data, we choose the product with the highest average cross-price elasticity across all markets.

For the automobile data, this is more difficult since the products change across markets (time periods). Our subset of elasticities is somewhat arbitrary, but we stress it is identical across all sets of parameters. We choose our subset as follows: We first define a product number from 1 to the  $J_M$  in each market using the product numbers from the original data sources. For each product within the market, we calculate the cross-price elasticity each of these  $J_M$  products (one of which is the own-price elasticity). For each product  $i$  in market  $\kappa$ , this gives us  $\eta_{i1}^\kappa, \dots, \eta_{ij}^\kappa, \dots, \eta_{iJ_M}^\kappa$ . We then choose the  $\eta_{i,j}$  which has the highest average cross-price elasticity across all markets,  $\kappa$ , and products  $i$  (ignoring the own-price elasticities within this calculation). This is identical to how we choose the cross elasticities for the cereal data, but because the products change across markets in the automobile data, the interpretation is not as clean.

#### 4.2.1 Automobile Data

Similar to own-price elasticities, cross-price elasticities can differ by an order of magnitude across the sets of parameters associated with convergence, sometimes two (Table 5). For product 1, the cross-price elasticity with its closest substitute ranges from 0.010 to 1.33 (the entries in Table 5 have been multiplied by 10). If we restrict ourselves to the best set of parameters for each algorithm, the elasticity ranges by a factor of four: 0.05 to 0.21. For product 2, the range across all parameters associated with convergence is 0.01 to 2.15; for the best set of parameters the range is 0.08 to 0.26. Product 3 exhibits the most variation. For the set of parameters associated with convergence, the range is 0.003 to 15. If we move to the best set of parameters, the range becomes 0.06 to 0.20. For product 4, the ranges are 0.013 to 0.87 and 0.08 to 0.20 across all sets of parameters associated with convergence and the best sets of parameters, respectively.

While Table 5 points to substantial variation, it is still possible that the true elasticity is found most of the time. Figure 9 plots the histogram of possible cross-price elasticities for each

of the four products. For visual ease, we truncate the horizontal axes of the graphs at unity. In some ways, the results for cross-price elasticities are even more dramatic compared to own-price elasticities. In each case, the truth is outside the mass (thick portion) of the distribution. Furthermore, the range of this thick portion of the distribution is large; it is from almost zero to just under 0.20 for three of the four products.

Figure 10 provides kernel density plots for the large set of cross-price elasticities described above using the best set of parameters. As with the histograms, the best of the best set of parameter values imply larger cross-price elasticities. This is more easily seen in the density plots for elasticities below 0.05 (Figure 11) and above 0.05 (Figure 12). In both cases, more of the distribution for the truth is to the right.

Finally, we calculate the within product-market standard deviation in the cross-price elasticities. We calculate the standard deviation across all routines that converged and across the best sets of parameters for each algorithm (ignoring those algorithms that never converge). The average cross-price elasticity across the over 230,000 elasticities and those routines that converged is 0.06. The average within product-market standard deviation in these estimates is 0.17, while the median is 0.04 with an interquartile range is 0.03 to 0.05. Therefore, it is not uncommon for estimates to vary by plus or minus 100 percent within a given product.

#### 4.2.2 Cereal Data

The cereal data also exhibit large variation in their cross-price elasticities. As Table 6 illustrates, for each of the products 1 through 4, it is possible to estimate a zero cross-price elasticity, or an elasticity that, in some cases, exceeds 1.5 (the entries in Table 5 are multiplied by 10). This is true even when we choose the best set of parameters implied by all 50 starting values. Product 1 exhibits the smallest variation among the four products: 0.20 to 0.36. The elasticities for product 2 lie between 0 and 0.35. For product 3, the range of elasticities is 0.07 to 0.52. Finally, product 4 has cross elasticities that range from 0.03 to 0.21. This variation seems to be larger than that of the automobile data.

Figure 13 shows that there is a significant mass of cross-price elasticities ranging from zero to over 0.5, especially in cases where the truth is at the high end of the distribution. Similar

to the own-price elasticities, the spike at the truth is driven by two of the ten algorithms. When we omit these algorithms from the histograms, we see that we are more likely to estimate elasticities significantly far from the truth, than the truth itself (see Figure 14).

When we plot the distribution of a large number of cross-price elasticities, as we did with the automobile data, the cereal data appear to exhibit even more variation across the best set of results for each algorithm (Figure 15). This finding may not be surprising given the variation in the nonlinear parameters reported in Table 2. For example, the standard deviation of the random coefficient term associated with price varies from 0.67 (Conjgrad) to 3.31 (SolvOpt). The price-income interaction term varies between -0.56 and 588. These terms will drive differences in estimated cross-price elasticities. Similarly, the nonlinear parameters associated with other characteristics also vary by over 100 percent. For example, the mushy-income interaction term ranges from 0.75 to 2.22.

Figures 16 and 17 focus on different ranges of the cross elasticities.

The within product-market standard deviation in the cross-price elasticities show similar patterns to the automobile data. The average cross-price elasticity across the over 50,000 elasticities and among those routines that converged, is 0.15. The average within product-market standard deviation for these estimates is 0.50, while the median is 0.20 with an interquartile range of 0.13 to 0.41. As a percentage of the average cross-price elasticity, this variation is even larger once compared to the variation in the automobile data; here, it is not uncommon for estimates to vary by plus or minus 300 percent within a given product. These ranges are similar even among the best outcomes for each algorithm (also ignoring those best outcomes that did not converge). The mean cross-price elasticity is 0.13, while the mean within product-market standard deviation is 0.19, with 0.13 as the median. Therefore, even if researchers try 50 starting values, but using one algorithm, a significant amount of variation is still present.

### 4.3 Merger Results

Estimation of a BLP-type demand model yields a system of equations that amounts to a matrix of price derivatives for each product. Using information on the ownership structure of the market, a researcher can exploit the first-order conditions under an assumed conduct mode

to yield estimates of marginal cost. For example, under static Bertrand-Nash behavior and constant marginal costs, the first-order conditions associated with the firms' profit-maximization problems imply:

$$p - mc = \Omega(p)^{-1} s(p),$$

where  $p$  is the price vector,  $s(\cdot)$  is the vector of market shares and  $mc$  denotes the corresponding marginal costs. The dimension of these vectors is equal to the number of the products available in the market, say  $J$ . The  $\Omega$  matrix is the Hadamard product of the (transpose) of the matrix of the share price derivatives and an ownership structure matrix. The ownership structure matrix is of dimension  $J \times J$ , with an  $(i, j)$  element equal to 1 if products  $i$  and  $j$  are produced by the same firm and zero, otherwise. Because prices are observed and demand estimation allows us to retrieve the elements of  $\Omega$ , marginal costs are directly obtained using the expression in ().

A simple change of 1s and 0s in the ownership structure matrix accompanied with a series of assumptions discussed below allows the econometrician to draw inferences regarding the welfare effects associated with a change in the industry's structure, as the one implied by mergers among competitors.<sup>17</sup> We analyze the range of possible welfare calculations for a hypothetical merger for both the automobile and the cereal data sets on the basis of post-merger equilibrium prices. Using the automobile data, we assume GM and Chrysler merge. In the case of the cereal data set, we assume Kellogg's and General Mills merge. The vector of post-merger prices  $p^*$  is the solution to the following system of non-linear equations:

$$p^* - \widehat{mc} = \widehat{\Omega}^{post}(p^*)^{-1} \widehat{s}(p^*). \quad (11)$$

The elements of  $\widehat{\Omega}^{post}$  reflect changes in the ownership structure implied by the hypothetical merger. To solve for the post-merger prices, we keep the share-price derivatives and shares at their use pre-merger levels instead of solving the above system of non-linear equations in (11).<sup>18</sup> Thus, we avoid dealing with issues related to the numerical instabilities of Newton routines used in the `fsolve` routine of Matlab for the solution of nonlinear equations, as well as with issues

---

<sup>17</sup>Exercises for evaluating the welfare effects associated with the introduction of new goods are performed in a very similar manner.

<sup>18</sup>This approximation is also discussed in Nevo (1997)

related to potentially multiple equilibria; we have found that these issues affect post-merger prices. Although we agree that the discussion of both issues is important, it is beyond the scope of this paper.

With the post-merger prices in hand, we can estimate expected consumer welfare changes due to the mergers under consideration. A consumer's expected change in utility due to a merger may be evaluated as the change in her logit inclusive value (McFadden [1981], Small and Rosen [1981]). Therefore, the compensating variation for individual  $i$  is the change in her logit inclusive value divided by the marginal utility of income. When prices enter the utility function linearly, which holds in our case, the compensating variation is given by:

$$CV_i = \frac{\ln \left[ \sum_{j=0}^{j=J} \exp(V_{ij}^{post}) \right] - \ln \left[ \sum_{j=0}^{j=J} \exp(V_{ij}^{pre}) \right]}{\alpha_i}, \quad (12)$$

where  $V_{ij}^{pre}$  and  $V_{ij}^{post}$  are defined in (3) using the pre- and post-merger prices. Integrating over the density of consumers yields the average change in consumer welfare from the merger.<sup>19</sup>

### 4.3.1 Automobile Data

The automobile data set is a repeated cross-section of virtually all cars sold in the US between 1971 and 1990 at an annual frequency. We assume that the GM and Chrysler merger takes place in the beginning of our sample, namely 1971. Therefore, we have an annual estimate of the merger's impact on prices. We report the average of the price impact across all years. The average price change across all optimization routines that achieved convergence is \$265, with a median of \$0. The average within-product standard deviation in this estimated price change is \$1,020 across all sets of parameters associated with convergence. The median of the standard deviations is zero, given all of the zero predicted price changes and the 75th percentile is \$427. Therefore, it is not uncommon to see variation in the estimated price changes of over 100 percent.

Figure 19 plots the price change densities for each of the best sets of parameters, conditional on a product having a price change. Large differences are observed across the distributions,

---

<sup>19</sup>Nevo (2000b) contains a typo in the reported formula; the paper omits the  $\exp()$  operators.



which translate into large differences in the merger effects on the consumers. Figure 20 plots the histogram of welfare changes assuming there are 265 million consumers. Even when we ignore uncertainty due to sample variation, the predicted consumer effects vary from near zero to over \$40 million, with a large mass above \$20 million. Given that there may be no statistical or economical way to choose among local minima, depending on which side of the merger one lies, a compelling argument can be made on her behalf. We take some comfort observing that the truth lies somewhat in the middle of this distribution; however, accounting for the uncertainties with respect to non-linear search issues and sampling error implies that we may learn very little about a given market following the merger counterfactual.

### 4.3.2 Cereal Data

In the cereal data, the average price change across all routines achieving convergence is 1.86 cents per serving, with a median of 1.62 cents per serving. The average within-product standard deviation in this estimated price change is 11.04 cents across all sets of parameters associated with convergence. Outliers drive this high mean as the median is 1.38 cents, nearly 100 percent of the typical price change. Across the best set of parameters, the average price change is 1.76 cents. The median price change is lower (1.61) cents, while the mean and median within-product standard deviation are 0.24 and 0.16 cents, respectively. The 75th-percentile standard deviation is 0.32 cents. Therefore, even among the best set of parameters, it is not uncommon to see price changes for a particular product vary by more than 50 percent.

Figure 21 plots the price change densities for each of the best sets of parameters, including all products for which a price change due to the merger takes place. As with the previous sets of results associated with own- and cross-price elasticities, the cereal data do not exhibit as much variation as in the automobile data. We do, however, observe large differences in the estimated welfare effects, especially when we omit the two algorithms that clearly outperform the others. Figure 22 plots the histogram of welfare changes assuming there are 265 million consumers for all algorithms. Figure 23 omits the algorithms 3 and 5. While there is considerable variation, the results are not as striking as with the automobile data.

## 5 Gradients and Eigenvalues

Any discussion of the first- and second-order conditions for a minimum has been carefully postponed to this point. We should have probably begun our analysis with this discussion because it would allow us to rule out certain candidate sets of parameter estimates. Our timing of the presentation has been deliberate for a number of reasons. To the best of our knowledge, first- and second-order conditions are rarely discussed by empirical economists in their work. Additionally, we want to stress the point that optimization algorithms may stop at points in the parameter space where these conditions are not satisfied. Finally, and most importantly, for both the cereal and the automobile data sets the “global” minimum found and discussed above does not meet these conditions.<sup>20</sup>

### 5.1 Automobile Data

Among the 500 combinations of starting values and optimization algorithms, 350 achieved convergence. Among these 350 combinations, 111 have gradients for which the  $\infty$  – *norm* is below 30, 92 have gradients with  $\infty$  – *norm* below 20 and 59 have gradients with  $\infty$  – *norm* below 10.<sup>21</sup> We realize that these are fairly lenient standards for a zero gradient, but any  $\infty$  – *norm* gradient cut-off will illustrate our message.

Using 30 as the  $\infty$  – *norm* gradient cut-off, we find that only 19 of the 500 sets of parameters corresponds to points that meet both the first- and second-order conditions. Most interestingly, *these results do not include the “global” minimum*, implying that this point is not a minimum. Despite trying 500 different starting value and algorithm combinations and having 350 of them “converge”, we know that we have not found the global minimum. Table 7 lists the values of the GMM objective function and parameter estimates for these 19 sets of results. It would appear that these results correspond to roughly 11 unique local minima. We note that we are

---

<sup>20</sup>To analyze the first- and second-order conditions, we calculate numerical gradients and Hessians using the Matlab routines `fminunc.m` and `eig.m`, respectively. Although `fminunc` is an optimization routine it provides gradients and Hessians as by-products.

<sup>21</sup>Among the entire set of 500 starting value and algorithm combinations, 152, 126 and 87 have gradient  $\infty$  – *norm* below 30, 20 and 10, respectively.

restricting ourselves to a fairly small neighborhood of starting values; there are probably other local minima to be uncovered if we expand this neighborhood.

Figures 24 and 25 provide kernel density plots of own- and cross-price elasticities that correspond to the sets of parameter estimates associated with the 11 local minima. If anything, it appears that the variation across these sets of results is even larger compared to the variation when focusing on the best set of results for each algorithm. Figure 26 is a histogram of consumer welfare changes from the hypothetical merger between Chrysler and GM discussed above. Again the variation is large; the consumer welfare effects vary between -\$2.13 million and -\$15.35 million. Therefore, even if the a researcher was diligent and checked her first- and second-order conditions, it is conceivable that she could converge on any one of the sets of parameters.

## 5.2 Cereal Data

Among the 301 combinations of parameter starting values and algorithms that implied convergence, 142 have gradients with  $\infty - norm$  below 30, 126 have gradients with  $\infty - norm$  below 20 and 113 have gradients with  $\infty - norm$  below 10.<sup>22</sup> Only 37 starting value and algorithm pairs with a gradient  $\infty - norm$  below 30 meet the second-order conditions. Once again, the “global” minimum discussed above is not a true minimum; it meets the first-order conditions, but the associated Hessian contains negative eigenvalues. Once again, we are left knowing that a lower point in the GMM objective function still exists.

The parameter estimates associated with the 37 local minima are listed in Table 8. It is more difficult to point out the number of distinct local minima that these estimates represent, so we plot the kernel densities associated with the implied own- and cross-price elasticities. The kernel density plots of the own elasticities are suggestive of two distinct local minima (Figure 27). However, when we plot the kernel densities of the cross-price elasticities in Figure 28, there is much more variation. Our take-away from Figures 27 and 28 is that the objective function contains plateaus or ridges that affect cross- but not own-price elasticities.

The welfare effect of the hypothetical merger between Kellogg’s and General Mills evaluated at the 37 sets of parameters corresponding to local minima do not exhibit the degree of variation

---

<sup>22</sup>Across all 500 pairs, 156 have gradients below 30, 137 have gradient  $\infty - norm$  below 20 and 118 have gradient  $\infty - norm$  below 10.

we experienced with the automobile data. Figure 29 shows that the welfare effects vary by 50 percent, as opposed to over 700 percent for the automobile data set.

### 5.3 Discussion of Eigenvalues

Given our fairly lenient standards for zero gradient norms, an obvious concern regarding our Hessian eigenvalue calculations is that while locally a given eigenvalue may be negative, perhaps small movements around the point under consideration would yield positive eigenvalues. If this were the case, for example, the lowest point for the cereal data may indeed be a minimum. In support of this argument, parameter estimates associated with values of the GMM objective function in the neighborhood of 14.9 have all of their Hessian eigenvalues positive. Against the same argument, none of the 82 starting value and algorithm pairs that implied values for the objective function around 4.56, the lowest value of the objective function we uncovered, have all of their Hessian eigenvalues positive. The sensitivity of the Hessian eigenvalues to very small movements around the proposed minimum is more worrisome, we would argue. This sensitivity may also imply that certain points with a near-zero gradient norm may have all positive Hessian eigenvalues in only a small neighborhood. Therefore, a researchers may wrongly stop at a point that isn't truly a minimum because her tolerance for a zero gradient is too large.

## 6 Conclusions

Empirical industrial organization has been increasingly relying on highly nonlinear structural models. Researchers are often concerned about econometric issues, such as endogeneity and the variation in the data that can, in principle, identify the parameters of interest. However, the actual process of finding the extremum of the underlying objective function involved in the empirical exercise undertaken is rarely discussed, and if it is, is often relegated to a terse footnote.

In this paper, we show that an econometrician's search for the extremum can have large consequences on the conclusions drawn for variables of interest in economic analysis. We believe that these issues deserve as much attention as the identification strategy. For a common class

of demand models for differentiated products, we show that depending on the search algorithm of the researcher, a wide range of policy implications exist. Furthermore, parameter estimates of “converged” routines may not satisfy the first- and second-order conditions for an extremum.

