

NBER WORKING PAPER SERIES

ESTIMATION OF RANDOM COEFFICIENT DEMAND MODELS:
CHALLENGES, DIFFICULTIES AND WARNINGS

Christopher R. Knittel
Konstantinos Metaxoglou

Working Paper 14080
<http://www.nber.org/papers/w14080>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2008

We have benefited greatly from conversations with Steve Berry, Severin Borenstein, Michael Greenstone, Phil Haile, Aviv Nevo, Hal White, Frank Wolak, Catherine Wolfram, and seminar participants at the University of Calgary, University of California at Berkeley, the University of California Energy Institute, and the 2008 NBER Winter IO meeting. Metaxoglou acknowledges financial support from Bates White, LLC. We are also grateful to Bates White, LLC for making their computing resources available. All remaining errors are ours. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2008 by Christopher R. Knittel and Konstantinos Metaxoglou. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Estimation of Random Coefficient Demand Models: Challenges, Difficulties and Warnings
Christopher R. Knittel and Konstantinos Metaxoglou
NBER Working Paper No. 14080
June 2008
JEL No. C1,C61,C81,L1,L4

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 non-trivial for a variety of reasons. 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, and more recently in public and labor. We find that convergence may occur at a number of local extrema, at saddles and in regions of the objective function where the first-order conditions are not satisfied. 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 can vary between positive values to negative seventy billion dollars depending on the set of parameter estimates used.

Christopher R. Knittel
University of California, Davis
Department of Economics
One Shields Ave
Davis, CA 95616
and NBER
crknittel@ucdavis.edu

Konstantinos Metaxoglou
Bates White LLC
kmetaxoglou@ucdavis.edu

1 Introduction

1.1 What this paper is about

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 nonlinear search algorithm along with sets of starting values and stopping rules. For a common class of demand models used in industrial organization, and more recently in labor and public economics, we find that termination of many popular nonlinear search algorithms may occur at local extrema, saddles points, as well as in regions of the objective function where the first-order conditions for optimality fail. Furthermore, parameter estimates and measures of market performance, such as price elasticities, exhibit notable variation depending on the combination of the algorithm and starting values in the optimization exercise at hand.

Our findings highlight the importance of verifying both the first- and second-order conditions for a local optimum, and emphasize that researchers often face the dilemma of choosing among local extrema—a non-trivial task for several reasons.¹ 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 and there is little or no statistical guidance as to which of the local extrema is the consistent one.² Additionally, consistency proofs are local in nature; they may say little about the behavior of the objective function outside of the neighborhood where the true parameter vector lies. Finally, even in cases where the global extremum is the consistent root and the econometrician is convinced she has uncovered it for her 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 if they ignore issues of multiple extrema.

According to our results, the design of the optimization exercise undertaken by an empirical economist may have large effects on the economic variables of interest. The difficulties of

¹For a compact discussion of consistency theorems regarding extremum estimators, see Section 5.3 in Cameron and Trivedi (2005). Amemiya (1985) using a mixture of normals provides an example of an inconsistent global MLE. In the same example, a root of the likelihood equation may be consistent (example 4.2.2, pg. 119). McFadden and Newey (1994) provide an in-depth discussion of the large-sample properties for popular classes of extremum estimators, such as MLE, NLS and GMM.

²Amemiya (1985) mentions two ways for gaining confidence that a local extremum is a consistent root: (1) if the solution gives a reasonable value from an economic-theoretic viewpoint and (2) if the iteration by which the local extremum was obtained started from a consistent estimator.

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, among empirical economists. We rarely discuss the details of our optimization exercises. In most cases, we fail to show how the conclusions of our work change under alternative sets of parameters. Gradient norms and Hessian eigenvalues are almost never reported. The purpose of this paper is to convey that a thorough optimization exercise and a clear documentation of its design are as important to the conclusion of an empirical study as the identification strategy.

We draw our examples from a specific class of discrete-choice demand models for differentiated products which have been particularly popular since 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 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, due to the inherent horizontal differentiation of products. Faced with a change in one of the characteristics of their most preferred product, consumers are more likely to switch to products with similar attributes.

Answers to a variety of important economic questions have been provided using BLP-type models. 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 BLP-type demand systems.³ 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.⁴ Finally, researchers modeling school and housing choices have used BLP-type models.⁵

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 term inferred by equating observed to estimated market shares of the products under consideration, which requires an empirical approximation. To highlight these difficulties, we

³For example: new good valuation (Petrin [2004]); trade policies (BLP [1999]); mergers (Nevo [2000b]); construction of price indices that account for quality changes and product introductions (Nevo [2003]). Brenkers and Verboven (2006) extend the individual-level logit of BLP to nested logit to examine the effects of the car industry restructuring in the EU.

⁴For dynamic extensions of the BLP model see, for example, Hendel and Nevo (2006), Gowrisankaran and Rysman (2008).

⁵For school choice, see Hastings, Kane and Staiger (2007). For housing choices, see Bayer and McMillan (2005) and Bayer, McMillan and Rueben (2005).

use two different publicly data sets. The first data set is from BLP. The second is the one used by Nevo (2000a). For each data set, we engage in a thorough estimation exercise employing 10 different optimization algorithms using 50 different starting values for each algorithm; that is, we estimate the same model 500 times for each data set.

All algorithms are prone to uncover local minima, saddles and to terminate at regions with non-zero gradients. The parameter estimates across the local minima and termination points vary greatly both within and across algorithms; more importantly, the variation in parameter values is also economically important. Price elasticities 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 yielding 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 even greater variation for estimated cross-price elasticities.

The variation across algorithms that exists 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, may not be sufficient. Depending on the termination criteria used, we find that one algorithms 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 two hypothetical mergers. The differences in parameter estimates can have equally large effects on the implied consumer welfare consequences of a merger; in short, we find anything is possible, from nearly no effect on consumers to large welfare consequences. Much of this variation remains even when we estimate the model 50 times for each algorithm and select the parameter estimates that correspond to the set of starting values that give rise to the smallest value of the objective function within the same algorithm.

Multiple extrema and saddles are the by-product of the nonlinear nature of the model; if the objective function is not globally concave (convex), by definition, multiple extrema and/or saddles will exist. We stress that our results are *not* due to weak identification. Weak identification leads to the objective function being relative flat at the consistent root (Stock and Wright [2000]). This may lead to algorithms stopping at multiple places around a local extremum, which will generate some of the variation that we observe. However, we are unaware of any research showing that weak identification *causes* multiple extrema or saddles; we clearly find

both in the data. For this to be true, strong identification would somehow have to lead the objective functions of nonlinear models to be globally concave/convex, as in a linear model.

We also stress that the variation we uncover due to multiple extrema is not captured in typical standard error calculations. 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 extremum 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 potentially disjoint; at the very least, they would be functions of these other local minima.

Our work is also related to a 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. Although, 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.

Finally, the paper is 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 studies have documented that GMM estimators in over-identified models may exhibit substantial biases.⁶ In addition, when instruments are either weak or many, the criterion

⁶See, for example, Hansen, Heaton and Yaron (1996), Altonji and Segal (1996), Burnside and Eichenbaum (1996) and Pagan and Robertson (1997).

function may exhibit 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.⁷ 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.

1.2 What this paper is not about

There are a lot of important topics that this paper does not address. First, this paper does not improve upon or teach existing nonlinear search methods. Our goal is simply to use existing techniques and apply them to a common class of models in industrial organization. Second, while we believe the topic to be of upmost importance, we say nothing about inference in the context of multiple extrema. When drawing inferences, the existing literature assumes away multiple extrema. Standard errors are calculated using only information about the local extremum. As noted above, we conjecture that consistent inference will, at least in part, be a function of the multiple extrema. The closer the objective function values are across the extrema, the more relevant these issues become.

Finally, our goal is not to discourage the use of structural econometric models that very often go at hand with nonlinear optimization methods. We want to encourage the use of thorough optimization designs, as well as the documentation of alternative sets of parameters meeting the conditions of local extrema. It would be unfortunate for our work to be viewed as a campaign against a structural approach in empirical analyses, that is invaluable addressing certain types of economic questions, such as those surrounding the welfare implications of market structures.

⁷See, 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).