## References

- [1] Altonji, Joseph G., and Lewis M. Segal. 1996. "Small-Sample Bias in GMM estimation of Covariance Structures." *Journal of Business and Economic Statistics*, 14(3):353-366.
- [2] Andrews, Donald W.K. 1997. "A Stopping Rule for the Computation of Generalized Methods of Moments Estimators," *Econometrica*, 65(4): 913:931.
- [3] Audet, Charles, and J.E. Dennis JR. 2006. "Mesh Adaptive Direct Search Algorithms for Constrained Optimization." *Siam Journal of Optimization*, 17(1): 188-217
- [4] Amemiya, Takeshi. 1985. *Advanced Econometrics*. Cambridge: Harvard University Press.
- [5] Angrist, Joshua D., and Alan.B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?." *Quarterly Journal of Economics*, 106(4):979-1014.
- [6] Bekker, Paul A. 1994. "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica*, 62(3): 657-681.
- [7] Berry, Steven. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics*, 25(2): 242-262.
- [8] Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841-890.
- [9] Berry, Steven, James Levinsohn, and Ariel Pakes. 1999. "Voluntary Export Restraints on Automobiles. Evaluating a Trade Policy." *American Economic Review*, 89(3): 400-430.
- [10] Bound, John, David A. Jaeger, and Regina M. Baker. 1995. "Problems with Instrumental Variable Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association*, 90(430): 443-450.
- [11] Burke, James V., Adrian S. Lewis, and Michael L. Overton. 2007. "The Speed of Shor's R-algorithm". Manuscript.

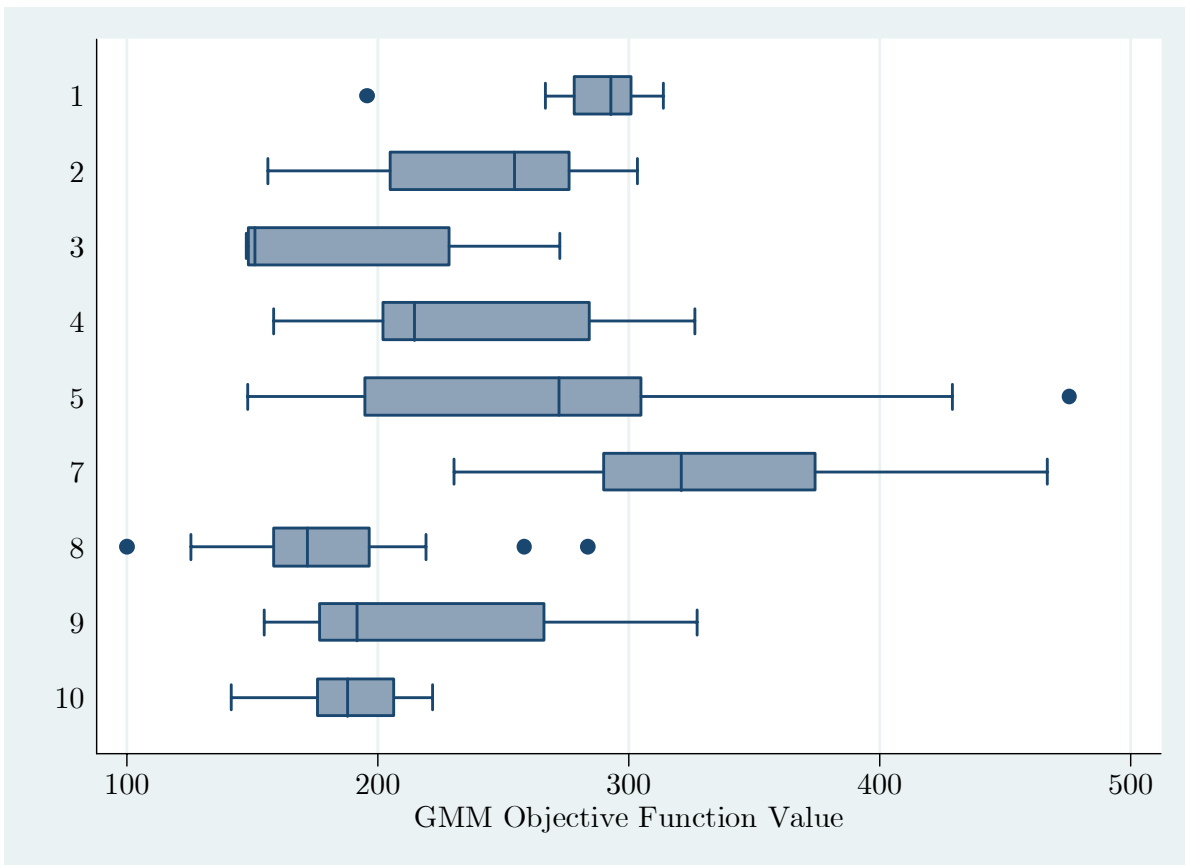
- [12] Burnside, Craig, and Martin Eichenbaum. 1994. "Small Sample Properties of Generalized Method of Moments based Wald Tests." National Bureau of Economic Research Technical Working Paper 155.
- [13] Stock, James H., and Jonathan H. Wright. 2000. "GMM with Weak Identification." *Econometrica*, 68(5): 1055-1096.
- [14] Dorsey, Robert E., and Walter J. Mayer, (1995). "Genetic Algorithms for Estimation Problems with Multiple Optima, Nondifferentiability, and Other Irregular Features." *Journal of Business and Economic Statistics*, 13(1): 53-66.
- [15] Drukker, David M., and Vince Wiggins. 2004. "Verifying the Solution from a Nonlinear Solver: A Case Study: Comment." *American Economic Review*, 94(1): 397-399.
- [16] Goffe, William L., Gary D. Ferrier, and John Rogers. 1994. "Global optimization of statistical functions with simulated annealing." *Journal of Econometrics*, 60(1-2): 65-99.
- [17] Hansen, Lars P., John Heaton, and Amir Yaron. 1996. "Finite-Sample Properties of Some Alternative GMM estimators." *Journal of Business and Economic Statistics*, 14(3): 262-280.
- [18] Hansen, Lars P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, 50(4):1029-1054.
- [19] Hendel, Igal and Aviv Nevo. 2006. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica*, 74(6): 1637-1673.
- [20] Hu, Wei-Min and Christopher R. Knittel. 2007. "Durable Hardware Demand in the Presence of Changing Software." Manuscript, University of California, Davis.
- [21] Gowrisankaran, Gautam and Marc Rysman. 2007. "Dynamics of Consumer Demand for New Durable Goods." mimeo.
- [22] Judd, Kenneth L. 1998. *Numerical Methods in Economics*. Cambridge: MIT Press.
- [23] Kappel, Franz, and Alexei Kuntsevich. 2000. "An Implementation of Shor's r-Algorithm." *Computational Optimization and Applications*, 15(2): 193-205.

- [24] Lagarias, Jeffrey C., James E. Reeds, Margaret H. Wright, and Paul E. Wright. 1998. "Convergence properties of the Nelder-Mead Simplex Method in Low Dimensions." *Siam Journal of Optimization*, 9(1): 112-147.
- [25] McCullough, B.D., and H.D. Vinod. 2003. "Verifying the Solution from a Nonlinear Solver: A Case Study: Reply." *American Economic Review*. 93(3): 873-892.
- [26] McFadden, Daniel. 1981. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data*, ed. C.F Manski and D. McFadden, 198-272. Cambridge, MIT Press.
- [27] Melnikov, Oleg. 2000. "Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market." Manuscript, Yale University.
- [28] Miranda, Mario J., and Paul L. Fackler. 2002. *Applied Computational Economics and Finance*. Cambridge: MIT press.
- [29] Nevo, Aviv. 1997. "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry." University of California, Berkeley Competition Policy Center Working Paper no. CPC 99-02.
- [30] Nevo, Aviv. 2000a. "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand." *Journal of Economics & Management Strategy*, 9(4): 513-548.
- [31] Nevo, Aviv. 2000b. "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry." *RAND Journal of Economics*, 31(3): 395-421.
- [32] Nevo, Aviv. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, 69(2): 307-342.
- [33] Nevo, Aviv. 2003. "New Products, Quality Changes, and Welfare Measures from Estimated Demand Systems." *Review of Economics and Statistics*, 85(2): 266-275.
- [34] Pagan, Adrian R., and J.C. Robertson. "GMM and its Problems." Manuscript, Australia National University.



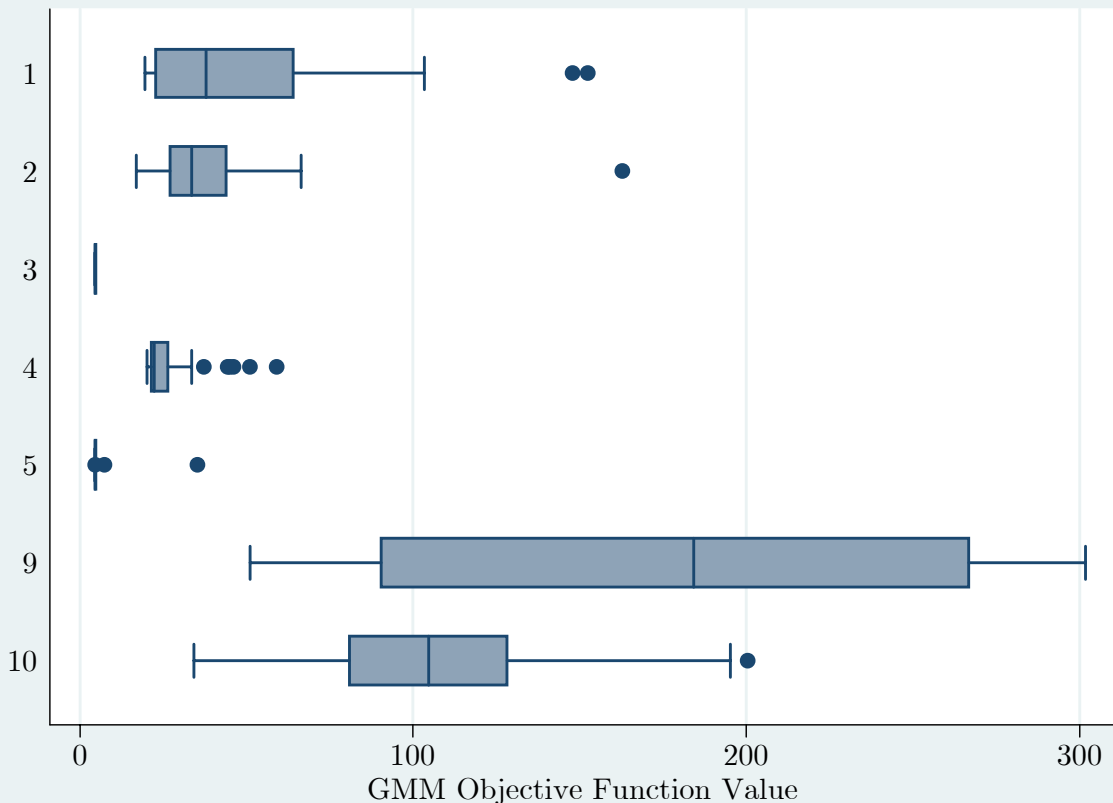
- [35] Petrin, Amil. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy*, 110(4): 705-729.
- [36] Shachar, Ron, and Barry Nalebuff. 2004. "Verifying the Solution from a Nonlinear Solver: A Case Study: Comment." *American Economic Review*, 94(1): 382-390.
- [37] Small, Kenneth, A., and Harvey S. Rosen. 1981. "Applied Welfare Economics with Discrete Choice Models." *Econometrica*. 49(1): 105-130.
- [38] Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65(3), 557-586.
- [39] Stock, James H., and Jonathan H. Wright. 2000. "GMM with Weak Identification." *Econometrica*, 68(5): 1055-1096.
- [40] Stock, James H., Jonathan H. Wright and Motohiro 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.
- [41] Torczon, Virginia. 1997. "On the Convergence of Pattern Search Algorithms." *Siam Journal of Optimization*, 7(1): 1-25.
- [42] Venkataraman, P. 2002. *Applied Optimization with Matlab Programming*. New York: Wiley.
- [43] Yang, Won Young, Wenwu Cao, Tae-Sang Chung, and John Morris. 2005. *Applied Numerical Methods using Matlab*. New York: Wiley

## B Figures



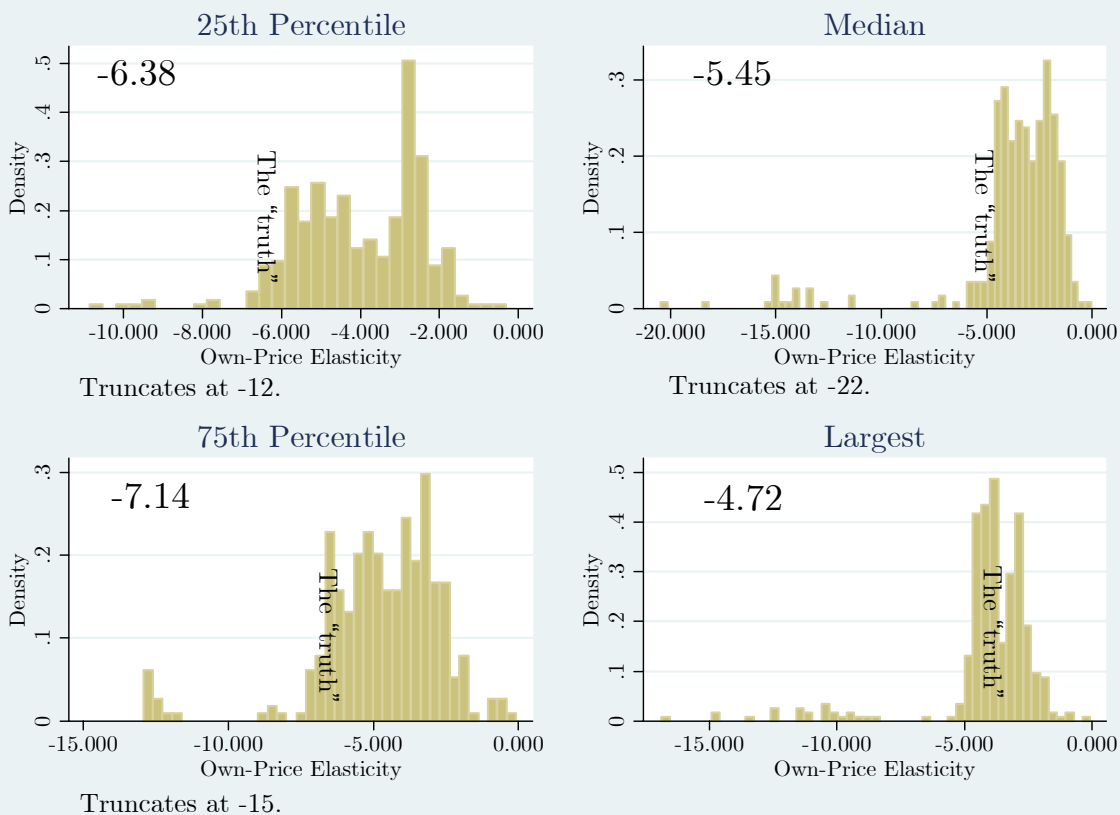
**Figure 1:** GMM objective values for converged algorithms using the automobile data. This truncates the upper 10% of the converged GMM objective values.

The box represents the 25th and 75th percentiles with a median line. Whiskers extend from the box to the upper and lower adjacent values and are capped with an adjacent line. The upper adjacent value is the largest data value that is less than or equal to the third quartile plus  $1.5 \times \text{IQR}$  and the lower adjacent value is the smallest data value that is greater than or equal to the first quartile minus  $1.5 \times \text{IQR}$ . Dots represent values outside these “adjacent values”.



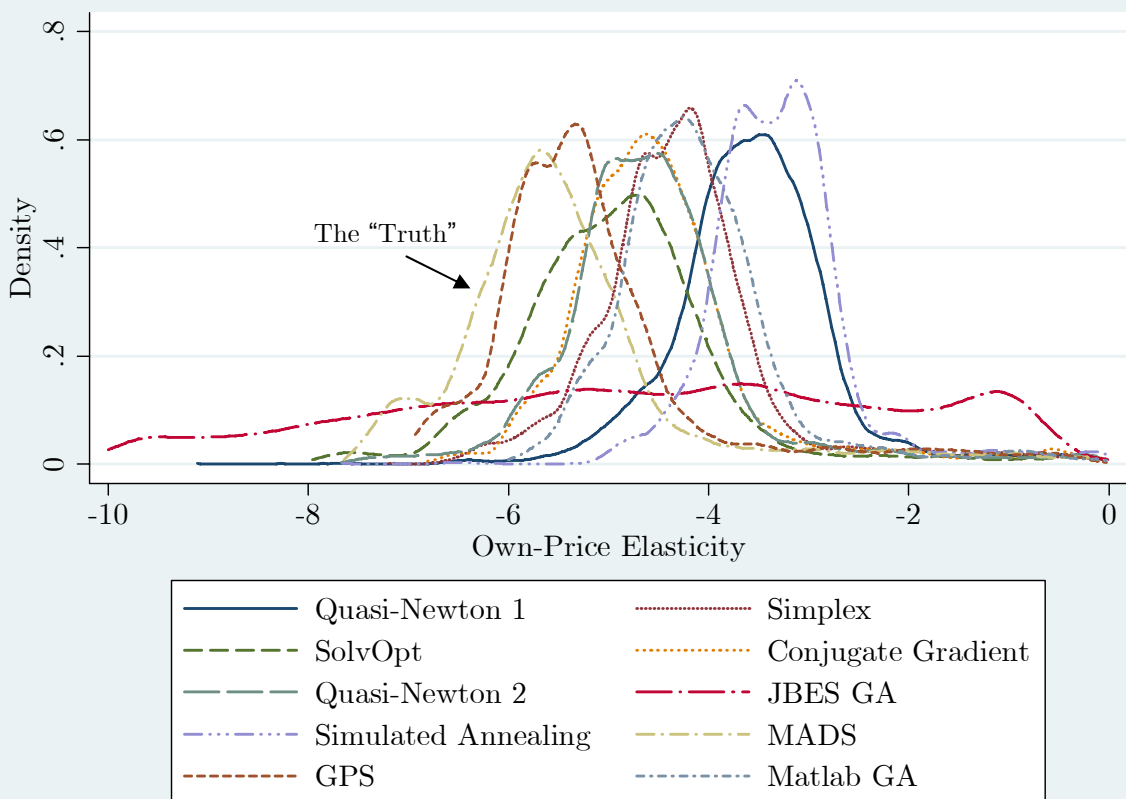
**Figure 2:** GMM objective values for converged algorithms using the cereal data. This truncates the upper 25% of the converged GMM objective values. Algorithms 6 (JBES Genetic Algorithm) and 7 (Simulated Annealing) never converge. Algorithm 8 (MADS) converges twice, but at objective values above the 75<sup>th</sup> percentile.

The box represents the 25th and 75th percentiles with a median line. Whiskers extend from the box to the upper and lower adjacent values and are capped with an adjacent line. The upper adjacent value is the largest data value that is less than or equal to the third quartile plus 1.5 X IQR and the lower adjacent value is the smallest data value that is greater than or equal to the first quartile minus 1.5 X IQR. Dots represent values outside these “adjacent values”.

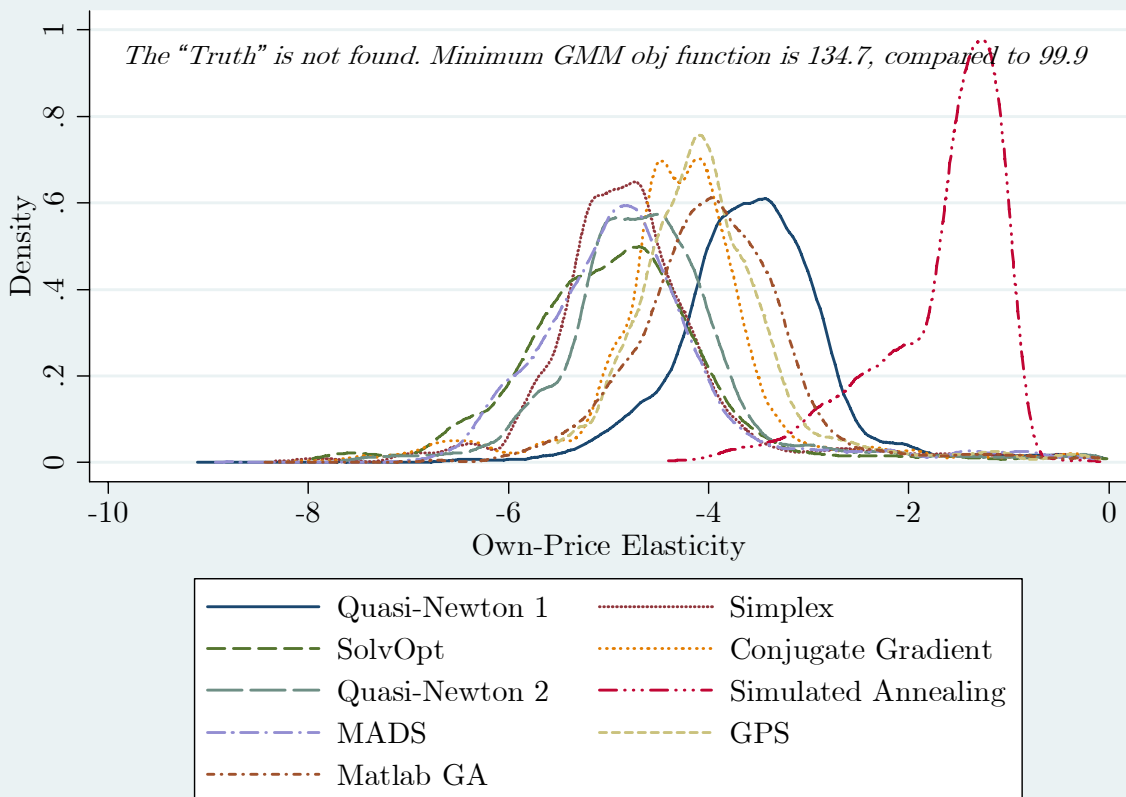


**Figure 3:** Histogram of candidate set of own-price elasticities for four products for the automobile data. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.

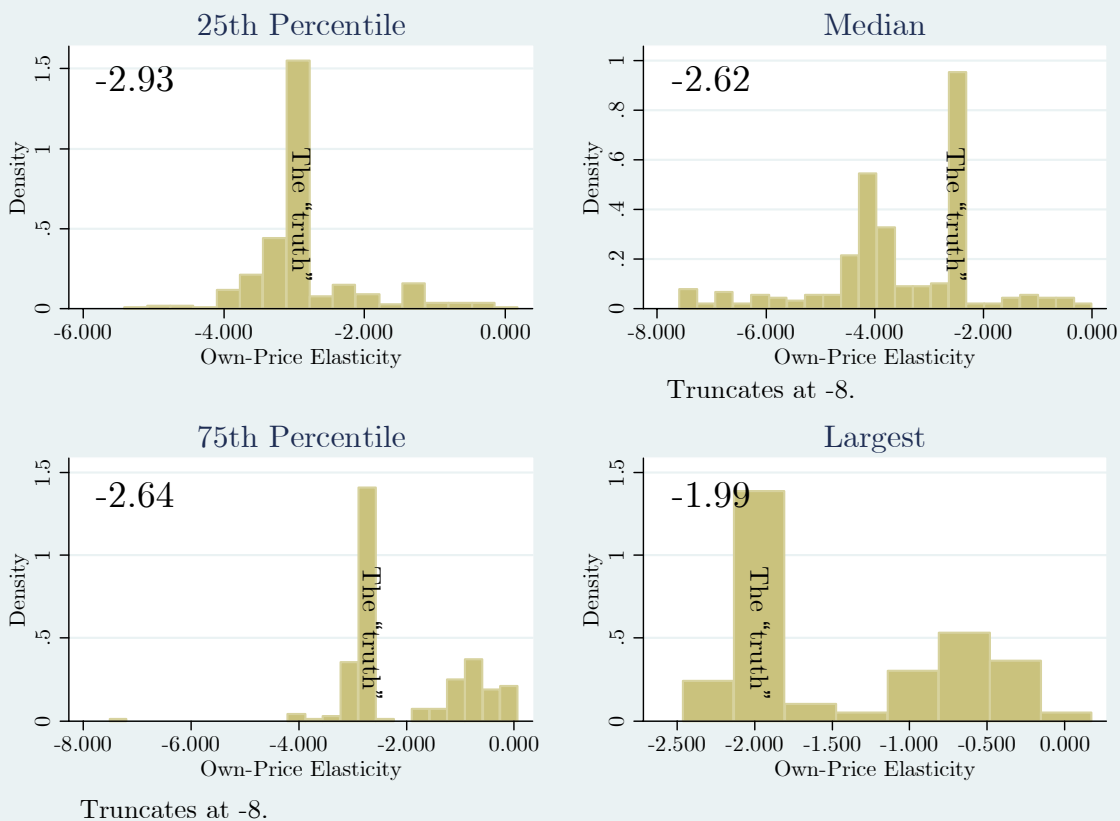
The average within-product-market standard deviation in the estimated elasticities across all converged results is 4.25; the median is 3.08. The average own-price elasticity is -3.66. Among the “best” sets of parameter values, the mean standard deviation is 0.84, the median is 0.81, and the mean own-price elasticity is -4.11.



**Figure 4:** Density estimates of the own-price elasticities across all products in the automobile data, using the “best” set of parameters for each algorithm.

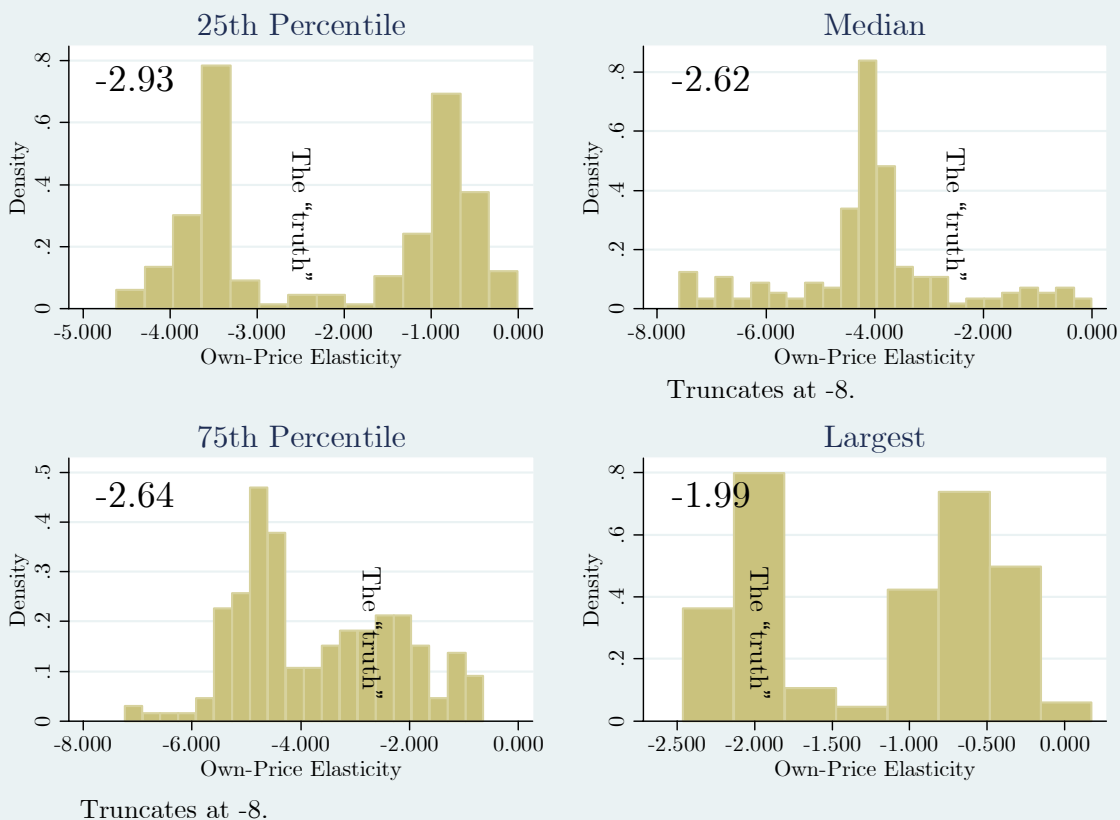


**Figure 5:** Density estimates of the own-price elasticities across all products in the automobile data, using the “best” set of parameters for each algorithm. This assumes the research tries 20 starting values, instead of 50.



**Figure 6:** Histogram of candidate set of own-price elasticities for four products for the cereal data. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.

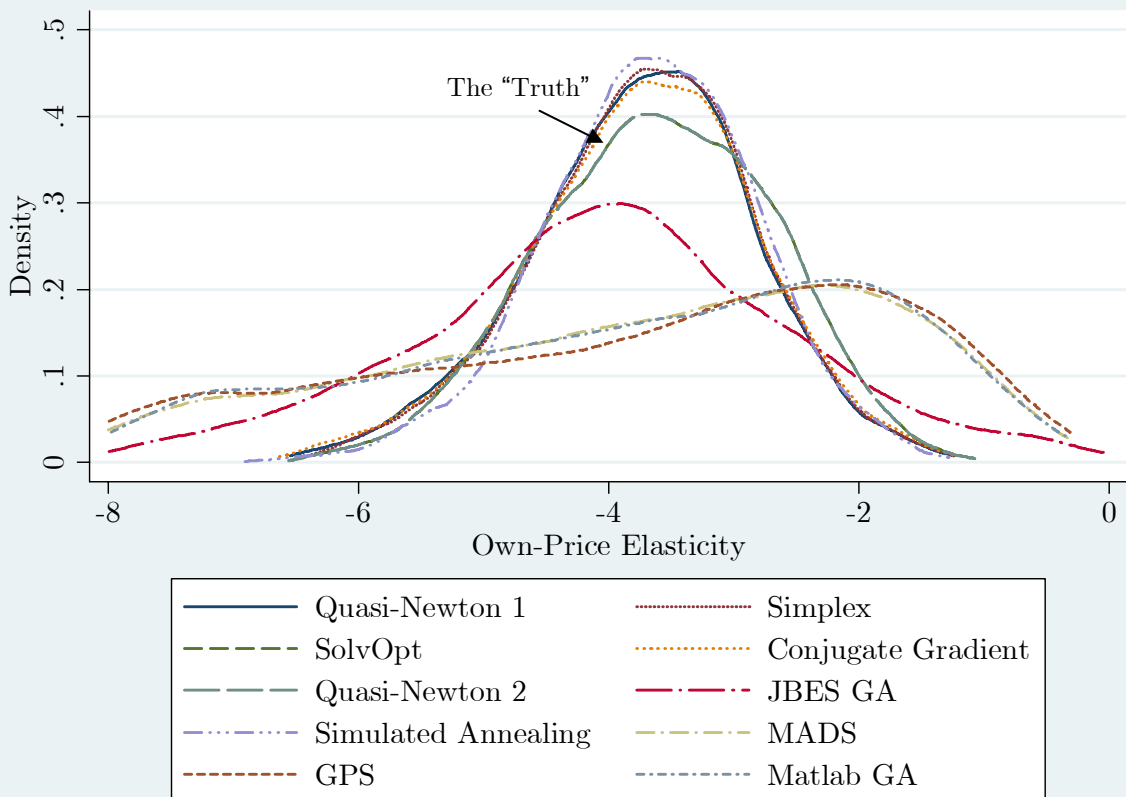
The average within-product-market standard deviation in the estimated elasticities across all converged results is 4.42; the median is 1.88. The average own-price elasticity is -4.92. Among the “best” sets of parameter values, the mean standard deviation is 2.66, the median is 1.17, and the mean own-price elasticity is -4.86.



**Figure 7:** Histogram of candidate set of own-price elasticities for four products for the cereal data without using algorithms 3 and 5. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.

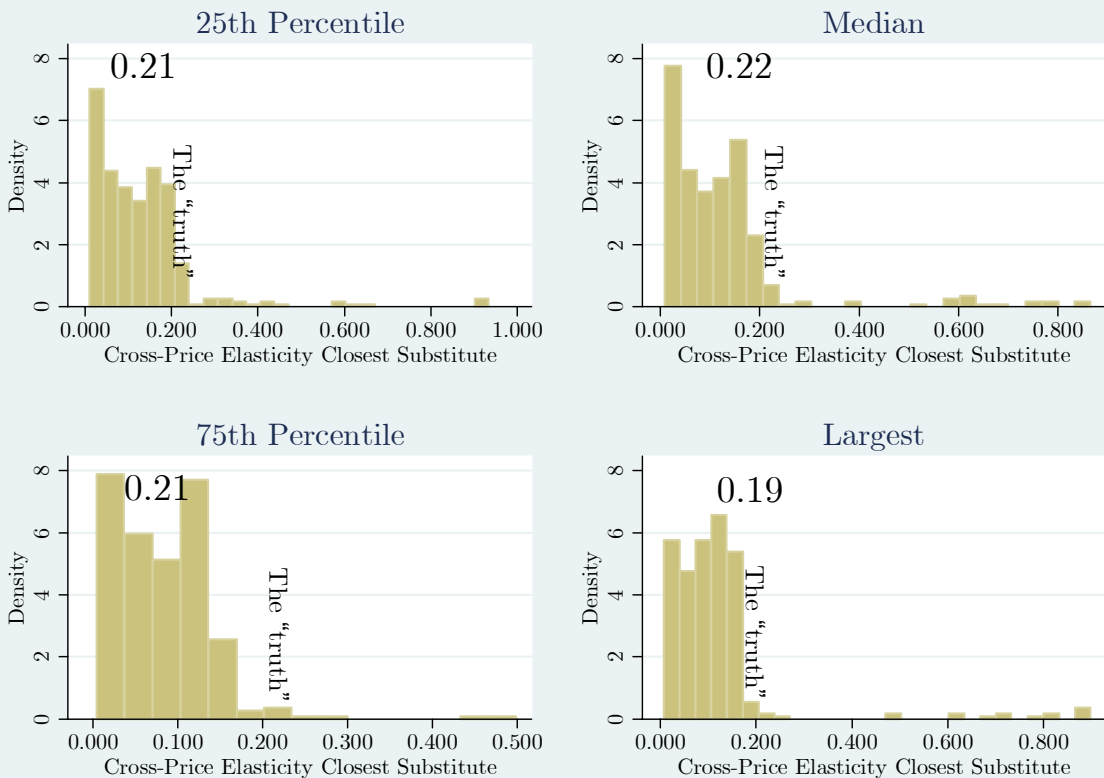
The average within-product-market standard deviation in the estimated elasticities across all converged results is 4.42; the median is 1.88. The average own-price elasticity is -4.92. Among the “best” sets of parameter values, the mean standard deviation is 2.66, the median is 1.17, and the mean own-price elasticity is -4.86.