2 The Model and Estimation

The starting point of a 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 non-price product characteristics, and ξ_{jt} includes the unobserved to the econometrician features of the product, such as after-sale services and image. The logit error term ε_{ijt} captures unobserved consumer heterogeneity.⁸ Consumer heterogeneity is also captured by the individual-specific coefficients associated with price and other observed product 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 Π and Σ are matrices of parameters. The elements of Π measure the importance of the demographic variables in shaping preferences, “observed variation”, while Σ captures “unobserved” variation in preferences. Demographics were introduced in Nevo’s work (e.g., Nevo [2001]); BLP captured heterogeneity only through the v s. Following Nevo (2000a), we decompose the expression in (1) into a component that is common across all consumers and a component that highlights their heterogeneity:

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

The first term with associated parameters θ_1 enters linearly in the GMM objective function, while the second, with associated parameters θ_2 enters in a nonlinear fashion. 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:

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

Therefore, the market share for product j is the probability that A_{jt} obtains, which under

⁸This specification ignores income effects, which may be important depending on the application.

appropriate independence assumptions may be expressed as:

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

For the purpose of estimation, distributional assumptions for ε , v and D are necessary. The vast majority of the literature assume extreme value and standard normal distributions for ε and v , respectively. Nevo draws demographics from the empirical distribution of the Current Population Survey as opposed to a parametric distribution.

Estimation follows Berry (1994). For a given θ_2 , there is a vector of mean utilities, δ , that equates predicted market with observed market shares. We can then decompose this vector of δ s into the mean level of observed utility, $x\beta - \alpha p$, and unobserved quality, ξ . The vector of unobserved product 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 ξ .⁹

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

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

where Φ is a consistent estimate of $E[Z' \xi \xi' Z]$. Integrating out the logit error ε , and assuming simulation of purchase decisions for ns individuals implies:

$$\begin{aligned} s_{jt}(x_{jt}, \delta_{jt}, P_{ns}; \theta_2) &= \frac{1}{ns} \sum_{i=1}^{ns} s_{ijt}, \\ s_{ijt}(x_{jt}, \delta_{jt}, P_{ns}; \theta_2) &= \frac{\exp \left[\delta_{jt} + \sum_{k=1}^K x_{jt}^k (\sigma_k v_i^k + \pi_{k1} D_{i1} + \dots + \pi_{kd} D_{id}) \right]}{1 + \sum_{m=1}^J \exp \left[\delta_{mt} + \sum_{k=1}^K x_{mt}^k (\sigma_k v_i^k + \pi_{k1} D_{i1} + \dots + \pi_{kd} D_{id}) \right]}. \end{aligned} \quad (7)$$

The individual market shares s_{ijt} are functions of δ_{jt} and θ_2 only. More precisely, for a given θ_2 , there is a δ_{jt} that equates observed market shares with predicted market shares for all j

⁹In principle, if the ξ 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.

and t . The difference $\delta_{jt} - x_{jt}\beta$ defines the structural error term:

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

As opposed to the simple and nested logit and the Generalized Extreme Value models, where analytical expressions for δ are readily available, the random coefficient logit employs a contraction mapping that solves for the vector of mean utilities δ that equates observed to predicted market shares with a k th iterate given by:

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

The empirical solution to the fixed point iteration in (9) yields an empirical approximation to the nonlinear market share functions used for estimation. Estimation requires a nonlinear search to find θ_1 and θ_2 that minimize the GMM objective function in (6). Because the structural error term is linear in θ_1 , concentrating out θ_1 and searching only over δ and θ_2 alleviates some of the computational burden associated with the nonlinear search. In more detail: a new iterate of θ_2 gives rise to a new vector of mean utilities δ , which implies a new iterate for θ_1 . The new iterate for θ_1 is derived using linear IVs.

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.¹⁰ 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 require a set of starting values and stopping rules.

The starting values for the mean utility vector δ 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; therefore, the variation in the starting values represents regression error plausibly obtained by across researchers. For the vector of coefficients θ_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 θ_2 *a priori*.

¹⁰ Available at <http://www.econ.ucdavis.edu/faculty/knittel/>

We employ ten search algorithms that are either derivative-based or direct search routines. The former utilize some information about the steepness and the curvature of the objective function, without necessarily keeping track information associated with the Hessian. 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. Four of our algorithms are derivative based; three are deterministic direct-search methods; and, three are stochastic direct-search algorithms. 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, the third is a conjugate gradient, while the fourth comes from the constrained optimization literature taking into account constraints with the method of exact penalization. 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 two quasi-Newton algorithms are available in the Mathworks Optimization Toolbox and on the website maintained by Hans Bruun Nielsen.¹¹ 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 the fourth routine based on Shor's r-algorithm.¹² Burke et al. (2007) provide a compact self-contained discussion of Shor's r-algorithm (see Kappel and Kuntsevich [2000] for additional details).

The three 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 the reader 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).

The stochastic direct-search routines include genetic algorithms and a simulated annealing algorithm. The codes for our two genetic algorithms are provided in the GADS toolbox and on the website maintained by Michael Gordy.¹³ Our simulated annealing code is our translation of the code originally developed for the Gauss matrix language by E.G. Tsionas available at the

¹¹ Available at <http://www2.imm.dtu.dk/~hbn/Software/>

¹² Available at: <http://www.uni-graz.at/imawww/>

¹³ Available at: <https://www.federalreserve.gov/research/staff/gordymichaelx.htm>

Gauss archive of the American University.¹⁴ 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 (Mathworks), Quasi-Newton 2 (Nielsen), 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, but not all, convergence is 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. We used a tolerance of 1E-3 for changes in both the parameter vector and the objective function.

We limited the number of function evaluations to 4,000. Imposing an upper bound on the number of function evaluations was largely dictated by the use of the direct-search algorithms, notably the Nelder-Mead simplex, which repeatedly appeared to “stall”: The algorithm 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 automobile data set of BLP, the algorithm did not improve upon the value of the objective function at the third decimal point after the first 94 iterations that required 152 function evaluations for the next 7,848 function evaluations.¹⁵ However, because the simplex method uses stopping rules that differ from most algorithms, convergence is never achieved.

In what follows, we assume that the 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. This admittedly arbitrary convergence assumption should only affects the results for the automobile data. The bulk of our reported results are broken up by algorithm, so the reader can see that this assumption does not affect our conclusions.¹⁶ For the remaining algorithms, we focus only on those sets of starting values that meet our convergence criteria and omit those results associated with termination implied by the constraint on the number of function evaluations.

¹⁴Available at: <http://www.american.edu/academic.depts/cas/econ/gaussres/optimize/optimize.htm>. The Matlab code is available from the authors upon request. Recently, Mathworks has included simulated annealing in its GADS toolbox.

¹⁵We imposed an upper bound of 8,000 to the number of function evaluations for this example.

¹⁶This 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.

Finally, we use the variable tolerance for the contraction mapping in Nevo (2000a).¹⁷ The tolerance for the contraction mapping begins at $1\text{E-}8$, it is reduced by a decimal point every 50 iterations of the contraction mapping and reset for every iteration of the nonlinear search. In section 7, we investigate the robustness of our results to the contraction mapping tolerance.

4 Results

In this section, we summarize our demand estimation results using the ten optimization algorithms described above for two 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.

In the case of the cereal data, our specification of the demand equation is the one used by Nevo. Specifically, the specification includes: brand dummies which subsume brand characteristics other than prices; random coefficients associated with a constant term, price, sugar content, and whether the cereal gets “mushy”; interaction terms between income and the constant term, income and price, income and sugar content, income and the mushy dummy variable; an interaction term between income squared and prices; interaction terms between age and the constant term, age and sugar content, and age and the mushy dummy variable; and, an interaction terms between whether the consumer has a child and price. We use the instruments included in the data set.

In the case of the automobile data, our specification 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. Our specific model includes a constant term, price, HP/weight, a dummy variable for whether the car has air conditioning, MPG and size as linear characteristics. The random coefficients are associated with: the constant, price, HP/weight, the dummy variable for whether the car has air conditioning and miles/gallon. Our instruments consist of the non-price automobile characteristics, their sums across other automobiles produced by the same firm, as well as their sums across automobiles produced by the rival firms.

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

¹⁷As in Nevo (2000a), we also draw 50 individuals. We increased the number of individuals to 500 and our conclusions remain.