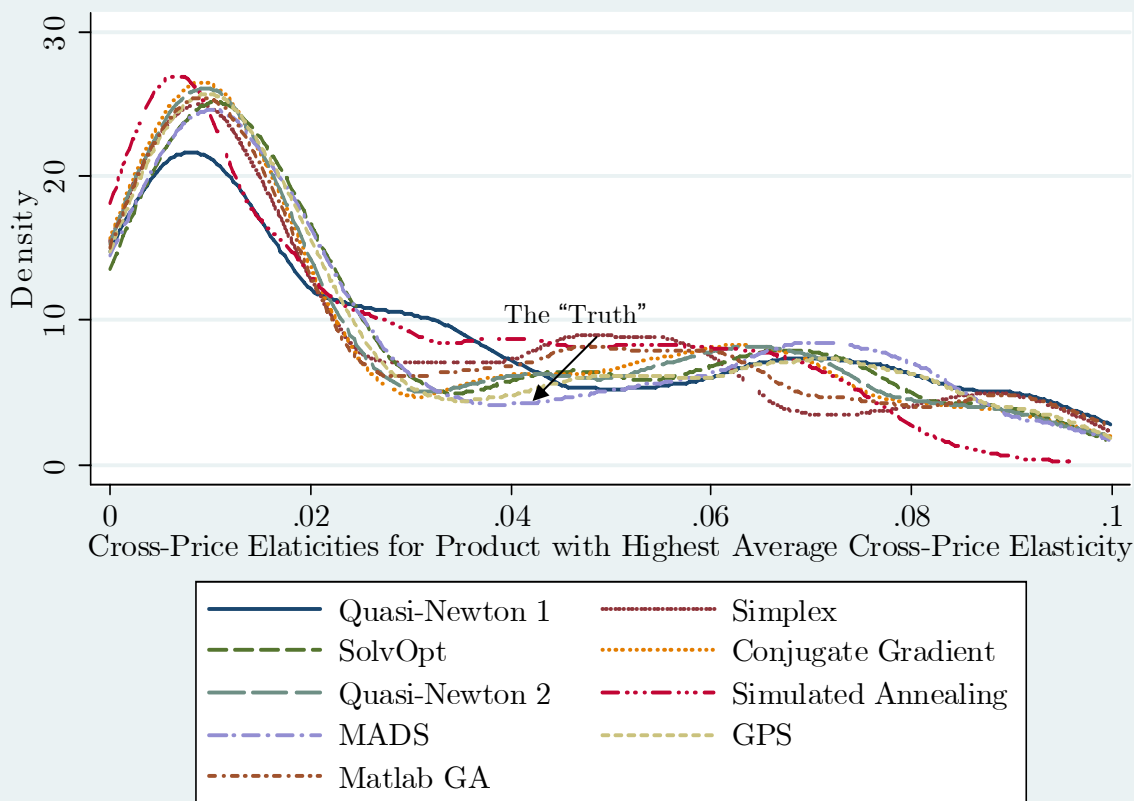


**Figure 8:** Density estimates of the own-price elasticities across all products in the cereal data, using the “best” set of parameters for each algorithm.

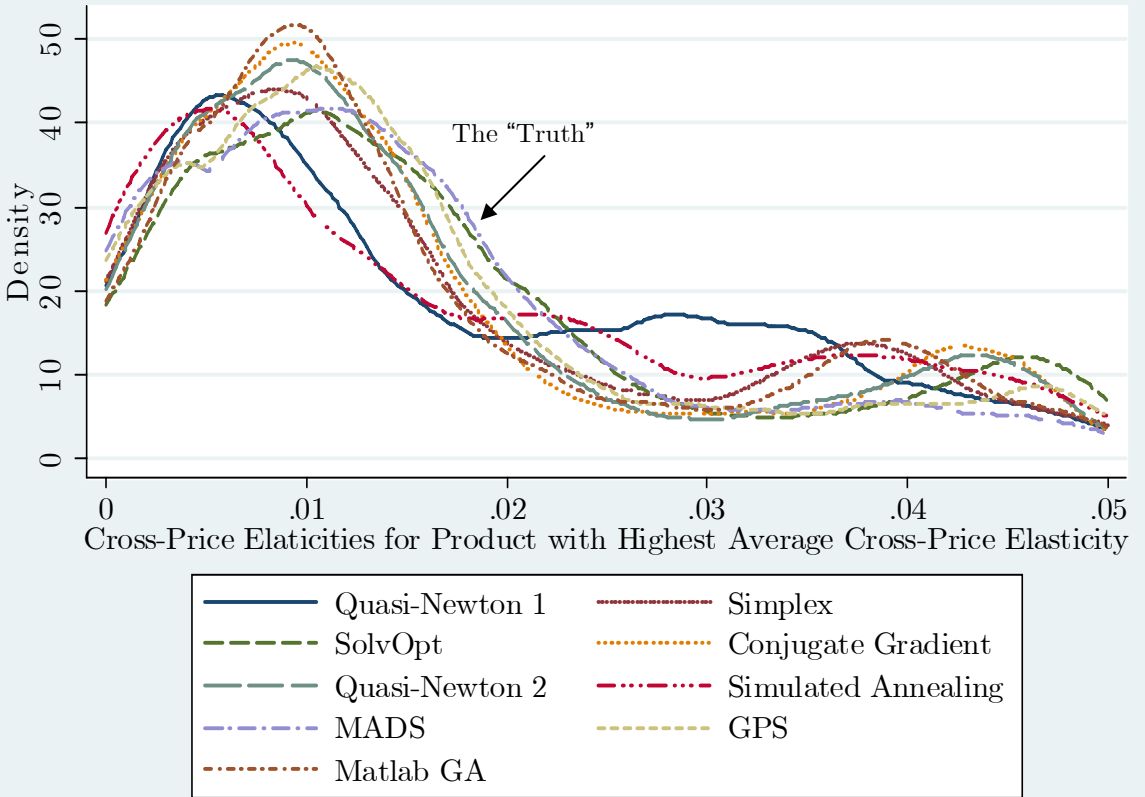
For the JBES GA and Simulated Annealing algorithms, the best set of parameters do not meet the convergence criteria.



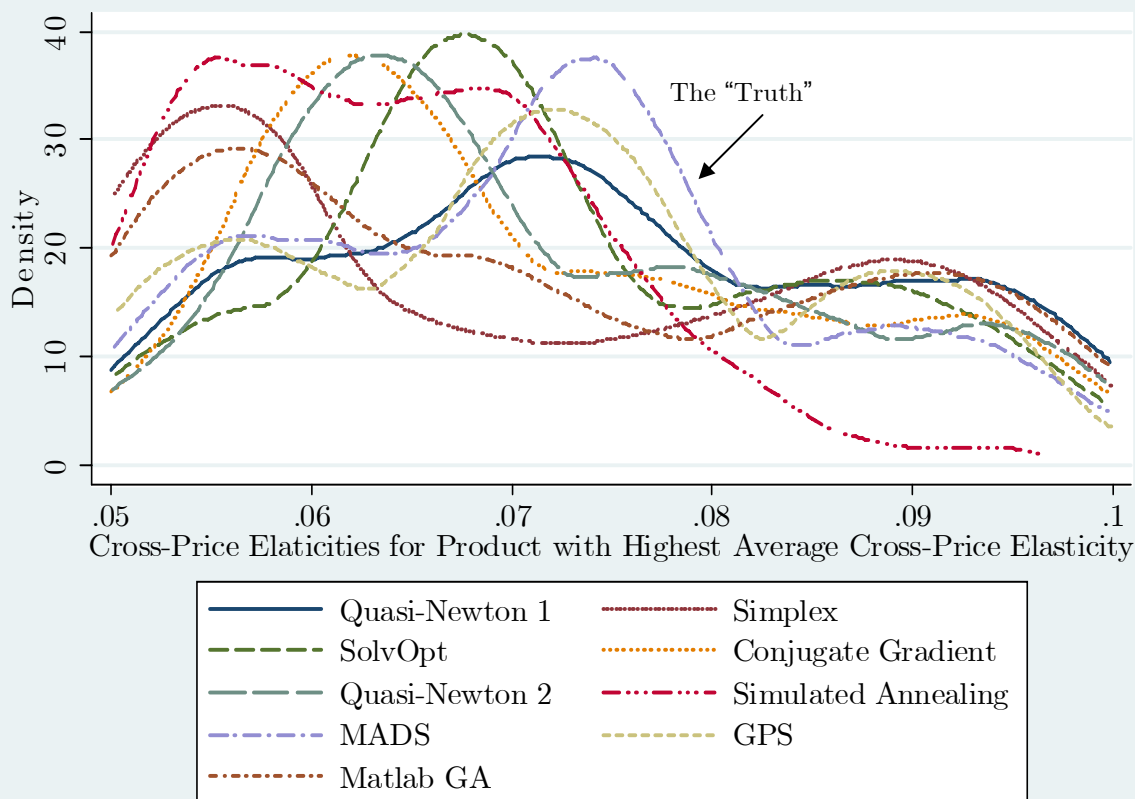
**Figure 9:** Histogram of candidate set of cross-price elasticities for four products in the automobile data. The cross product is chosen as the product that is the closest substitute. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value. All graphs are truncated above by 1.00.



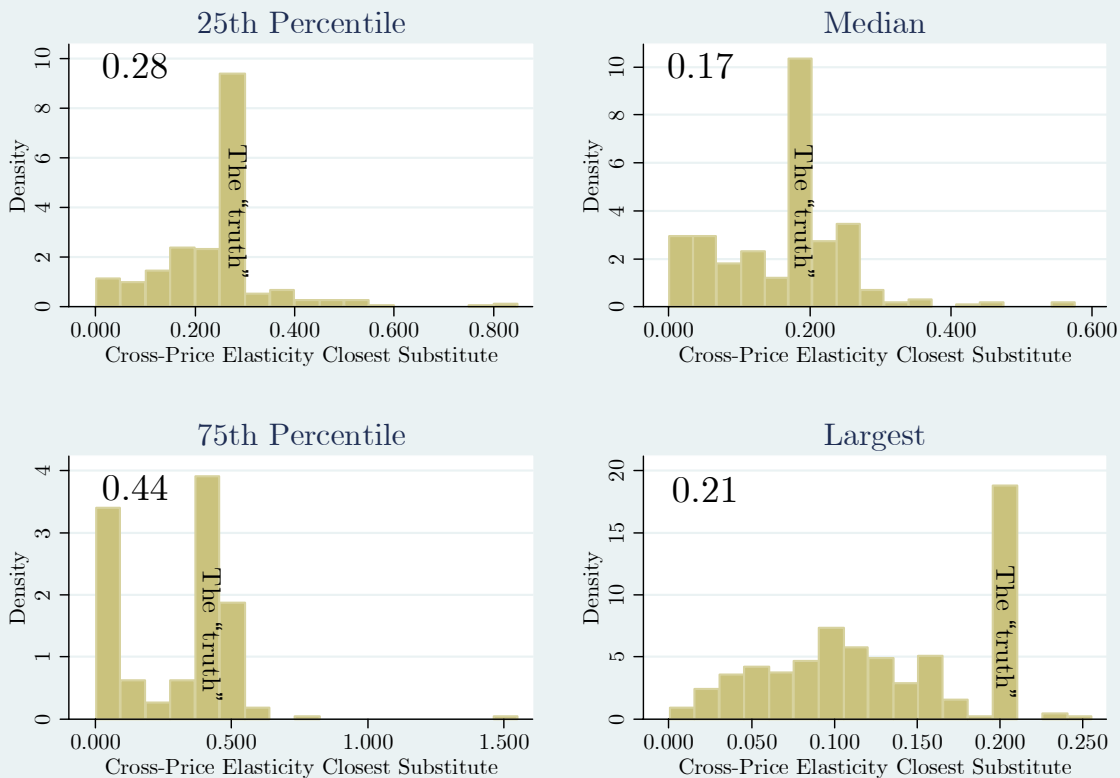
**Figure 10:** Density estimates of the cross-price elasticities across all products in the automobile data for the product that has the highest average cross-price elasticity. These results use the “best” set of parameters for each algorithm and truncate at zero and above at .1.



**Figure 11:** Density estimates of the cross-price elasticities across all products in the automobile data for the product that has the highest average cross-price elasticity focusing on the elasticities ranging from 0 to .05. These results use the “best” set of parameters for each algorithm.

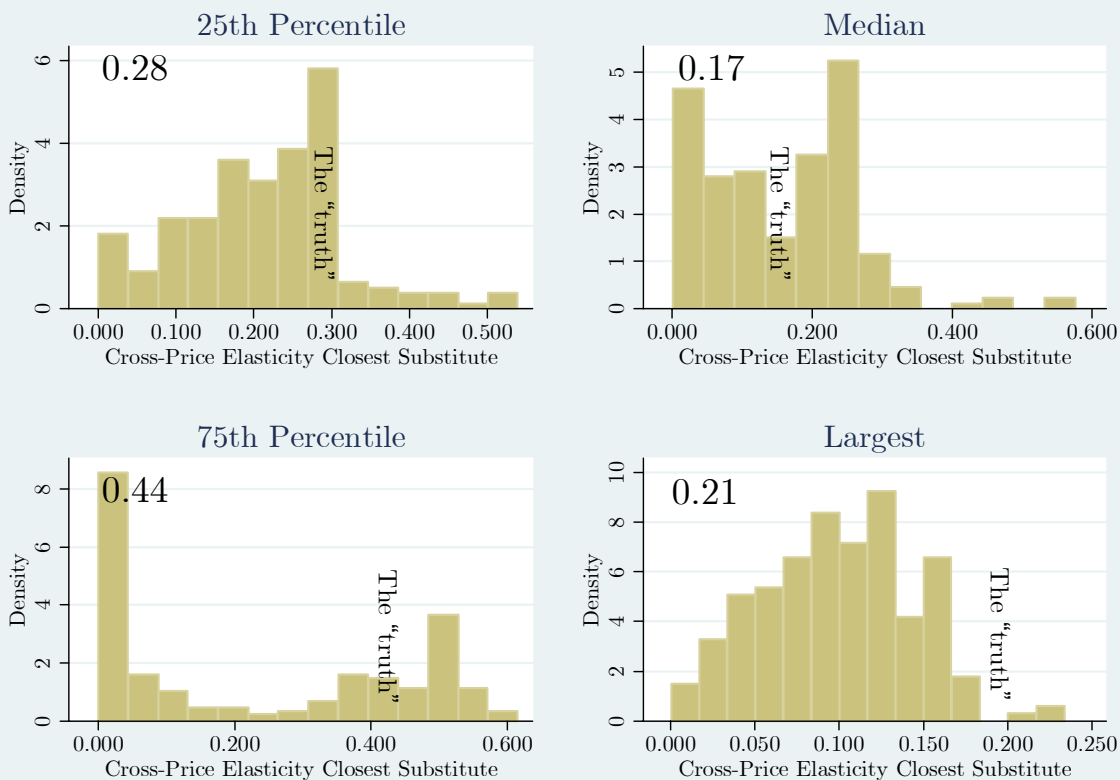


**Figure 12:** Density estimates of the cross-price elasticities across all products in the automobile data for the product that has the highest average cross-price Elasticity focusing on the elasticities ranging from .05 to .1. These results use the “best” set of parameters for each algorithm.

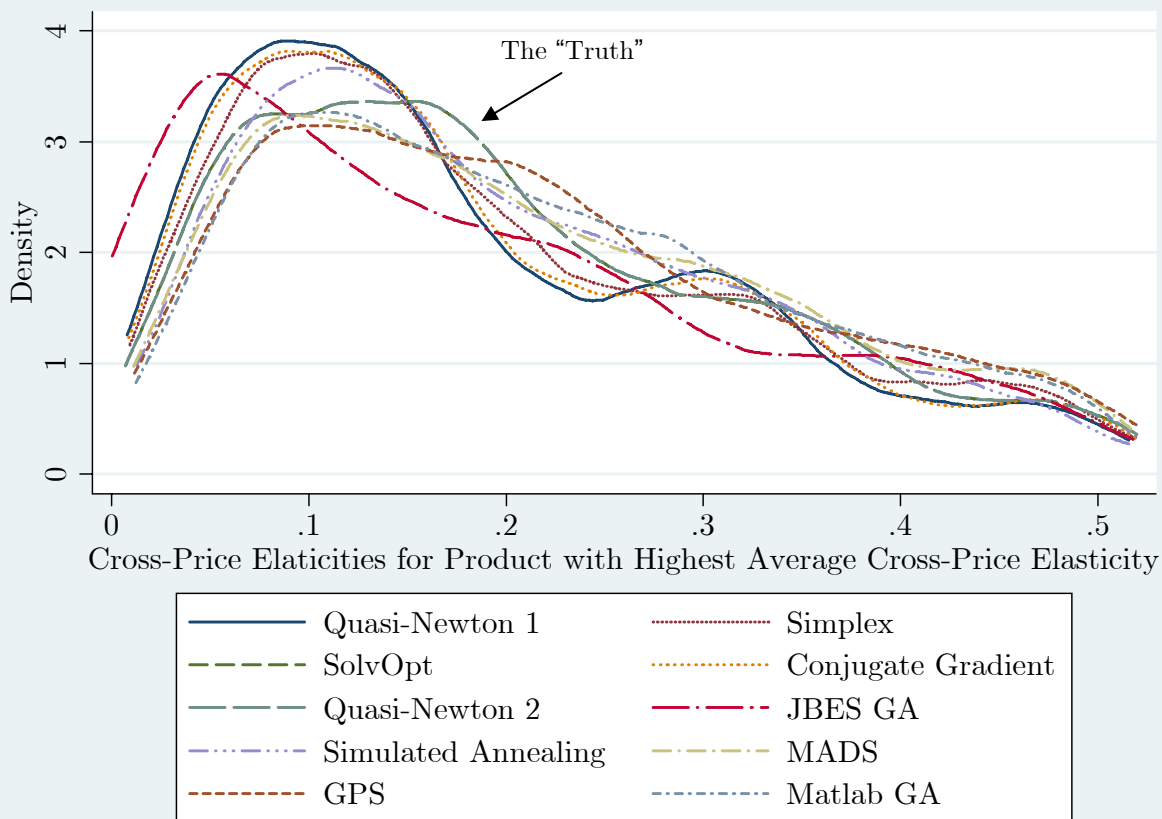


**Figure 13:** Histogram of candidate set of cross-price elasticities for four products in the cereal data. The cross product is chosen as the product that is the closest substitute. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.

These graphs truncate the larges 10 percent.

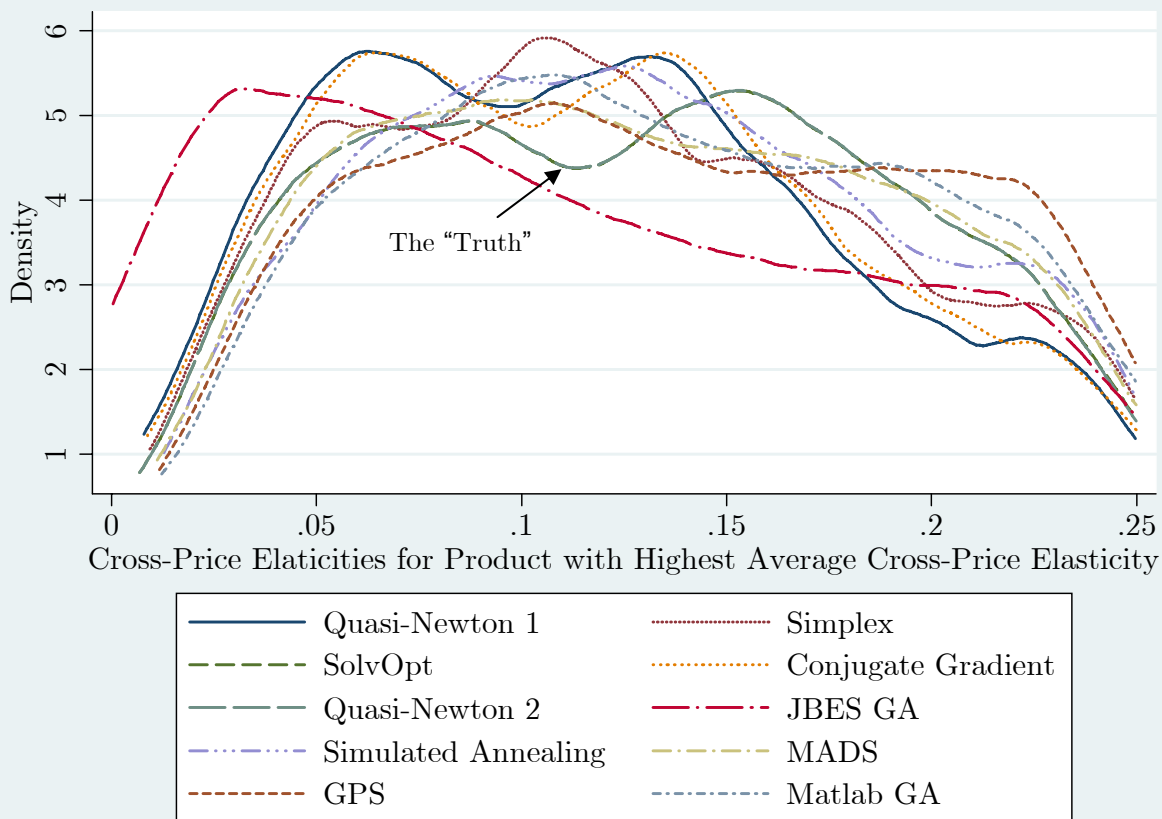


**Figure 14:** Histogram of candidate set of cross-price elasticities for four products in the cereal data without using algorithms 3 and 5. The cross product is chosen as the product that is the closest substitute. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.

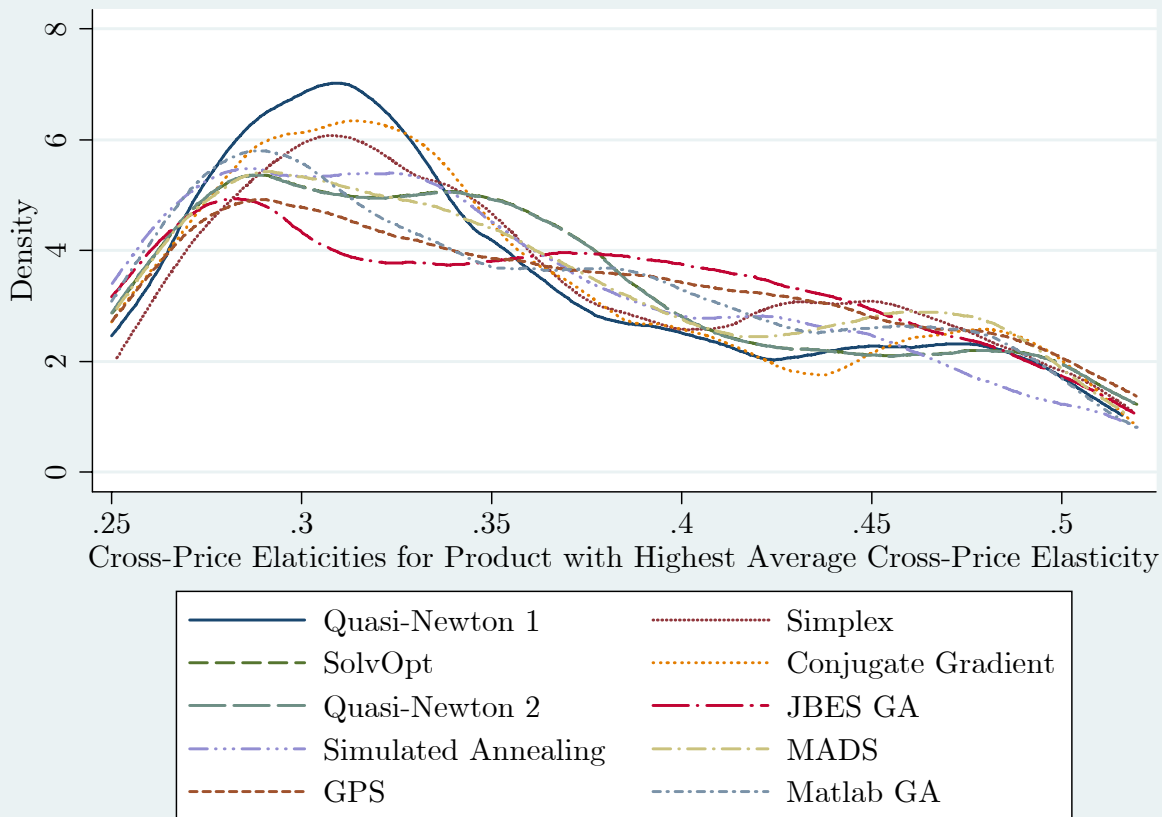


**Figure 15:** Density estimates of the cross-price elasticities across all products in the cereal data for the product that has the highest average cross-price elasticity. These results use the “best” set of parameters for each algorithm and truncate at zero and above the 90<sup>th</sup> percentile.

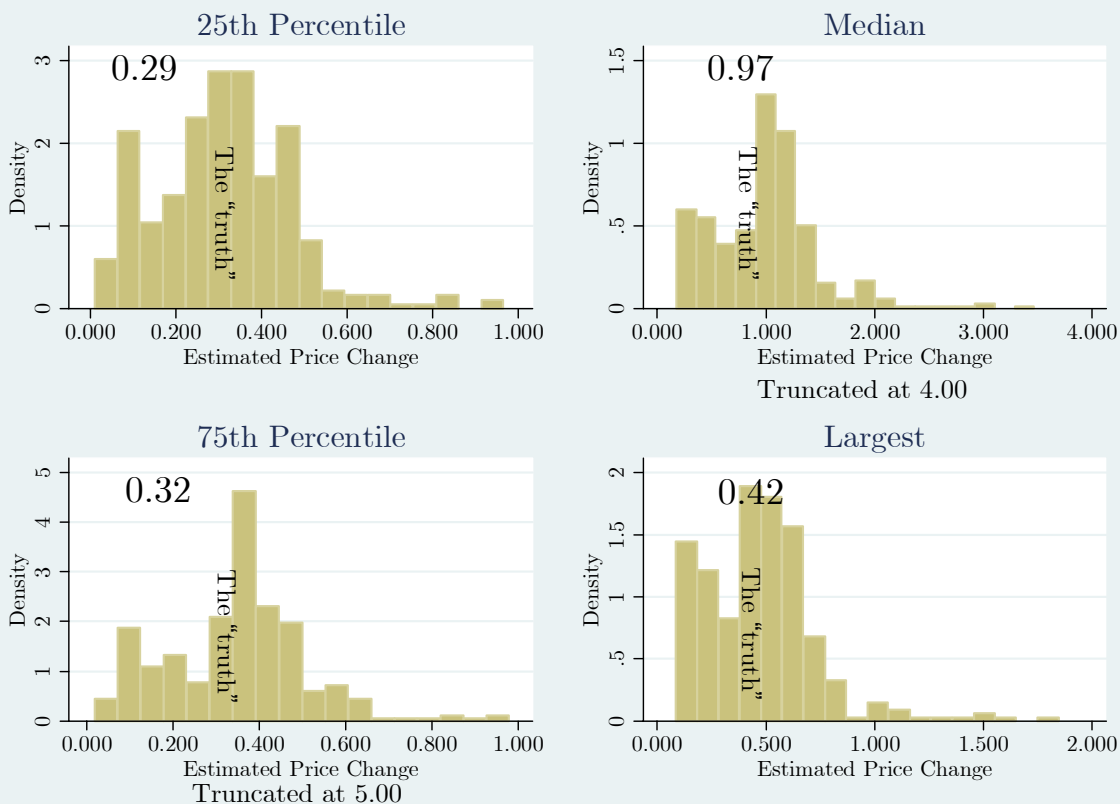




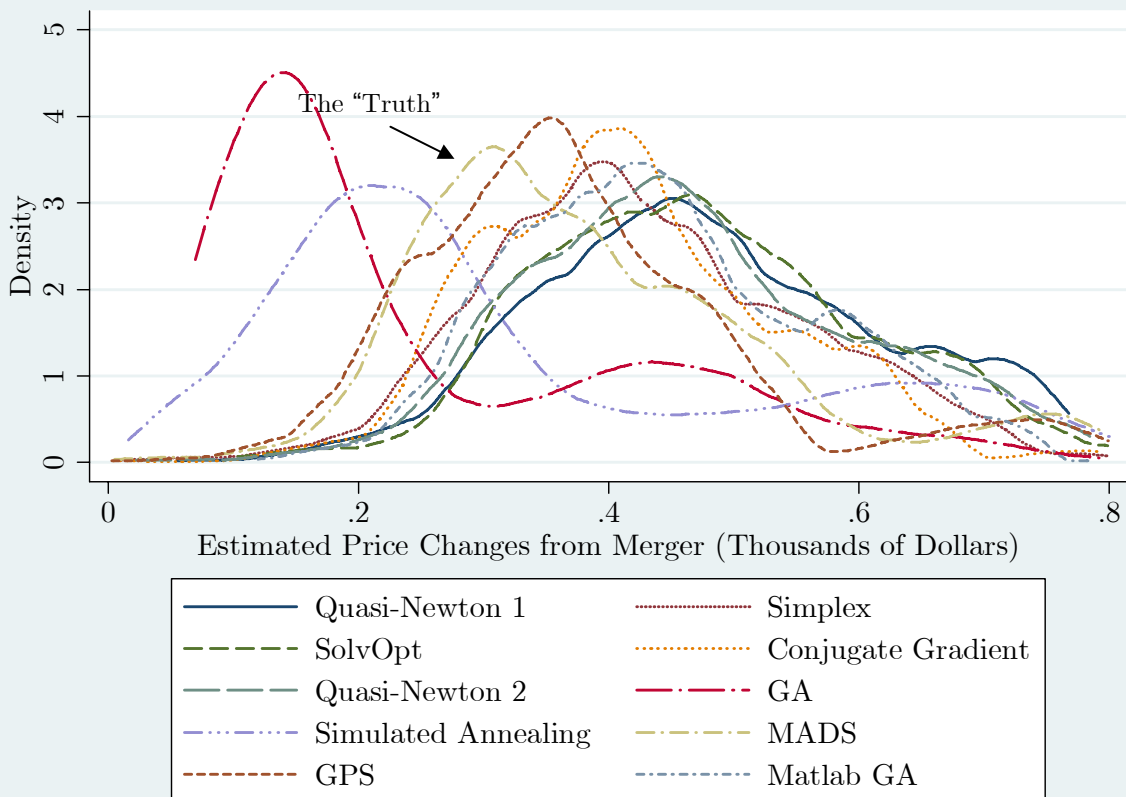
**Figure 16:** Density estimates of the cross-price elasticities across all products in the cereal data for the product that has the highest average cross-price elasticity focusing on the elasticities ranging from 0 to .25. These results use the “best” set of parameters for each algorithm.



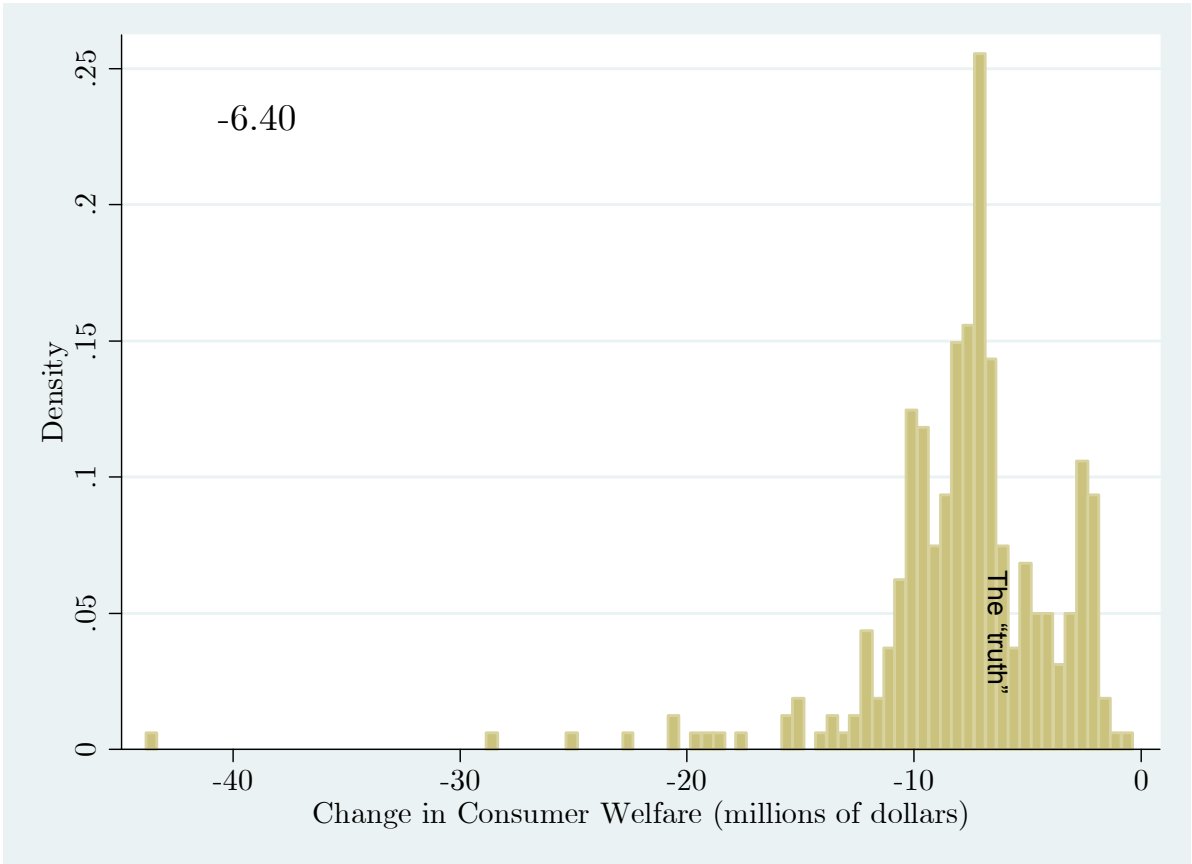
**Figure 17:** Density estimates of the cross-price elasticities across all products in the cereal data for the product that has the highest average cross-price elasticity focusing on the elasticities ranging from .25 to .52. These results use the “best” set of parameters for each algorithm.