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 converges and minimizes the GMM objective function across the 50 sets of starting values. We define the “best of the best” set of parameters as 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 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.¹⁸ We truncate the GMM values at their 90th and 75th percentiles for the automobile and cereal data sets, 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 both data sets, the JBES GA never converged and led to extremely unstable results; therefore, we omit these results in our discussion.

For the automobile data, the GMM objective function values for each of the ten algorithms lies between 125.5 and over 100,000 for those starting values of the parameters that allowed the algorithms to converge. The number of parameter starting values that led to convergence is 379. The range of GMM values implied by the best sets of parameters is between 125.5 and 215.6. Only the simulated annealing algorithm gave rise to a GMM value equal to 125.5 even after using 50 starting values for each algorithm. Figure 2 plots a histogram of the GMM objective values uncovered; again, the figure truncates the upper 10% of values. We see that algorithms often converge in a region where the objective value is 215 and above; rarely does the process converge in regions with objective values below 200. For the cereal data, the GMM values lie between 4.56 and 50.99, 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

¹⁸The 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”.

of the GMM values is between 4.56 and 2,241.51. Convergence was achieved for 299 of the sets of parameter starting values. Figure 4 plots a histogram of the resulting GMM objective values. Unlike the automobile data, convergence is often achieved near the minimum of the distribution; although this is true for only two of the ten algorithms.

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 in our discussion below.

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 economics. For the automobile data, SolvOpt is dominated by the simulated annealing algorithm; in other specifications MADS and GPS have uncovered the best sets of results for the automobile data. This suggests that 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 10 algorithms may not be enough.

Tables 1 and 2 report the best set of parameter value across algorithms for the automobile and cereal data, 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 0.22 and 0.55; sometimes the mean marginal utility for horsepower is positive, while other times it is negative. Table 1b reports “naive” estimates of standard errors, in the sense that they ignore multiple extrema. These standard errors suggest that formal tests of the parameters would reject equality across algorithms. In the cereal data, the coefficient associated with price lies between 30.2 and 114.1 (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 -0.79 and 588.56.

We note that all of the parameter values seem “reasonable.” This is important because if only the parameter values associated with one 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 or lead a researcher to continue her search until she has found this point in the parameter space.

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.¹⁹ 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 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 24×24 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 the 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

¹⁹While 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.

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; we omit those results corresponding to GMM objective values in the upper decile of all results that led to convergence. 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 SolvOpt achieved the lowest GMM value. Because the JBES genetic algorithm never converges, we omit it from our discussion of the implied elasticities.

The variation in the estimates of the utility parameters significantly affects the own-price elasticity estimates for each of the four products. If we focus on parameter starting values that implied convergence, the own-price elasticity for product 1 varies between -40.47 and -0.79.²⁰ The GPS algorithm finds the extremes of this range. The other three products exhibit greater variation. The elasticity estimates for product 2 lie between -49.43 and 0.06. Moving to products 3 and 4, the corresponding ranges are -91.43 to -0.53, and -87.04 to 1.84, respectively.

To get a feel for whether these extremes are outliers, Figure 5 plots the histogram of elasticities for the 379 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 across all 500 estimation exercises. For all four products, a significant amount of the distribution falls outside what would appear to be reasonable variation in the estimates; it is not uncommon to see variation

²⁰We omit those sets of results that lead to an objective value above the 75th percentile of the results leading to convergence. This corresponds to an objective value of roughly 277.

exceeding 200 percent. We admit that given these differences do not represent sampling variation, one could argue that reasonable variation is no variation. For the first two products the true elasticity falls outside of the bulk of the distribution and is estimated to be more elastic than other estimates. For the final two products, the truth lies roughly in the middle of the distributions and in between two modes for the product with the largest market share.

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 379 sets of parameter starting values that allowed convergence. Ideally the distribution of these standard deviations should be degenerate with a mass at zero; it is not. The mean of this standard deviation is 16.8; its median is 11.12. To put these numbers in context, the mean elasticity is -6.63.

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. Table 3 shows that for product 1, elasticities vary by a factor of two, from -3.72 to -1.57. The other three products exhibit larger variation: -3.76 to -1.01 (product 2), -5.00 to -0.53 (product 3), -3.42 to 1.08 (product 4). The mean within-product-market standard deviation among the best set of parameter starting values is 8.42, while its median is 6.05; therefore, the standard deviation in the elasticity estimate for the product with the median level of variation is greater than the average elasticity.

Finally, we turn to summaries of the entire set of own-price elasticities. Figure 6 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 time. 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 7. *The “truth” is never found.* The place in the parameter space leading to an objective value of 125.47 was uncovered only once and on the 29th set of starting values. 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 four specific products within the cereal data are more robust, although the variation is still notable. Across all products, we notice similar variation to that found in the automobile data. 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 the results for automobiles. For product 1, the range of elasticities across all sets of parameter associated with convergence is between -47.15 and 8.65. For product 2, the range is -2.41 to -0.09. The range for products 3 and 4 is -11.22 to 0.24, and -1.47 to -0.11, respectively. Assuming a researcher tries 50 starting values, a single algorithm and reports the results from the best set of parameters, the range remains wide: -45.35 to -12.55, -1.92 to -1.55, -6.72 to -2.34 and -1.46 to -0.75 for products 1, 2, 3 and 4, respectively.

The mean standard deviation of the elasticity estimates for a given product-market combination, is 7.50 among estimates associated with convergence, while the mean elasticity is -9.93. Using a normal approximation, this suggests that 34 percent of the time the own-price elasticity will vary by over 80 percent. Interestingly, although the four specific products exhibit less variation compared to their counterparts in the automobile data, their within product-market combination standard deviations are similar.

Histograms of own-price elasticities are provided in Figure 8. Largely because of two algo-

rithms, 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 so for 32 of 50 sets of starting values. To isolate the effects of SolvOpt and Quasi-Newton 2, we also provide histograms using the remaining six algorithms that exhibited convergence and had GMM objective values below the upper quartile. Omitting SolvOpt and Quasi-Newton 2 changes the shape of the histograms considerably. The remaining algorithms rarely find elasticities 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 10. 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. 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. As with our analysis of own-price elasticities, we first focus on four specific product and then include densities for a larger number of elasticities for the cereal data. For the product-specific analysis, the choice of a substitute is necessary; we opt for the closest one. Our measure of closest substitutability is the largest average cross-price elasticity across all sets of results that led to convergence.²¹

Although our density plots can include all of the estimated cross elasticities, 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 . Additionally denote the elasticity of quantity demanded for good i with respect to the price of good j by η_{ij} . 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 plot the cross-price elasticities of a

²¹We again omit results associated with GMM objective values in the upper quartile.

single product chosen independent of the results. Because only the cereal data have products that repeat across markets, we show only density plots for the cereal data.

4.2.1 Automobile Data

The variation in cross-price elasticities is larger than that of own-price elasticities; cross-price elasticities can differ by two orders of magnitude across the sets of parameters associated with convergence, sometimes three (Table 5). For product 1, the cross-price elasticity with its closest substitute ranges from 0.006 to 2.74 (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: 0.10 to 0.31. For product 2, the range across all parameters associated with convergence is 0.006 to 1.71; for the best set of parameters the range is 0.036 to 0.15. Products 3 and 4 exhibit even more variation. With the set of parameters associated with convergence, the range in the cross-price elasticity for Product 3 is -0.05 to 4.61. If we move to the best set of parameters, the range becomes -0.04 to 0.19. Product 4 ranges from -0.027 to 3.93 and -0.015 to 0.09 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 11 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 the 90th percentile. In some ways, the results for cross-price elasticities are even more dramatic compared to own-price elasticities. For Products 2, 3 and 4, 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 0.10.

Finally, we calculate the within product-market standard deviation in the cross-price elasticities. We calculate the standard deviation separately across all routines that converged and across the best sets of parameters for each algorithm (ignoring those algorithms that never converge). The median cross-price elasticity across the over 230,000 elasticities and those routines that converged is 0.02. The average within product-market standard deviation in these estimates is 0.23, while the median is 0.04. Therefore, it is not uncommon for estimates to vary by plus or minus 200 percent within a given product. The variation is reduced for the best sets of parameter values because four of the nine routines find points in the parameter space in the same area (leading to a GMM objective value of 215). The mean standard deviation is 0.02, the median is 0.003, and the median cross-price elasticity is .01.

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 8 (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 most variation among the four products: 0.48 to 8.11. The elasticities for product 2 lie between 0.25 and 0.37. For product 3, the range of elasticities is 0.015 to 0.38. Finally, product 4 has cross elasticities that range from 0.012 to 0.18. This variation seems to be larger than that of the automobile data.

Figure 12 shows the histogram of estimates for each of the four specific products. The figure illustrates that, despite 82 of 299 sets of results converge at the same location of the parameter space, a significant amount of variation remains. 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 13).

Figure 14 plots densities of cross-price elasticities across the 10 best sets of parameter values. Specifically, let $\eta_{AB} = \% \Delta Q_A / \% \Delta p_B$, Figure 14 plots the cross-price elasticities associated with product B that has the highest average elasticity across all markets, using all sets of parameters that lead to convergence. We truncate the elasticities below by zero and above by the 75th percentile. When we plot the distribution of a large number of cross-price elasticities, the cereal data appear to exhibit large variation across the best set of results for each algorithm. 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 (Quasi-Newton 1) and 588.2 (SolvOpt). 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 15 and 16 focus on different ranges of the cross elasticities. Focusing on a smaller range of elasticities magnifies the differences.

The within product-market standard deviation in the cross-price elasticities show similar patterns to the automobile data. The median cross-price elasticity across the over 50,000 elasticities and among those routines that converged, is 0.10. The average within product-market standard deviation for these estimates is 0.67, while the median is 0.10. These ranges are similar even among the best outcomes for each algorithm (also ignoring those best outcomes

that did not converge). The mean within product-market standard deviation is 0.44, with 0.07 as the median. Therefore, even if researchers try 50 starting values, but using one algorithm, a significant amount of variation persists.

4.3 Merger Simulations

With demand estimates available, the construction of a matrix of price derivatives emerging from the first-order conditions implied by profit maximization is straightforward. Combined with information on the ownership structure of the market and a conduct mode, inferring marginal cost is possible. For example, under static Bertrand and constant marginal costs, the first-order conditions associated with the firms' profit-maximization problems imply:

$$p - mc = \Omega(p)^{-1} s(p), \quad (10)$$

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 Ω 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 its (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 Ω , marginal costs are directly obtained using (10).

A simple change of 1s and 0s in the ownership structure matrix along a series of additional assumptions (see Nevo [2001]) allows the simulation of a change in the industry's structure, as the one implied by mergers among competitors, as well as the evaluation of its welfare effects.²² In what follows, we analyze the range of values for a measure of consumer welfare on the basis of post-merger equilibrium prices. For 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 nonlinear equations:

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

The elements of $\hat{\Omega}^{post}$ 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

²²Exercises for evaluating the welfare effects associated with the introduction of new goods are performed in a very similar manner (e.g., Petrin 2002).

their use pre-merger levels instead of solving the above system of nonlinear equations in (11).²³ Thus, we avoid dealing with issues related to the numerical instabilities of Newton routines used in the solution of nonlinear equations, as well as with issues related to potentially multiple equilibria. Although we agree that the discussion of both issues is important, it is beyond the scope of our discussion here.

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.

4.3.1 Automobile Data

The automobile data set is a repeated cross-section of virtually all cars sold annually in the US between 1971 and 1990. We assume that the GM and Chrysler merger takes place in the middle of our sample, namely in 1981. Therefore, we have an annual estimate of the merger's impact on prices. We report the average of the price impact from 1981 to 1990.

The average price change across all optimization routines that achieved convergence is \$1,660, with a median of \$0. The average within-product standard deviation in this estimated price change is \$15,128 across all sets of parameters associated with convergence. Truncating at the 90th percentile gives rise to an average standard deviation of \$10,641, still very large. The median of the standard deviations is zero given the large number of zero predicted price changes and the value of their 75th percentile is \$1,446. Therefore, it is not uncommon to see variation in the estimated price changes that as large as their mean.

Having discussed our finding for price changes, in what follows, we focus on changes in consumer welfare. Figure 17 plots the histogram of annual welfare changes assuming there

²³This approximation is also discussed in Nevo (1997)

are 265 million consumers calculated over the second half of our available sample. The figure truncates the lower 1% of values and also positive values. Even when we ignore uncertainty due to sample variation, among the best sets of results for each algorithm, the predicted consumer effects range from positive to negative \$52.76 billion. Across all sets of results that led to convergences, we see large masses near zero and between negative \$10 billion. The parameters that yield the lowest objective function value lead to a welfare estimate at the lower end of this mass with a value of negative \$9.45 billion.

4.3.2 Cereal Data

In the cereal data, the average price change across all routines achieving convergence is 2.66 cents per serving, with a median of 2.07 cents per serving. The average within-product standard deviation in this estimated price change is 748.7 cents across all sets of parameters associated with convergence. Once again, outliers drive some of this, but the median is still large—326.3 cents. Across the best set of parameters, the average price change is 2.77 cents. The median price change is lower (2.07 cents), while the mean and median within-product standard deviation are 492 and 187 cents, respectively.

The merger counterfactual results with the cereal data are more robust. We do, however, observe large differences in the estimated welfare effects, especially when we omit the two algorithms that clearly outperform the others. Figure 18 plots the histogram of welfare changes assuming there are 265 million consumers for all algorithms. Figure 19 omits the algorithms 3 and 5. While there is considerable variation, the results are not as striking as in 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.²⁴

5.1 Automobile Data

Among the 500 combinations of starting values and optimization algorithms, 379 achieved convergence. Among these 379 combinations, 192 have gradients for which the $\infty - norm$ is below 30, 173 have gradients with $\infty - norm$ below 20 and 128 have gradients with $\infty - norm$ below 10.²⁵ 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 60 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; this point in the parameter space does not meet the first order conditions. The GMM objective value ranges from 215.0 to 252.46 across these points in the parameter space. Despite trying 500 different starting value and algorithm combinations and having 379 of them “converge”, we know that we have not found the global minimum. We also note that we are restricting ourselves to a fairly small neighborhood of starting values; there are probably other local minima to be uncovered if we expand this neighborhood. Across the 379 sets of converged results, nine yield GMM objective function below 215.0 and do not meet the first- and second-order conditions.²⁶

From the GMM objective and parameter values it is difficult to determine the unique set of minima; the GMM objective values suggest that there are a number of local minima represented and not one plateau. We plot the own-price elasticities for all sets of parameters in Figure 20. While various sets of parameters appears to correspond to the same local minimum, a significant amount of variation exists. 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 21 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 -\$3.83 billion and -\$6.02 billion; this range significantly differs from the welfare effects implied

²⁴To 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.

²⁵Among the entire set of 500 starting value and algorithm combinations, 216, 193 and 141 have gradient $\infty - norm$ below 30, 20 and 10, respectively.

²⁶The objective function values for these nine points are 125.47, 178.06, 178.15, 187.17, 188.5, 196.7, 204.47, 207.49 and 213.27.

by the best set of results. 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. Of course, we stress that if the consistent route is the global minimum, none of these correspond to the consistent set of parameters.

5.2 Cereal Data

Among the 299 combinations of parameter starting values and algorithms that implied convergence, 140 have gradients with $\infty - norm$ below 30, 126 have gradients with $\infty - norm$ below 20 and 110 have gradients with $\infty - norm$ below 10.²⁷ Only 23 starting value and algorithm pairs with a gradient $\infty - norm$ below 30 meet the second-order conditions. As with the automobile data, our “global” minimum does not meet the first- and second-order conditions; all of the converged sets of result sin the neighborhood of 4.56 have at least one negative eigenvalue.

It is difficult to point out the number of distinct local minima from the parameter estimates, 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 three distinct local minima (Figure 22). The welfare effect of the hypothetical merger between Kellogg’s and General Mills evaluated at the 26 sets of parameters corresponding to local minima do not exhibit the degree of variation we experienced with the automobile data. Figure 23 shows that the welfare effects vary by 50 percent, as opposed to over 100 percent for the automobile data set.

5.3 Discussion of Gradients and 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 a positive eigenvalue. 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 can sometimes have all of their Hessian eigenvalues positive, but other times not. 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

²⁷Across all 500 pairs, 152 have gradients below 30, 134 have gradient $\infty - norm$ below 20 and 115 have gradient $\infty - norm$ below 10.