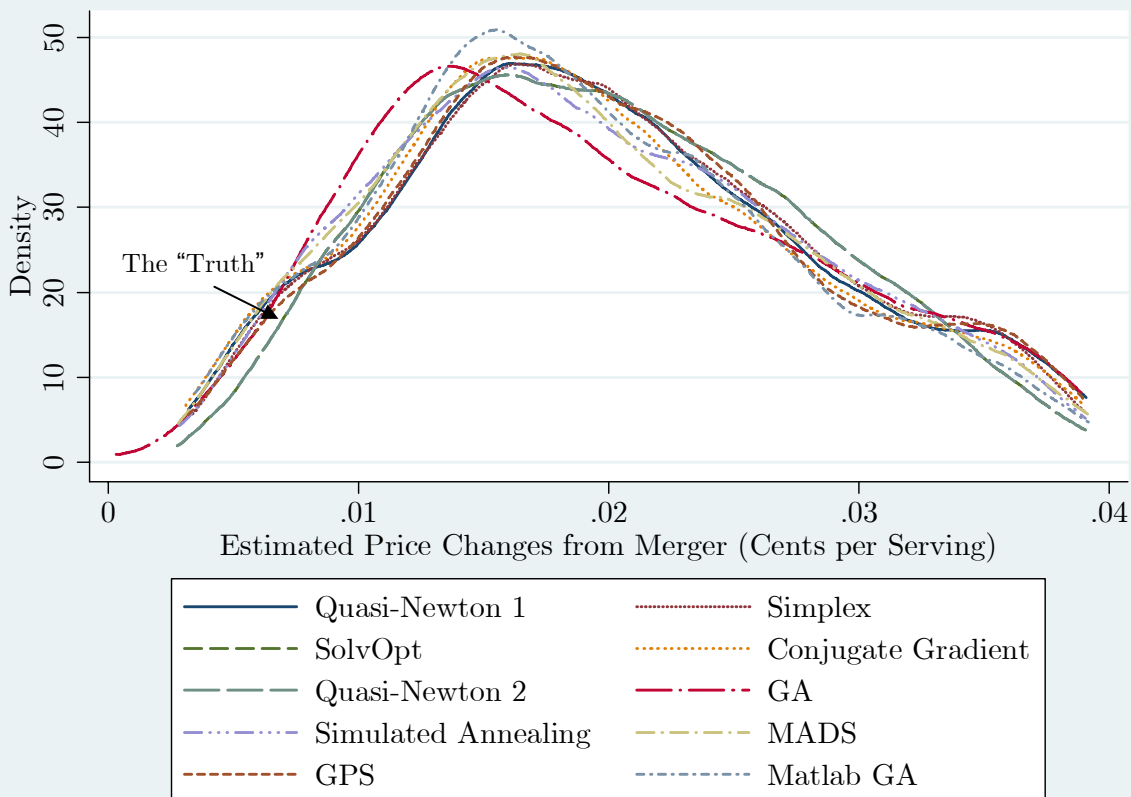
**Figure 18:** Histogram of candidate set of price changes for four products in the automobile data. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.



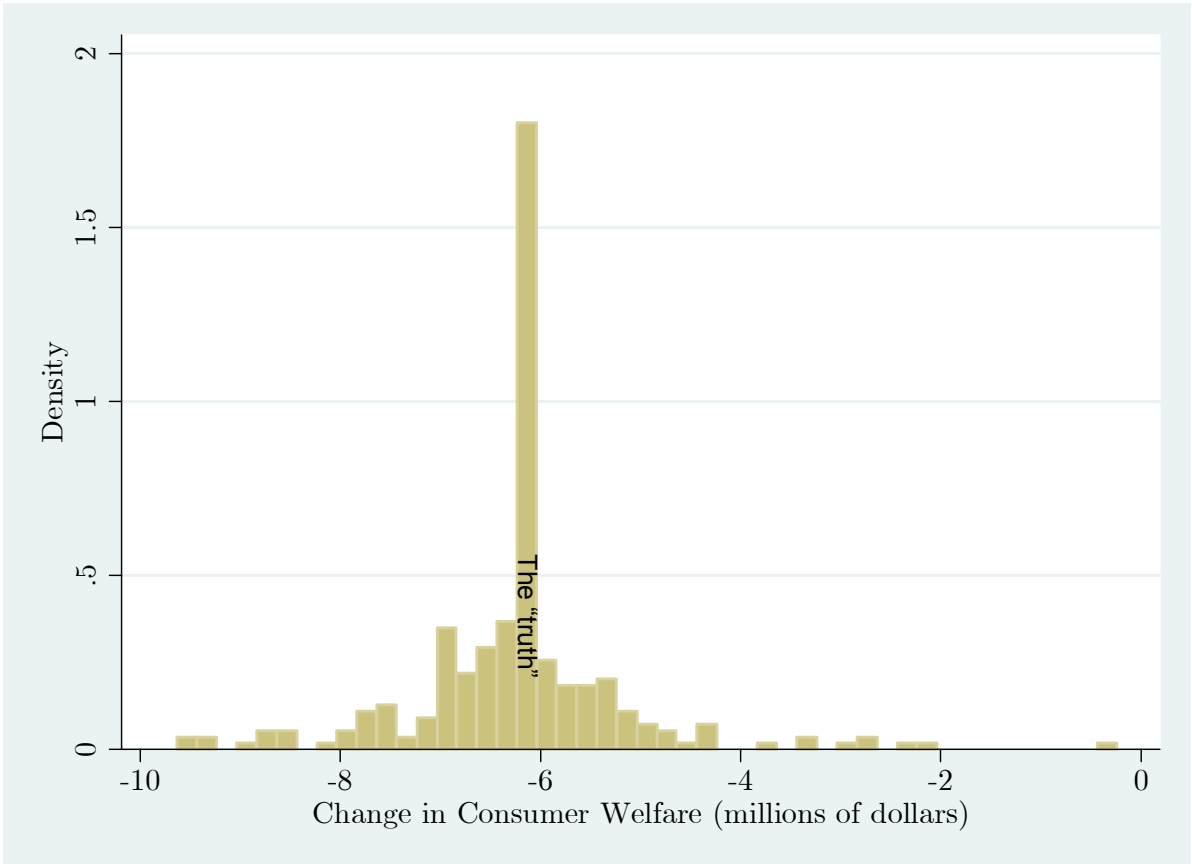
**Figure 19:** Density estimates of the estimated price changes following the merger for the automobile data. These results use the “best” set of parameters for each algorithm and truncate at zero and above .80 percentile (roughly the 90<sup>th</sup> percentile).



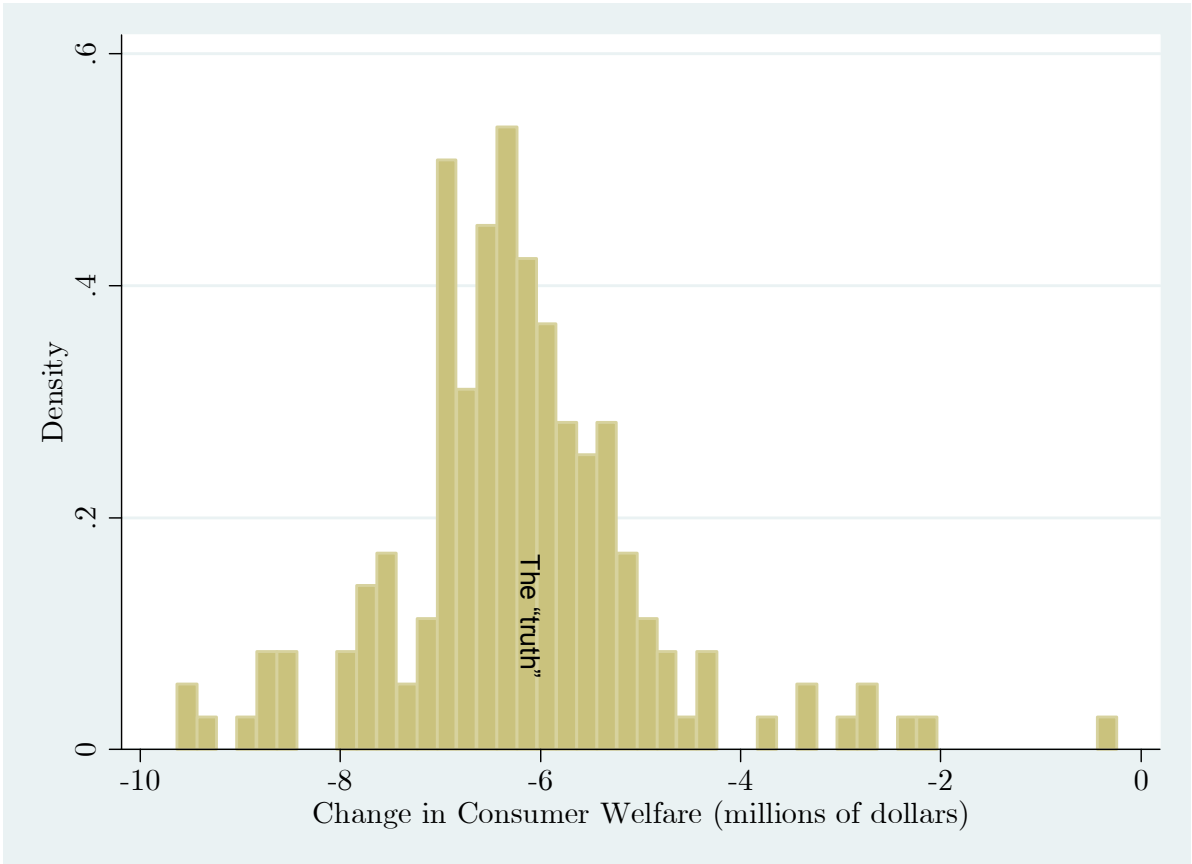
**Figure 20:** Histogram of estimated change in total welfare from hypothetical merger using the automobile data. The range across eight “best” estimates is -4.67 to -9.94.



**Figure 21:** Density estimates of the estimated price changes following the merger for the cereal data. These results use the “best” set of parameters for each algorithm and truncate below the 5<sup>th</sup> and above the 95<sup>th</sup> percentile.

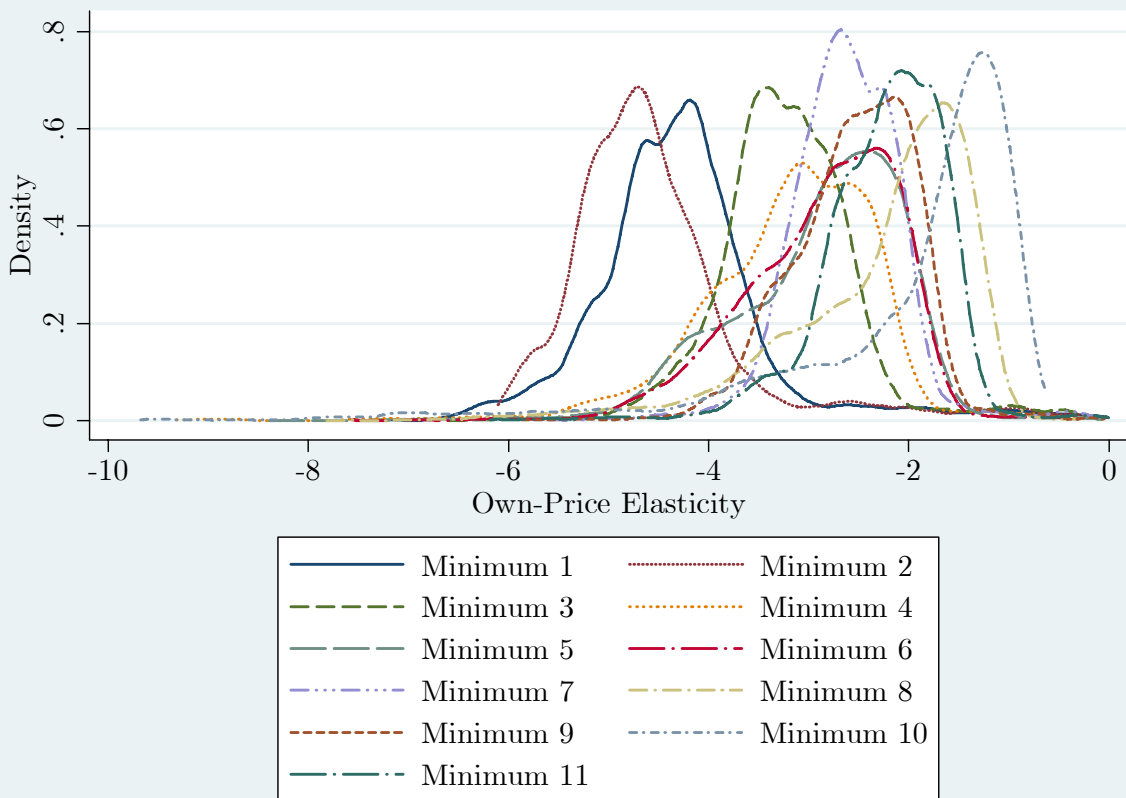


**Figure 22:** Histogram of estimated change in total welfare from hypothetical merger using the cereal data. The range across eight “best” estimates is -6.16 to -7.05. This truncates the results at -10.

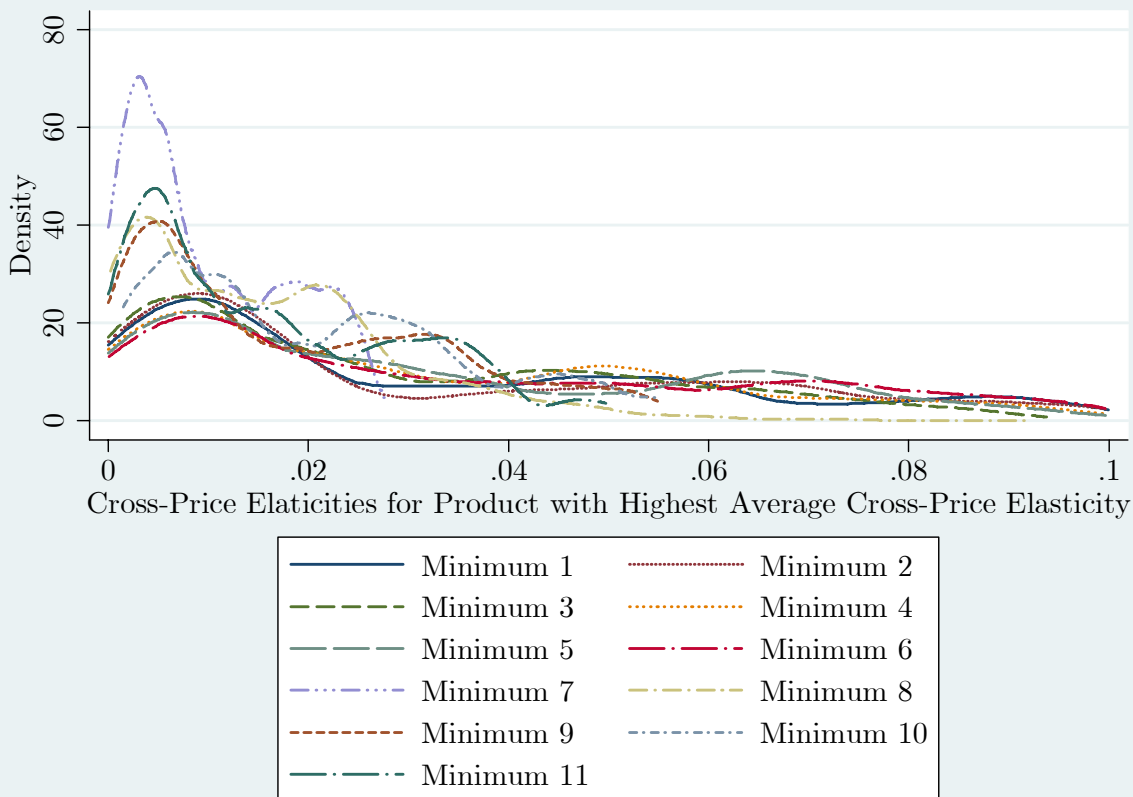


**Figure 23:** Histogram of estimated change in total welfare from hypothetical merger using the cereal data. The range across eight “best” estimates is -6.16 to -7.05. This truncates the results at -10.