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 Contraction Mapping Tolerance

Unlike many other nonlinear models, the BLP-type error term is not additively separable in the objective function.²⁸ It is instead retrieved by means of a contraction mapping adding a layer of computational burden, given the linear rate of convergence of the implied fixed-point iterations. As with each iterative procedure, this contraction mapping requires a tolerance level to declare convergence. In the results reported to this point, we adopted a variable contraction mapping tolerance similar to that in Nevo (2000a).^{29,30}

To assess the sensitivity of our findings to the contraction mapping tolerance, we fix the tolerance associated with the automobile and cereal data to 1E-16 and 1E-12, respectively. Using a tolerance of 1E-16 with the cereal data proved to be extremely time consuming, with some sets of starting values taking over 24 hours to converge. Because our goal is to replicate how practitioners employ nonlinear search methods, and our results suggest that many algorithm/starting-value pairs are required for an exhaustive search, we reduced the tolerance to 1E-12 for the cereal data.

For the automobile data, 373 sets of starting values led to convergence. The objective function values lie between 178.1 and over 96,000. The range in objective values among the best sets of results for each algorithm is from 178.1 to 238.66 (restricting to those results that converged). Five of the algorithms converge a point near 178.1; their parameter estimates are extremely similar suggesting this is indeed one local minima. Interestingly, the previous best point, yielding a GMM objective value of 125.5 is no longer uncovered. If we expand to all

²⁸Assume a simple NLS example, where $y = f(\theta, X) + \varepsilon$, for a nonlinear function $f(\cdot)$ of the parameter vector θ and explanatory variables X for a dependent variable y . The error term may be written as $\varepsilon = y - f(\theta, X)$. With a potential abuse of notation, a BLP-type model may be written as $y = f(X, \theta, \varepsilon)$, hence a non-additive error term emerges.

²⁹The code in Nevo (2000a) increases the tolerance (thus, making it more lenient) as the number of iterations in the contraction mapping increase. The contraction mapping takes less time during early iterations when the mean utilities exhibit larger changes.

³⁰For comparison, BLP (1999) use 1E-4.

of the estimates, including those that did not achieve convergence, the Simulated Annealing routine locates a point in the parameter space with an objective value of 133.8. It finds this region of the parameter space only once; no other algorithm finds a GMM objective value under 178 even when we do not restrict ourselves to sets of results that converged.

Many more of the converged sets of results met the first-and second-order condition; 265 have an $\infty - norm$ below 30 with all of their Hessian eigenvalues being positive. Again, the range in objective values (178.1 to 260.6) suggest multiple extrema, not a plateau.

For the cereal data, 257 sets of starting values converged with GMM objective values ranging from 4.56 to over 69,000. The objective value ranges from 4.56 to 327.7 among the best sets of results. Of these converged sets of results, 41 meet the first- and second-order conditions with objective function values ranging from 17.5 to over 11,000. Once again, the region of the parameter space corresponding to an objective function value of 4.56 does not meet the second-order conditions; this is despite the fact that 79 sets of starting values converge in and around this area.³¹

We reproduced all of the tables and figures constructed with variable tolerance level using fixed ones. The results are strikingly similar. To provide an idea to the reader regarding our findings, Figures 24 through 27 repeat the information in Figures 2, 4, 5 and 8, using fixed as opposed to variable tolerance levels. Increasing the contraction mapping tolerance *does not* change the conclusions; the figures are remarkably similar. Algorithms are still likely to converge at multiple places in the parameter space and the variation in elasticities remains large, both for the four specific products and across all products. The summary measures focusing on sets of results that meet the first- and second-order conditions show more variation because of the increase in the number of parameter estimates that satisfy these conditions, as the discussion below illustrates.

For the automobile data, the mean standard deviation of the elasticity estimates across converged sets of results is 11.9; the median is 8.88. For comparison, using variable tolerance, the mean and median are 15.13 and 9.65, respectively. For the cereal data, the average and median within-product-market standard deviation in the estimated elasticities across all converged results are 7.78 and 3.37, respectively. They are almost identical to the mean and median of 7.58 and 3.24, respectively, using variable tolerance.

Using the automobile data, we do find, however, that the truth has changed. More importantly, the elasticity estimates change dramatically. For product 1, the elasticity associated

³¹As before only SolvOpt and Quasi-Newton 2 reach this region.

with the lowest objective value changes from -3.00 to -1.81. Similar changes occur for the other products.

7 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, 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] Amemiya, Takeshi. 1985. *Advanced Econometrics*. Cambridge: Harvard University Press.
- [3] Andrews, Donald W.K. 1997. "A Stopping Rule for the Computation of Generalized Methods of Moments Estimators," *Econometrica*, 65(4): 913:931.
- [4] Angrist, Joshua D., and Alan.B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?." *Quarterly Journal of Economics*, 106(4):979-1014.
- [5] Audet, Charles, and J.E. Dennis JR. 2006. "Mesh Adaptive Direct Search Algorithms for Constrained Optimization." *Siam Journal of Optimization*, 17(1): 188-217
- [6] Bayer, Patrick and Robert McMillan. 2006. "Racial Sorting and Neighborhood Quality." NBER Working Paper #11813.
- [7] Bayer, Patrick, Robert McMillan and Kim Rueben. 2004. "An Equilibrium Model of Sorting in an Urban Housing Marke." NBER Working Paper #10865.
- [8] Bekker, Paul A. 1994. "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica*, 62(3): 657-681.
- [9] Berry, Steven. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics*, 25(2): 242-262.
- [10] Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841-890.
- [11] Berry, Steven, James Levinsohn, and Ariel Pakes. 1999. "Voluntary Export Restraints on Automobiles. Evaluating a Trade Policy." *American Economic Review*, 89(3): 400-430.
- [12] 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.
- [13] Brenkers, Randy, and Frank Verboven. 2006. "Liberalizing a distribution system: The European Car Market." *Journal of the European Economic Association*, 4(1): 216-251.

- [14] Burke, James V., Adrian S. Lewis, and Michael L. Overton. 2007. "The Speed of Shor's R-algorithm". Manuscript.
- [15] 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.
- [16] Cameron, A. Colin., and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] Hansen, Lars P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, 50(4):1029-1054.
- [22] Hendel, Igal and Aviv Nevo. 2006. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica*, 74(6): 1637-1673.
- [23] Gowrisankaran, Gautam and Marc Rysman. 2007. "Dynamics of Consumer Demand for New Durable Goods." mimeo.
- [24] Hasting, Justine, Thomas Kane, and Douglas Staiger. 2007. "Parental Preferences and School Competition: Evidence from a Public School Choice Program." NBER Working Paper #11805.
- [25] Judd, Kenneth L. 1998. *Numerical Methods in Economics*. Cambridge: MIT Press.
- [26] Kappel, Franz, and Alexei Kuntsevich. 2000. "An Implementation of Shor's r-Algorithm." *Computational Optimization and Applications*, 15(2): 193-205.

- [27] 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.
- [28] 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.
- [29] McFadden, Daniel L. 1981. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data*, ed. C.F Manski and D.L. McFadden, 198-272. Cambridge, MIT Press.
- [30] McFadden, Daniel L., and Whitney K. Newey. 1994. "Large Sample Estimation and Hypothesis Testing." In *Handbook of Econometrics*, ed. R.F Engle and D.L. McFadden, 2113-2245.
- [31] Miranda, Mario J., and Paul L. Fackler. 2002. *Applied Computational Economics and Finance*. Cambridge: MIT press.
- [32] 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.
- [33] 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.
- [34] Nevo, Aviv. 2000b. "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry." *RAND Journal of Economics*, 31(3): 395-421.
- [35] Nevo, Aviv. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, 69(2): 307-342.
- [36] Nevo, Aviv. 2003. "New Products, Quality Changes, and Welfare Measures from Estimated Demand Systems." *Review of Economics and Statistics*, 85(2): 266-275.
- [37] Pagan, Adrian R., and J.C. Robertson. "GMM and its Problems." Manuscript, Australia National University.
- [38] Petrin, Amil. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy*, 110(4): 705-729.

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

A Figures

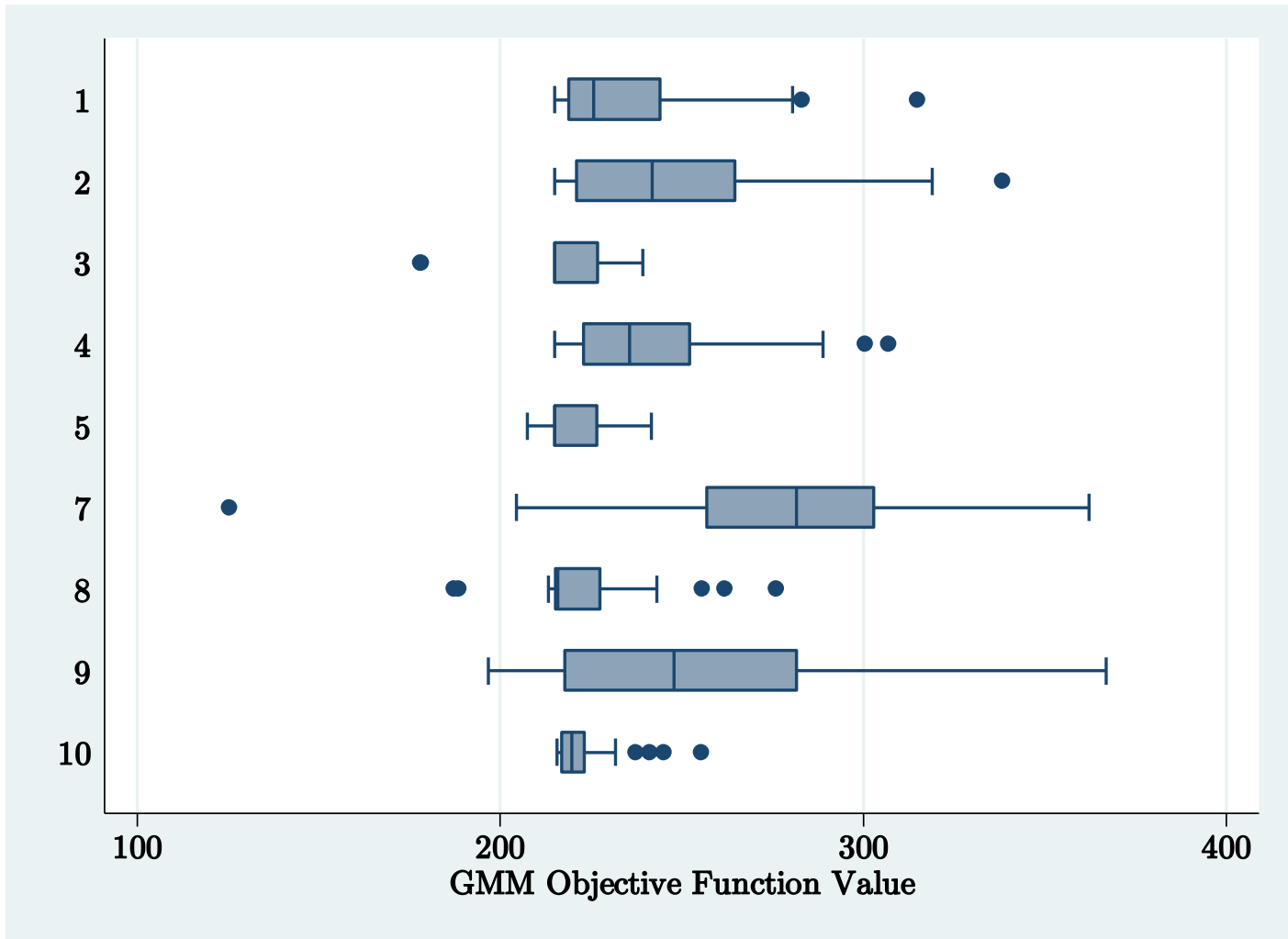


Figure 1: GMM objective values for converged algorithms using the automobile data. This truncates the upper 10% of the converged GMM objective values. Algorithm 6 (JBES Genetic Algorithm) never converges.

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”.

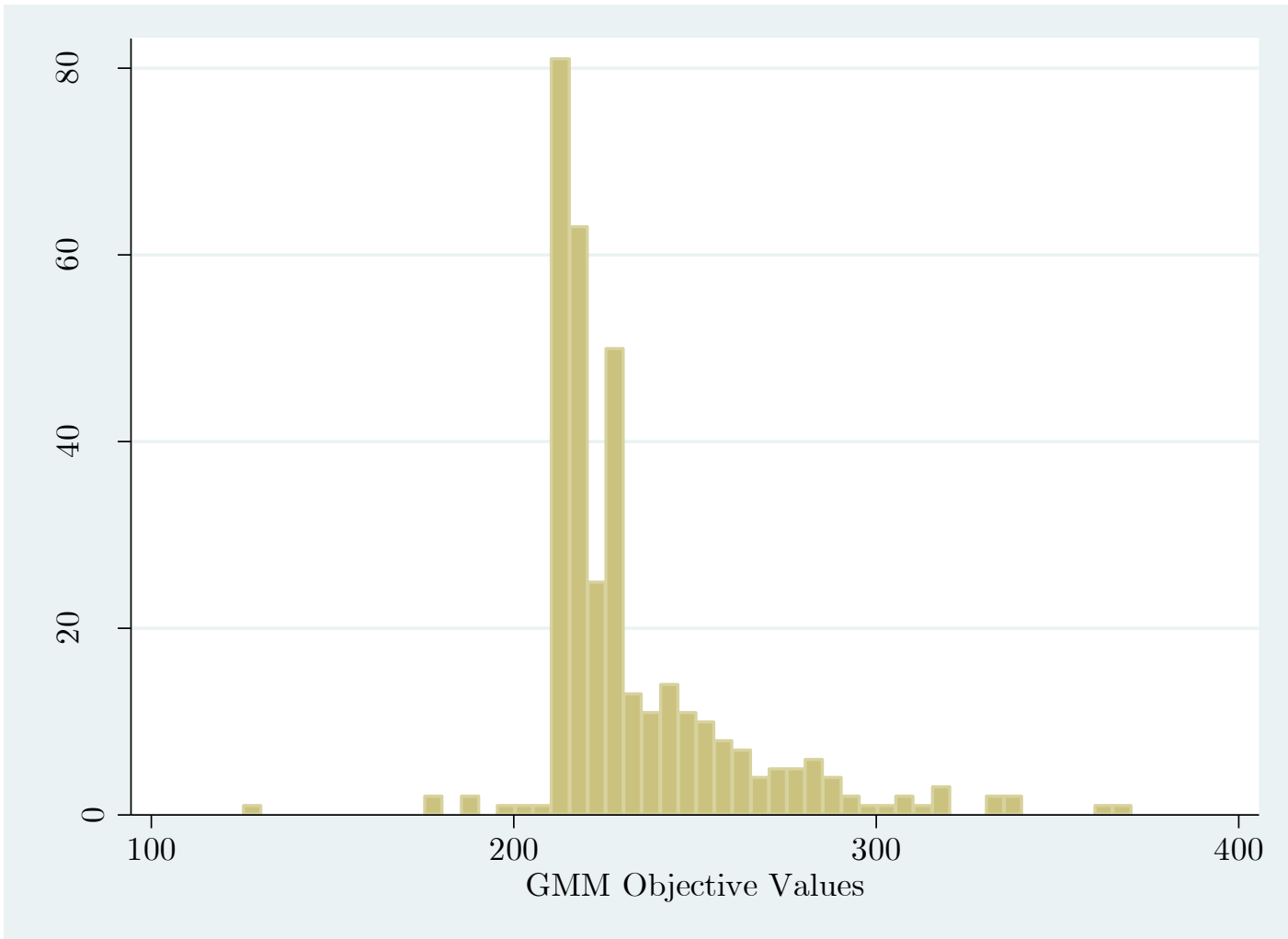


Figure 2: A histogram of GMM objective values for converged algorithms using the automobile data. This truncates the upper 10% of the converged GMM objective values.