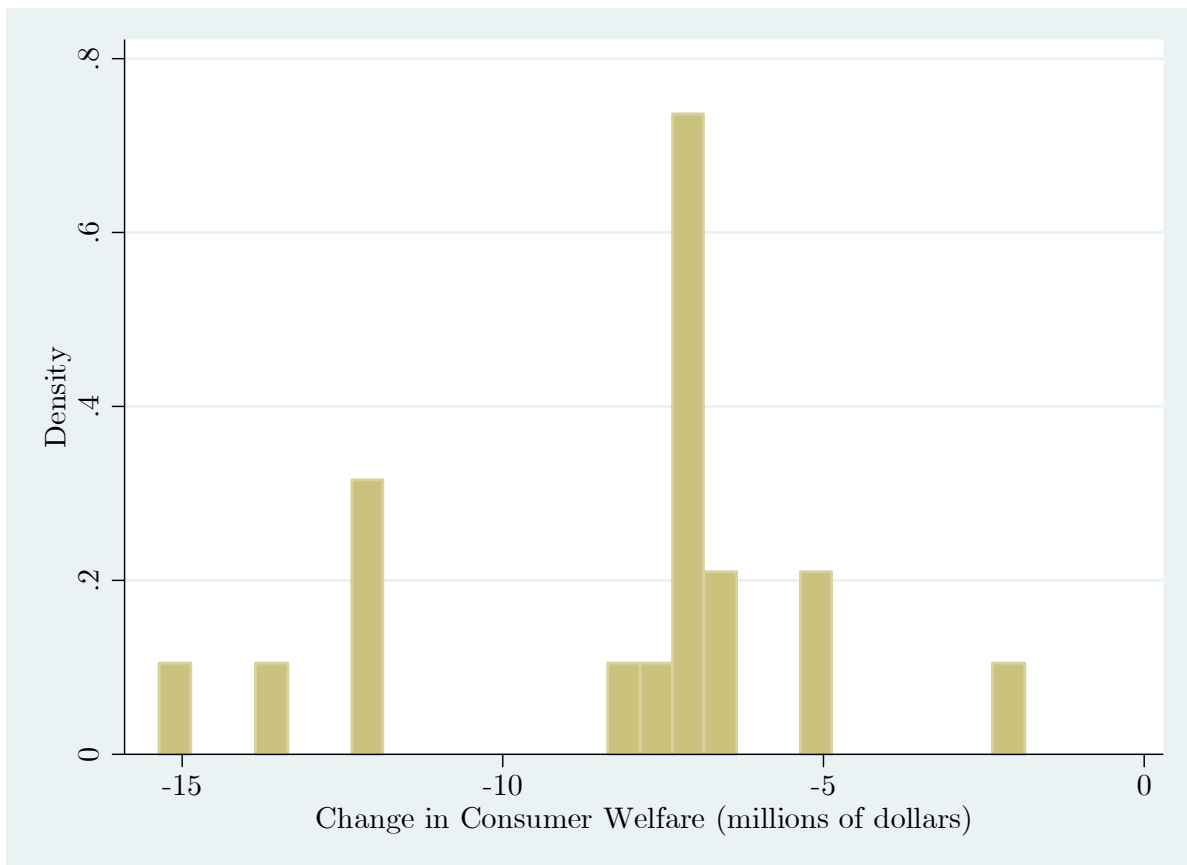




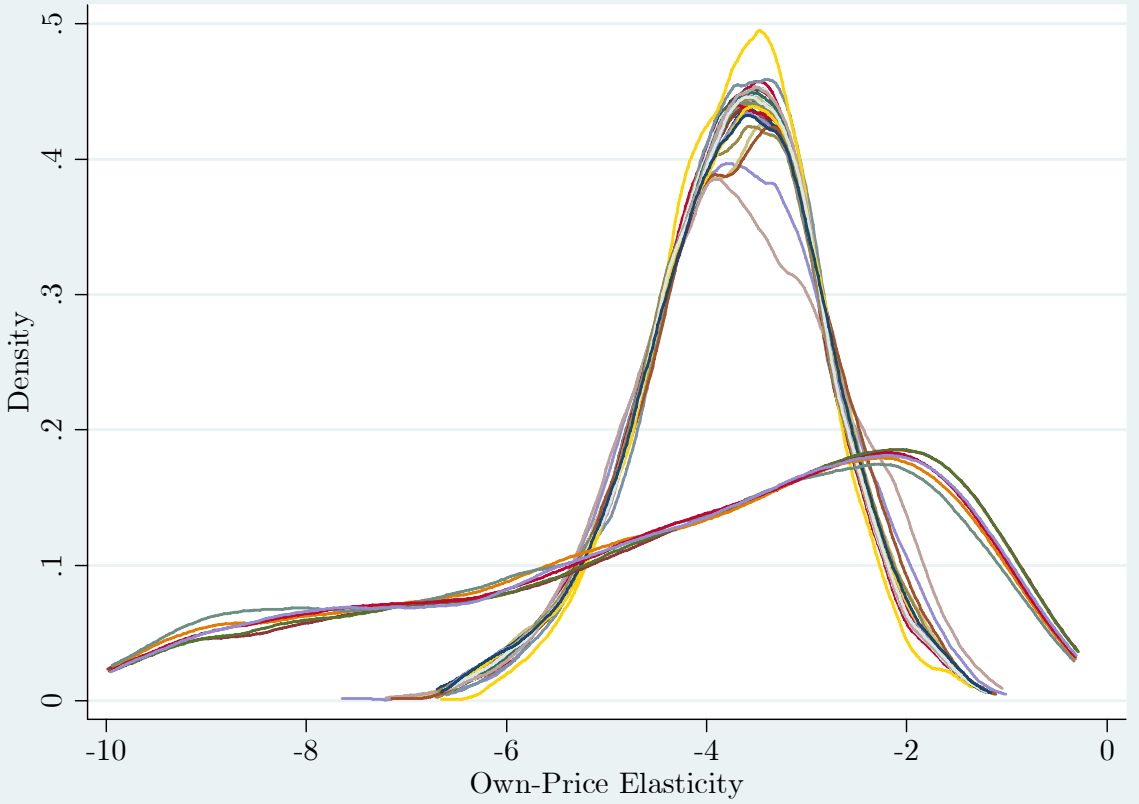
**Figure 24:** Density estimates of the estimated own-price elasticities for the 11 unique local minima for the automobile data.



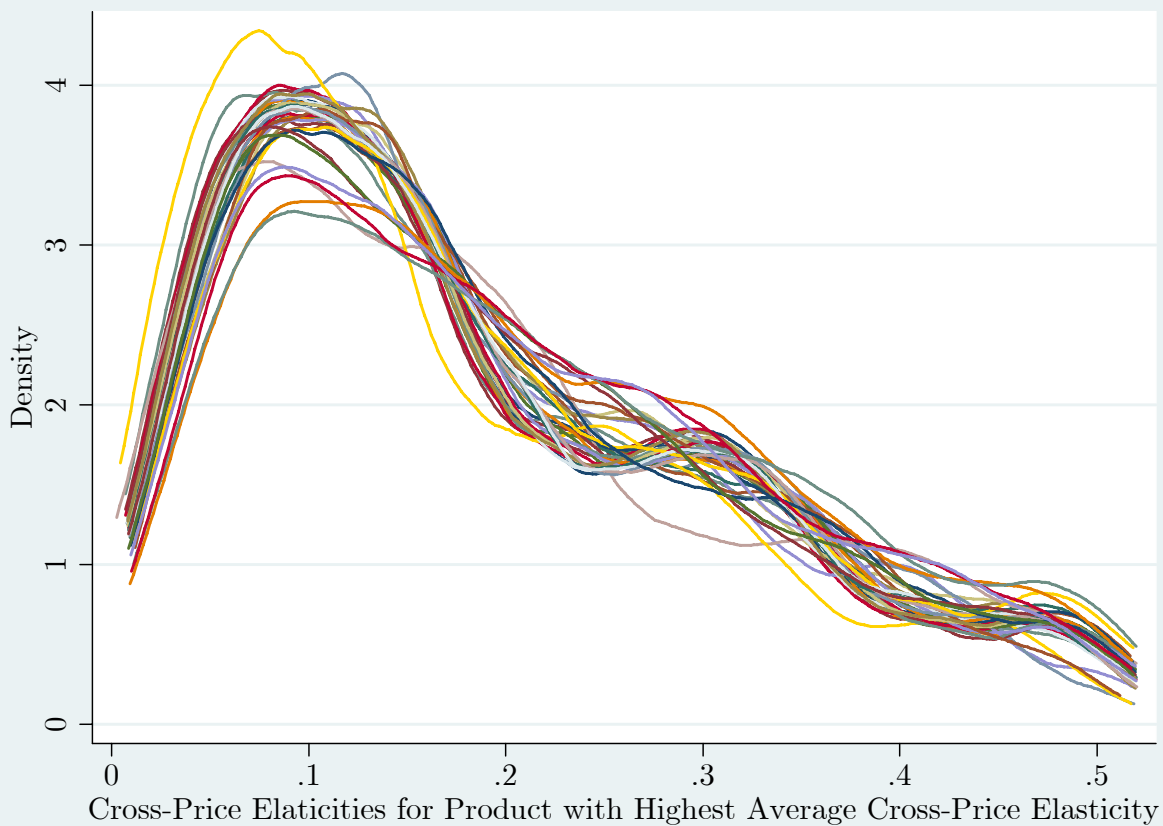
**Figure 25:** Density estimates of the estimated cross-price elasticities for the 11 unique local minima for the automobile data. These are truncated at 0.1.



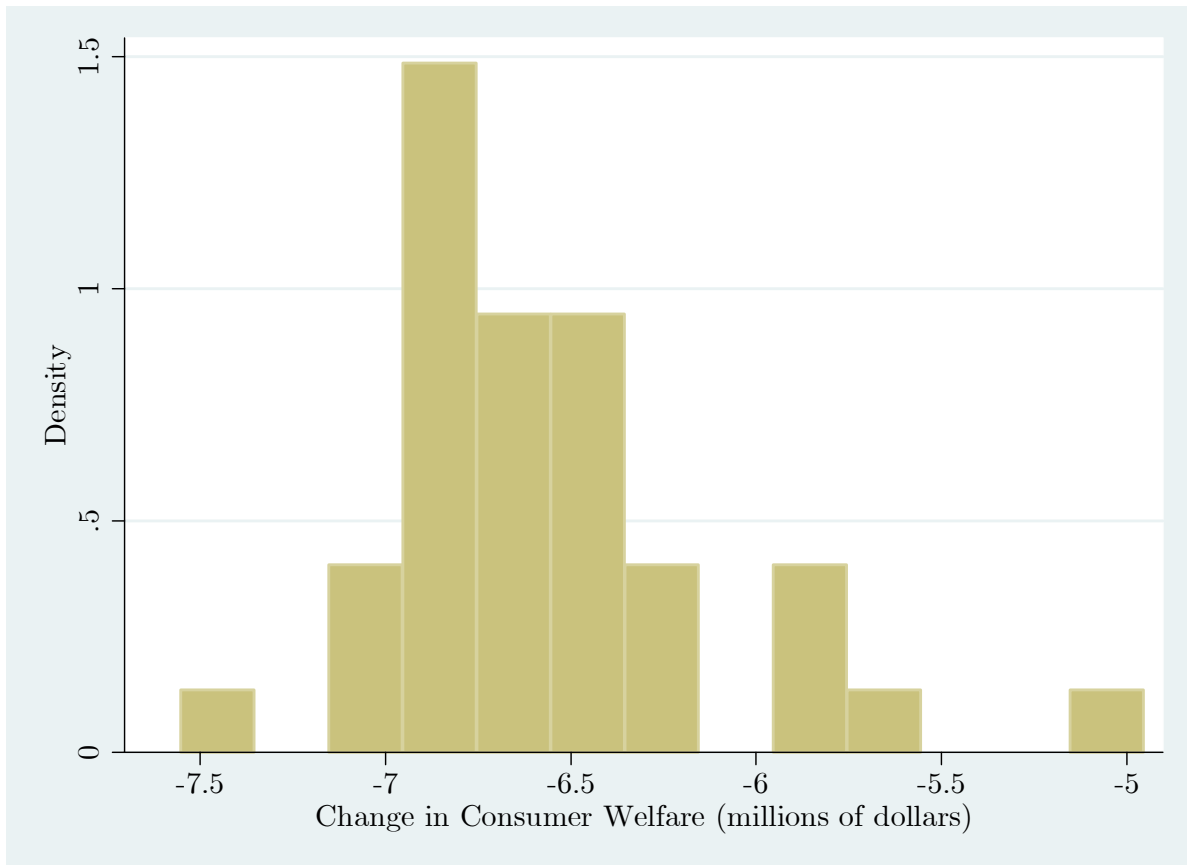
**Figure 26:** Histogram of estimated change in total welfare from hypothetical merger using the automobile data for the 11 unique local minima. The range is from -\$ million to -\$15.35 million.



**Figure 27:** Density estimates of the estimated own-price elasticities for the 37 local minima for the cereal data. These are truncated at 0.1.



**Figure 28:** Density estimates of the estimated cross-price elasticities for the 37 local minima for the cereal data. These are truncated at 0.52 (90<sup>th</sup> percentile).



**Figure 29:** Histogram of estimated change in total welfare from hypothetical merger using the cereal data for the 37 local minima. The range is from  $-\$10.00$  million to  $-\$15.50$  million.

	QNewton1	Simplex	Solvopt	Confgrad	QNewton2	JBES GA	S A	MADS	GPS	Matlab GA
Price	-0.573	-0.644	-0.666	-0.725	-0.664	-0.587	-0.593	-0.898	-0.896	-0.646
Constant	-10.800	-13.520	-17.602	-14.643	-15.885	-8.798	-7.949	-13.299	-14.418	-15.237
HP/Weight	-1.306	1.571	3.911	3.628	3.635	2.062	-1.475	3.636	3.803	3.241
Air Conditioning	1.441	1.221	1.647	1.740	1.281	1.172	0.818	2.068	2.009	1.743
Mile/\$	0.164	-0.099	-0.411	-0.382	-0.297	0.149	-0.106	-0.575	0.025	-0.085
Size	3.341	3.399	3.459	3.424	3.385	2.982	3.072	3.700	3.903	3.390
Sigma_price	-0.206	-0.265	-0.256	-0.277	-0.271	-0.218	-0.237	-0.342	-0.342	-0.246
Sigma_C	3.552	6.124	9.564	7.553	8.146	-1.237	-0.794	7.189	7.189	7.377
Sigma_HP/Weight	-5.914	-4.632	-2.667	-1.286	-1.932	0.213	-4.998	-2.720	-0.470	-1.923
Sigma_AC	-0.819	1.572	1.219	0.875	1.638	0.504	0.888	0.147	0.084	0.481
Sigma_Mile/\$	-0.158	-0.500	-0.862	-0.714	-0.674	-0.194	0.513	-0.933	-0.183	-0.390
GMM Obj	195.6	156.0	147.5	158.3	148.2	292.3	230.2	99.9	154.7	141.6

**Table 1:** Parameter estimates and GMM objective values for the 10 “best” set of results. While the results for the automobile data are not directly comparable original paper, below we include the results from BLP for comparison. Our model differs in two key respects: 1. We do not include supply side moments. 2. Our functional form for demand is slightly different. Standard errors are not included because it is uncertain how consistent standard errors would be generated. “Standard”, but likely inconsistent, standard errors available upon request.

	Reported	SEs
Price	--	--
Constant	-7.061	0.941
HP/Weight	2.883	2.019
Air Conditioning	1.521	0.891
Mile/\$	-0.122	0.32
Size	3.46	0.61
Sigma_price	--	--
Sigma_C	3.612	1.485
Sigma_HP/Weight	4.628	1.885
Sigma_AC	1.818	1.695
Sigma_Mile/\$	1.05	0.272
GMM Obj		

## C Tables

	QNewton1	Simplex	Solvopt	Conigrad	QNewton2	JRES GA	S A	MADS	GPS	Matlab GA	Reported	SEs
Price	-30.239	-30.556	-62.717	-30.842	-62.742	-30.337	-35.182	-32.254	-32.821	-31.016	-32.433	7.743
constant sigma	0.344	0.343	0.558	0.361	0.558	-0.093	0.085	0.393	0.407	0.273	0.377	0.129
price_sigma	1.374	1.806	3.312	0.672	3.313	0.008	1.848	1.702	1.060	2.127	1.846	1.075
sugar_sigma	-0.008	-0.003	-0.006	-0.009	-0.006	-0.073	-0.011	-0.006	-0.096	0.010	0.004	0.012
mushy_sigma	0.030	0.053	0.093	0.066	0.093	0.434	-0.886	0.112	0.014	0.325	0.081	0.205
C*inc	3.792	3.760	2.292	3.262	2.292	1.456	0.916	2.934	3.208	2.437	3.089	1.213
P*inc	-0.557	9.736	588.108	-0.787	588.556	-0.680	0.507	-0.068	2.369	0.348	16.598	172.334
S*inc	-0.202	-0.199	-0.385	-0.184	-0.385	0.291	-0.023	-0.191	-0.213	-0.121	-0.193	0.005
M*Inc	1.840	1.451	0.749	1.740	0.748	1.659	0.466	1.692	2.223	1.224	1.468	0.697
P*inc2	-0.049	-0.546	-30.181	0.079	-30.204	-1.441	0.507	0.275	-0.008	0.066	-0.659	8.955
C*Age	0.919	0.756	1.284	1.091	1.284	-0.173	-0.315	1.461	0.631	0.645	1.186	1.016
S*Age	0.030	0.032	0.052	0.028	0.052	0.512	0.131	0.028	0.004	0.018	0.029	0.036
M*Age	-1.782	-1.313	-1.354	-1.889	-1.353	-0.181	0.866	-1.831	-1.865	-0.950	-1.514	1.103
P*Child	1.016	4.455	11.055	2.614	11.053	0.718	0.056	13.563	-3.395	0.377	11.625	5.207
GMM Obj	19.55	17.00	4.56	20.04	4.56	530.12	131.32	15.84	50.99	34.24	14.9	

**Table 2:** Parameter estimates and GMM objective values for the 10 “best” set of results. We include the results from Nevo (2000) for comparison; here the models are identical. Standard errors are not included because it is uncertain how consistent standard errors would be generated. “Standard”, but likely inconsistent, standard errors available upon request.



Algorithm	25th			Median			75th			Largest		
	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max
Quasi-Newton 1	-4.15	-4.15	-2.43	-3.31	-3.31	-1.73	-4.79	-4.79	-2.71	-4.02	-4.02	-2.63
Simplex	-6.67	-5.02	-1.93	-14.14	-4.06	-1.15	-12.67	-5.62	-2.07	-10.01	-4.21	-2.04
SolveOpt	-6.41	-6.00	-2.39	-4.76	-4.30	-1.52	-7.03	-6.52	-2.67	-4.95	-4.64	-2.52
Conj. Grad.	-5.85	-5.49	-2.46	-13.20	-4.22	-1.76	-7.05	-5.96	-2.77	-8.69	-4.36	-2.66
Quasi-Newton 2	-5.99	-5.46	-1.78	-8.48	-4.19	-1.33	-11.68	-6.15	-2.25	-6.39	-4.37	-2.09
JBES GA	--	--	--	--	--	--	--	--	--	--	--	--
Simulated Annealing	-10.11	-2.54	0.12	-20.46	-3.20	-0.92	-8.78	-4.37	-1.68	-16.89	-3.95	-1.72
MADS	-6.38	-6.38	-0.63	-7.09	-5.45	-0.41	-8.37	-7.14	-0.78	-5.02	-4.72	-0.77
GPS	-5.84	-5.84	-0.43	-7.33	-5.33	-0.26	-7.03	-6.17	-0.19	-5.09	-4.59	-0.38
Matlab GA	-5.23	-5.06	-3.00	-4.64	-3.85	-2.24	-5.97	-5.26	-3.54	-4.40	-4.15	-3.21
<b>Total</b>	<b>-10.11</b>	<b>-6.38</b>	<b>0.12</b>	<b>-20.46</b>	<b>-5.45</b>	<b>-0.26</b>	<b>-12.67</b>	<b>-7.14</b>	<b>-0.19</b>	<b>-16.89</b>	<b>-4.72</b>	<b>-0.38</b>

**Table 3:** Own-price elasticities for the automobile data. These results report the minimum and maximum estimated elasticity obtained across converged parameter values for each algorithm. The “best” is defined as the set of parameters that achieves the lowest GMM objective value. The products are chosen based on their market shares, with the 25<sup>th</sup> representing the product with the market share equal to the 25<sup>th</sup> percentile, etc.