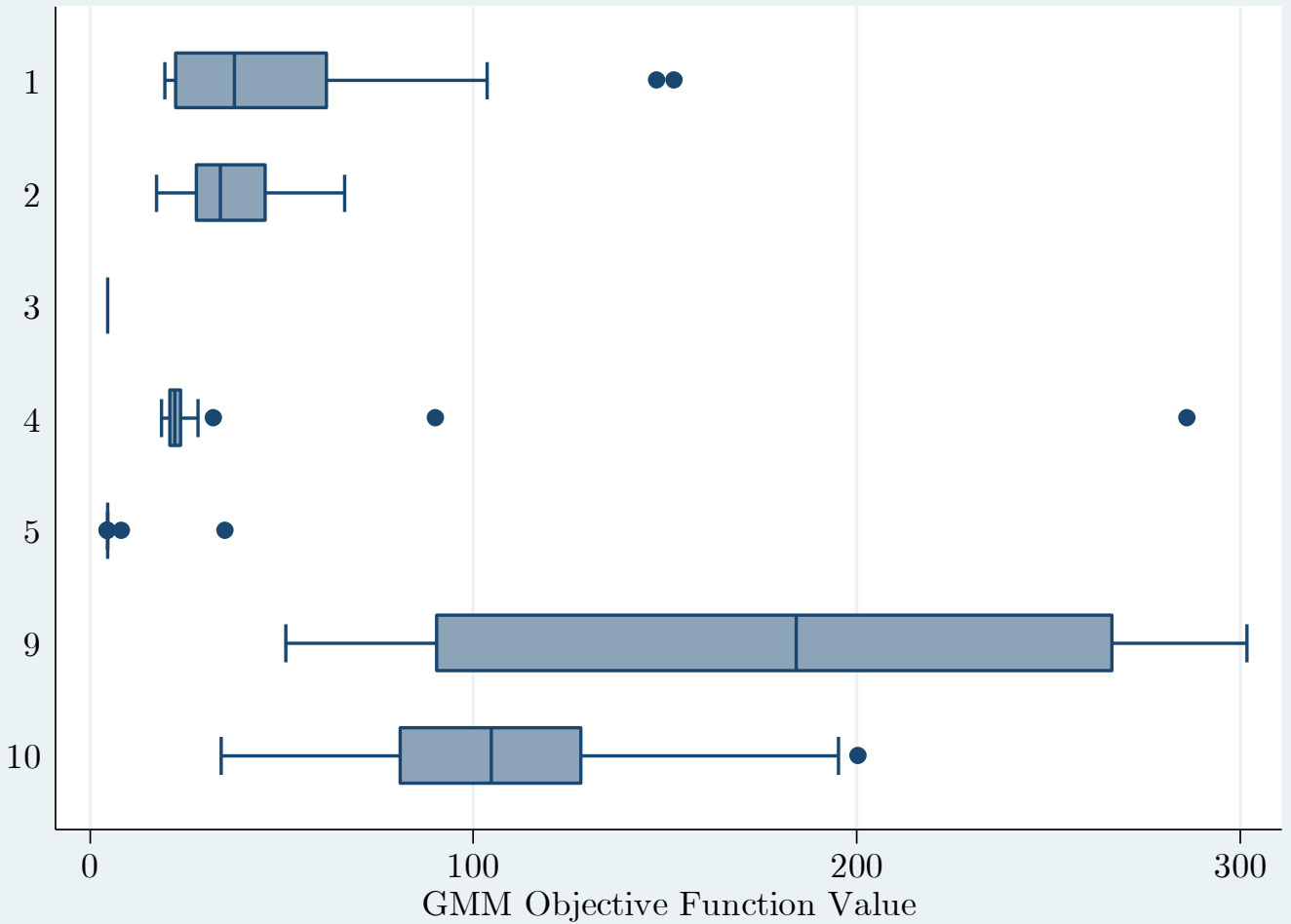


Figure 3: GMM objective values for converged algorithms using the cereal data. This truncates the upper 25% of the converged GMM objective values. Algorithms 6 (JBES Genetic Algorithm) and 7 (Simulated Annealing) never converge. Algorithm 8 (MADS) converges twice, but at objective values above the 75th percentile. A large fraction of the GMM objective values associated with non-converged sets of parameters for algorithms 7 and 8 lie below 300, but are omitted.

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”.

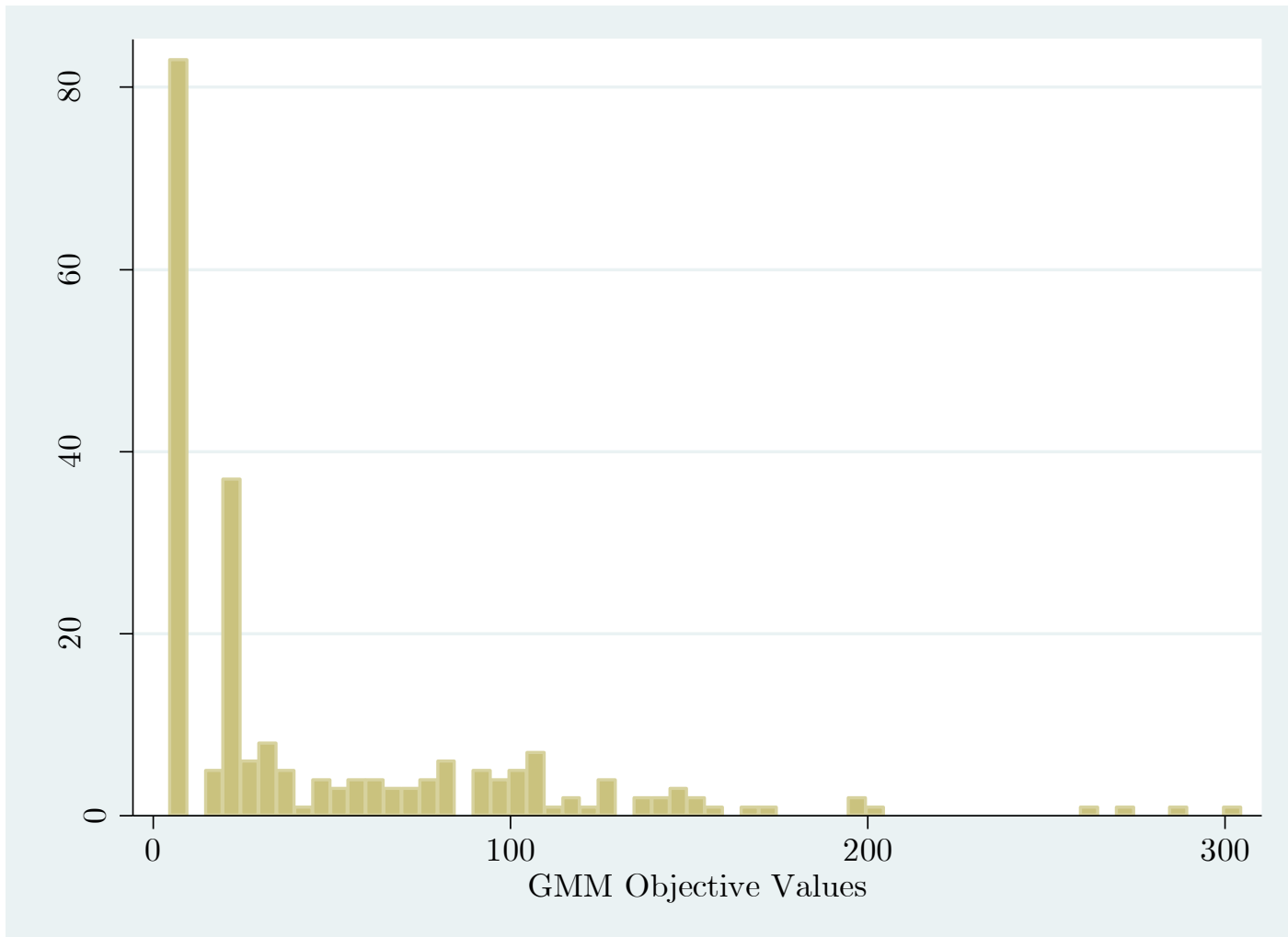


Figure 4: A histogram of GMM objective values for converged algorithms using the cereal data. This truncates the upper 25% of the converged GMM objective values.