These results truncate the worst 5% of the sets of results, in terms of the GMM objective value. The JBES GA never converges, so its results are omitted.

The average within-product-market standard deviation in the estimated elasticities across all converged results is 4.25; the median is 3.08. The average own-price elasticity is -3.66. Among the “best” sets of parameter values, the mean standard deviation is 0.84, the median is 0.81, and the mean own-price elasticity is -4.11.

Algorithm	25th			Median			75th			Largest		
	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max
Quasi-Newton 1	-5.42	-3.03	-0.87	-5.30	-4.23	3.45	-3.96	-2.91	-1.17	-2.46	-1.86	-0.09
Simplex	-5.45	-3.04	0.42	-6.83	-4.03	3.80	-4.07	-2.92	-0.32	-2.17	-1.91	-0.07
SolveOpt	-2.93	-2.93	-2.93	-2.63	-2.62	-2.62	-2.64	-2.64	-2.64	-1.99	-1.99	-1.99
Conj. Grad.	-3.42	-3.07	-0.46	-5.40	-4.09	3.45	-3.97	-2.89	-0.37	-2.35	-1.90	-0.09
Quasi-Newton 2	-5.42	-2.93	-1.15	-5.69	-2.62	5.00	-7.51	-2.64	-0.81	-2.00	-1.99	-0.09
JBES GA	--	--	--	--	--	--	--	--	--	--	--	--
Simulated Annealing	--	--	--	--	--	--	--	--	--	--	--	--
MADS	-5.77	-4.43	-0.19	-16.50	-10.94	4.23	-2.39	-1.62	-0.05	-1.73	-0.85	-0.16
GPS	-4.72	-3.50	-0.04	-17.97	-13.08	29.41	-1.88	-1.09	-0.01	-0.98	-0.76	-0.01
Matlab GA	-5.08	-3.96	-1.80	-13.21	-8.53	-1.71	-1.77	-1.23	-0.26	-1.07	-0.77	-0.48
<b>Total</b>	-5.77	-2.93	0.42	-17.97	-10.94	29.41	-7.51	-1.62	-0.01	-2.46	-0.85	-0.01

**Table 4:** Own-price elasticities for the cereal data. These results report the minimum and maximum estimated elasticity obtained across converged parameter values for each algorithm. The “best” is defined as the set of parameters that achieves the lowest GMM objective value. For the MADS algorithm, this does not correspond to a converged set of parameters. The JBES GA and Simulated Annealing algorithms never converge, so their results are omitted.

The products are chosen based on their market shares, with the 25<sup>th</sup> representing the product with the market share equal to the 25<sup>th</sup> percentile, etc.

The average within-product-market standard deviation in the estimated elasticities across all converged results is 0.82; the median is 0.80. The average own-price elasticity is -3.58. Among the “best” sets of parameter values, the mean standard deviation is 0.28, the median is 0.23, and the mean own-price elasticity is -3.65.

Algorithm	25th (x10)			Median (x10)			75th (x10)			Largest (x10)		
	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max
Quasi-Newton 1	0.18	1.05	1.05	0.18	1.26	1.26	0.12	0.60	0.60	0.24	0.87	0.87
Simplex	0.24	2.05	3.07	0.23	2.34	3.39	0.16	1.06	60.48	0.34	1.98	5.98
SolvOpt	0.33	2.03	2.22	0.43	1.71	2.26	0.24	1.86	2.45	0.42	1.94	2.01
Conj. Grad.	0.18	1.81	3.36	0.18	1.75	4.13	0.13	1.21	39.72	0.25	1.90	6.07
Quasi-Newton 2	0.18	1.85	9.21	0.21	1.80	21.46	0.11	1.97	150.10	0.23	1.85	7.31
JBES GA	--	--	--	--	--	--	--	--	--	--	--	--
Simulated Annealing	0.11	0.50	5.84	0.13	0.83	3.86	0.09	0.64	70.58	0.17	0.81	7.82
MADS	0.16	2.10	10.33	0.13	2.00	21.03	0.07	1.12	150.00	0.28	1.98	8.54
GPS	0.10	2.10	10.21	0.13	2.55	21.03	0.03	0.87	150.00	0.13	2.03	8.69
Matlab GA	0.57	1.66	1.73	0.69	1.79	2.20	0.37	0.81	1.67	0.57	1.64	1.80
<b>Total</b>	0.10	2.10	10.33	0.13	2.00	21.46	0.03	1.12	150.10	0.13	1.98	8.69

**Table 5:** Cross-price elasticities for the automobile data. These results report the minimum and maximum estimated elasticity obtained across converged parameter values for each algorithm. The cross-product is the closest substitute for the particular product. The “best” is defined as the set of parameters that achieves the lowest GMM objective value. The products are chosen based on their market shares, with the 25<sup>th</sup> representing the product with the market share equal to the 25<sup>th</sup> percentile, etc.

The JBES GA never converges, so its results are omitted.

The average within-product-market standard deviation in the estimated elasticities across all converged results is 0.17; the median is 0.04. The average cross-price elasticity is 0.06. Among the “best” sets of parameter values, the mean standard deviation is 0.04, the median is 0.04, and the mean cross-price elasticity is 0.03.

Algorithm	25th (x10)			Median (x10)			75th (x10)			Largest (x10)		
	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max
Quasi-Newton 1	1.02	2.81	8.14	0.11	2.72	2.83	0.28	5.23	5.53	0.12	1.25	2.88
Simplex	0.54	2.76	9.01	0.03	2.74	3.06	0.09	5.06	6.16	0.02	1.41	3.10
SolvrOpt	2.83	2.83	2.83	1.71	1.71	1.71	4.42	4.42	4.42	2.07	2.07	2.07
Conj. Grad.	0.46	2.80	5.31	0.10	2.48	2.67	0.08	5.07	5.55	0.12	1.22	2.29
Quasi-Newton 2	0.57	2.83	8.50	0.08	1.71	3.37	0.06	4.42	15.54	0.12	2.07	2.56
JIBES GA	--	--	--	--	--	--	--	--	--	--	--	--
Simulated Annealing	--	--	--	--	--	--	--	--	--	--	--	--
MADS	0.01	3.61	6.43	0.08	3.54	3.83	0.00	1.63	3.42	0.25	0.29	1.08
GPS	0.00	2.29	3.77	0.01	1.67	5.77	0.00	1.00	2.32	0.01	0.40	2.34
Matlab GA	0.79	2.79	5.40	0.10	0.00	4.68	0.05	0.74	3.39	0.24	0.37	1.74
<b>Total</b>	<b>0.00</b>	<b>2.83</b>	<b>9.01</b>	<b>0.01</b>	<b>1.71</b>	<b>5.77</b>	<b>0.00</b>	<b>4.42</b>	<b>15.54</b>	<b>0.01</b>	<b>2.07</b>	<b>3.10</b>

**Table 6:** Cross-price elasticities for the cereal data. These results report the minimum and maximum estimated elasticity obtained across converged parameter values for each algorithm. The cross-product is the closest substitute for the particular product. The “best” is defined as the set of parameters that achieves the lowest GMM objective value. The products are chosen based on their market shares, with the 25<sup>th</sup> representing the product with the market share equal to the 25<sup>th</sup> percentile, etc.

The JIBES GA and Simulated Annealing algorithms never converge, so their results are omitted.

The average within-product-market standard deviation in the estimated elasticities across all converged results is 0.50; the median is 0.20. The average cross-price elasticity is 0.15. Among the “best” sets of parameter values, the mean standard deviation is 0.19, the median is 0.13, and the mean cross-price elasticity is 0.19.

GMM Obj	Local						Sigma					
	Minimum	Price	Constant	HP/Weight	AC	Mile/\$	Size	Price	Sigma Constant	Sigma HP/Weight	Sigma AC	Sigma Mile/\$
156.0	1	-0.644	-13.520	1.571	1.221	-0.099	3.399	-0.265	6.124	-4.632	1.572	-0.500
169.3	2	-0.774	-13.690	3.493	1.751	-0.203	3.536	-0.292	6.714	0.359	-0.523	-0.475
208.7	3	-0.526	-9.586	-0.116	1.412	0.094	3.071	-0.191	2.274	-4.218	-0.216	-0.218
223.5	4	-0.445	-8.648	-10.481	1.533	0.272	3.385	-0.154	-1.991	-10.069	-0.360	-0.031
228.4	5	-0.358	-14.184	-4.091	1.129	0.316	3.377	0.124	5.784	7.340	-0.822	-0.138
228.4	5	-0.358	-14.183	-4.109	1.127	0.316	3.377	0.124	5.785	7.350	-0.824	-0.137
228.4	5	-0.358	-14.180	-4.116	1.129	0.317	3.377	0.124	5.783	7.353	-0.820	-0.136
260.0	6	-0.333	-18.391	2.563	1.198	0.264	3.210	0.117	7.969	-2.240	0.087	-0.255
268.1	7	-0.465	-8.516	2.072	1.134	-0.054	2.768	-0.158	-0.255	-0.334	-0.082	-0.358
272.5	8	-0.259	-8.646	-6.986	1.008	0.249	2.906	0.075	0.223	-6.615	0.111	-0.158
272.5	8	-0.259	-8.646	-6.985	1.008	0.249	2.906	0.075	0.223	-6.615	0.112	-0.158
272.5	8	-0.259	-8.646	-6.986	1.008	0.249	2.906	0.075	0.223	-6.615	0.111	-0.158
272.5	8	-0.259	-8.646	-6.982	1.008	0.249	2.906	0.075	0.224	-6.613	0.112	-0.158
272.5	8	-0.259	-8.646	-6.983	1.008	0.249	2.906	0.075	0.223	-6.614	0.112	-0.158
272.5	8	-0.259	-8.646	-6.985	1.008	0.249	2.906	0.075	0.223	-6.615	0.111	-0.158
272.5	8	-0.259	-8.646	-6.986	1.008	0.249	2.906	0.075	0.223	-6.615	0.111	-0.158
282.7	9	-0.381	-10.985	1.937	1.058	0.256	2.931	0.127	2.310	1.008	0.306	-0.059
302.7	10	-0.291	-15.833	1.510	0.888	0.273	3.053	0.028	5.971	-2.987	-0.772	-0.205
313.8	11	-0.308	-10.266	1.304	0.848	-0.063	2.608	0.107	-2.289	-0.713	0.422	-0.447

**Table 7:** List of GMM Objective Values and parameter estimates for the 19 sets of results that have gradients below 30 in absolute value and all positive eigenvalues.

GMM Obj <sub>i</sub>	Price	Constant	Price	Sugar	Musly	Contant	Price	Sugar	Musly	Price	Constant	Sugar	Musly	Price	Constant	Sugar	Musly	Price
		Sigma	Sigma	Sigma	Sigma	*Inc	*Inc	*Inc	*Inc	*Inc2	*Age	*Age	*Age	*Age	*Age	*Age	*Age	*Child
16.07	-31.03	0.37	1.91	-0.01	0.11	3.41	-4.66	-0.20	1.64	0.36	1.09	0.03	-1.60	8.85				
16.34	-31.88	0.37	1.83	-0.01	0.05	3.05	0.00	-0.18	1.21	0.23	0.88	0.03	-1.12	10.16				
17.06	-32.70	0.42	1.51	-0.01	0.13	2.56	-12.42	-0.17	1.70	1.04	1.74	0.02	-1.99	15.91				
17.57	-30.95	0.34	2.16	0.00	0.04	3.96	3.05	-0.22	1.38	-0.10	0.50	0.05	-1.01	8.05				
19.55	-30.24	0.34	1.37	-0.01	0.03	3.79	-0.56	-0.20	1.84	-0.05	0.92	0.03	-1.78	1.02				
19.80	-30.96	0.34	1.33	-0.01	0.06	3.43	0.14	-0.20	1.96	0.00	1.16	0.03	-2.08	1.16				
20.04	-30.84	0.36	0.67	-0.01	0.07	3.26	-0.79	-0.18	1.74	0.08	1.09	0.03	-1.89	2.61				
20.06	-30.67	0.33	1.15	-0.01	0.03	3.57	-1.18	-0.20	1.73	0.04	0.95	0.03	-1.85	0.87				
20.28	-29.67	0.34	0.27	-0.01	0.10	3.92	2.97	-0.20	1.39	-0.21	0.62	0.04	-1.06	6.75				
20.29	-30.39	0.36	0.64	-0.01	0.06	3.50	-0.78	-0.19	2.02	0.00	1.32	0.02	-2.16	2.60				
20.33	-31.61	0.31	1.33	-0.01	-0.01	2.83	0.05	-0.16	1.36	0.14	1.26	0.01	-1.80	2.78				
20.40	-31.83	0.32	1.25	-0.01	0.05	3.04	0.14	-0.18	1.74	0.13	1.15	0.02	-1.94	1.12				
20.89	-32.07	0.30	1.35	-0.01	-0.02	2.90	0.05	-0.17	1.38	0.19	1.30	0.00	-1.79	2.79				
20.98	-30.13	0.34	1.15	-0.01	0.05	3.83	-0.73	-0.20	1.93	-0.08	0.94	0.03	-1.90	-0.67				
21.07	-31.19	0.33	1.29	-0.01	0.04	3.26	-0.35	-0.18	1.56	0.06	0.86	0.03	-1.69	-0.54				
21.16	-31.61	0.37	1.73	-0.01	-0.07	3.15	-0.64	-0.18	1.55	0.11	1.01	0.02	-1.81	-0.50				
21.22	-31.41	0.31	1.79	-0.01	0.02	3.01	-7.92	-0.17	1.56	0.52	1.05	0.01	-1.73	0.17				
21.38	-30.95	0.34	0.93	-0.01	0.01	3.34	0.22	-0.19	1.70	0.01	0.71	0.03	-1.48	-0.05				
21.45	-30.83	0.39	0.65	-0.01	0.03	3.33	-1.42	-0.19	1.93	0.08	1.16	0.02	-2.02	0.52				
21.46	-30.96	0.30	0.89	-0.01	0.00	3.48	-0.90	-0.20	1.82	0.05	0.86	0.03	-1.84	-0.39				
21.54	-30.16	0.36	0.62	-0.01	0.03	3.73	-1.40	-0.20	2.09	-0.02	1.06	0.03	-2.00	0.52				
21.68	-30.15	0.33	0.63	-0.01	0.08	3.51	-0.15	-0.18	1.86	-0.07	0.84	0.03	-1.70	0.46				
21.73	-31.77	0.37	1.26	-0.01	0.05	2.90	-1.35	-0.17	1.35	0.23	0.80	0.02	-1.41	0.67				
21.76	-30.81	0.36	0.59	-0.01	0.08	3.41	-0.14	-0.20	2.20	-0.02	1.38	0.03	-2.42	0.58				
21.83	-30.18	0.42	0.94	-0.01	-0.10	3.62	0.16	-0.18	1.36	-0.07	1.11	0.01	-1.63	1.15				
21.91	-30.35	0.37	1.06	-0.01	0.06	3.65	-0.73	-0.19	1.68	-0.02	1.03	0.01	-1.69	-0.70				
21.91	-30.87	0.36	0.43	-0.01	0.02	3.10	1.12	-0.17	1.47	-0.02	1.16	0.02	-1.84	1.66				
22.25	-30.48	0.33	0.79	-0.01	0.00	3.37	0.65	-0.18	1.89	-0.07	0.90	0.02	-1.57	0.33				
22.25	-31.65	0.32	0.70	-0.02	-0.16	2.91	-0.25	-0.18	1.49	0.11	1.06	0.02	-1.91	0.33				
23.21	-34.00	0.27	3.10	0.00	-0.07	2.25	5.54	-0.17	1.25	0.08	0.59	0.02	-1.13	2.32				
24.11	-26.23	0.30	0.92	-0.01	0.22	4.84	-10.08	-0.19	2.22	0.01	1.36	0.01	-2.12	3.88				
26.34	-30.75	0.43	-0.19	-0.02	0.26	3.16	-0.70	-0.17	2.23	0.05	1.27	0.01	-2.09	0.41				
27.14	-29.35	0.27	-0.05	0.00	-0.06	3.95	-8.23	-0.19	0.85	0.41	0.23	0.04	-0.35	6.07				
27.36	-33.29	0.27	1.01	-0.01	-0.09	2.21	-1.35	-0.16	1.11	0.43	0.39	0.04	-1.15	0.57				
32.47	-35.79	0.36	0.46	-0.04	0.41	0.54	5.41	-0.10	1.94	0.45	2.80	-0.07	-3.07	9.17				
37.29	-34.03	0.36	0.02	-0.02	-0.18	1.73	-0.13	-0.13	1.00	0.50	0.56	-0.02	-0.56	1.63				
53.61	-36.25	0.29	-0.68	-0.03	-0.35	0.44	-0.96	-0.11	1.58	0.82	0.59	0.02	-1.46	-0.95				

**Table 8:** List of GMM Objective Values and parameter estimates for the 37 sets of results that have gradients below 30 in absolute value and all positive eigenvalues.