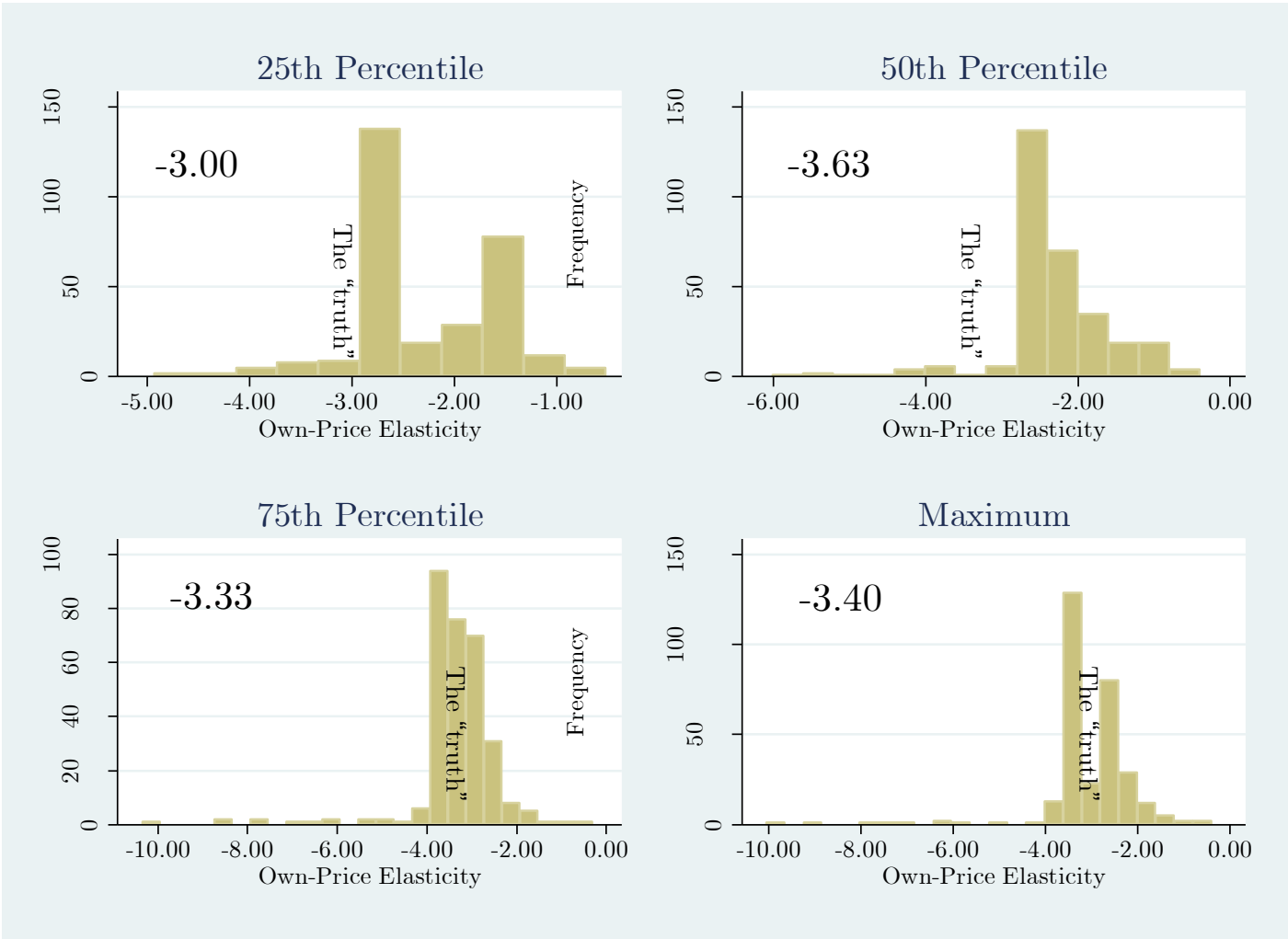


Figure 5: Histogram of candidate set of own-price elasticities for four products for the automobile data. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.

The graphs omit all sets of results that converged but had GMM objective values in the upper decile. The graphs also truncate above at zero and below the 10th percentile.

The average within-product-market standard deviation in the estimated elasticities across all converged results is 15.13; the median is 9.65. The average own-price elasticity is -6.12. Among the “best” sets of parameter values, the mean standard deviation is 1.75, the median is 1.19, and the mean own-price elasticity is -2.92.

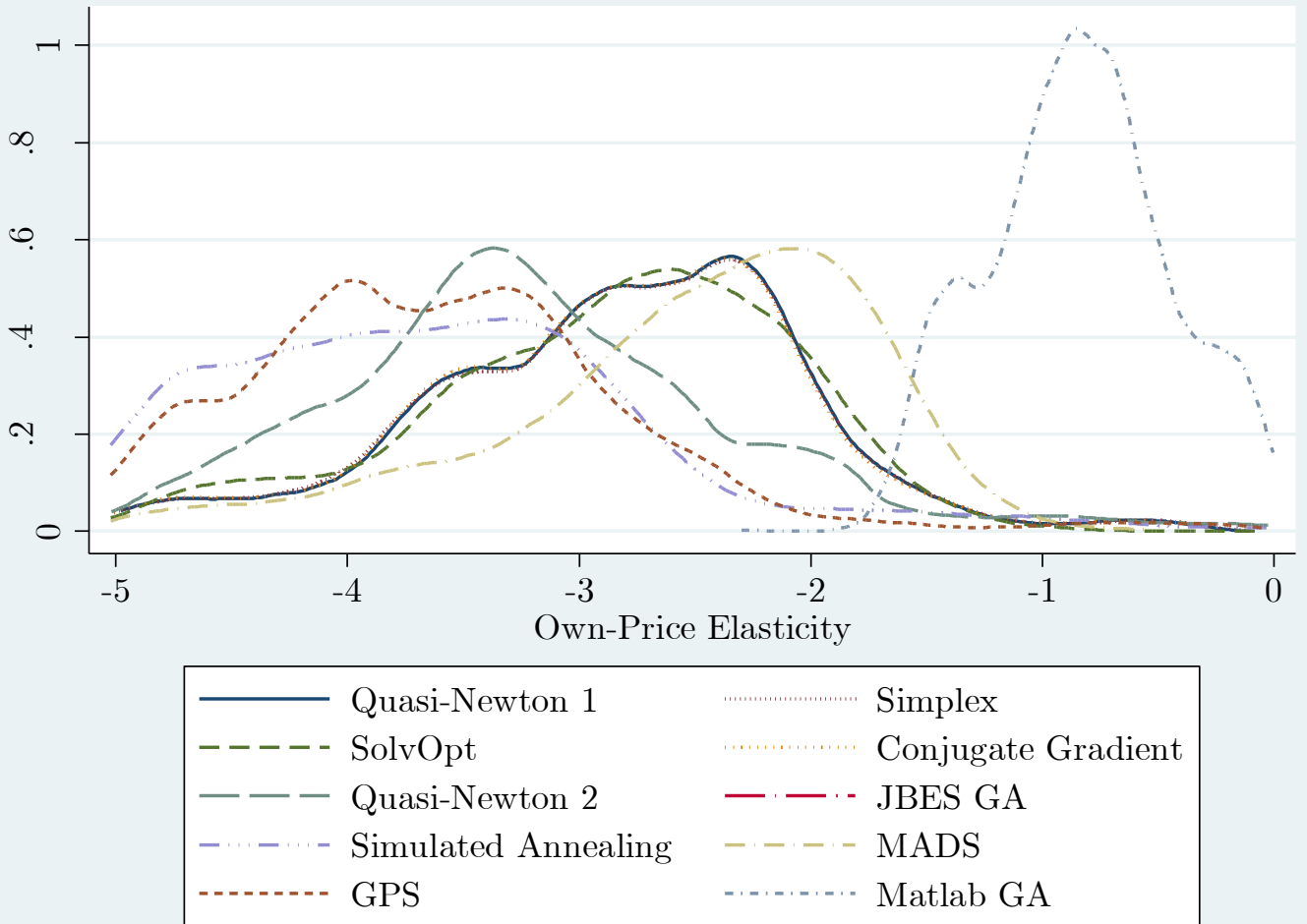


Figure 6: Density estimates of the own-price elasticities across all products in the automobile data, using the “best” set of parameters for each algorithm.

These densities truncate the lower 10% among converged sets of results.

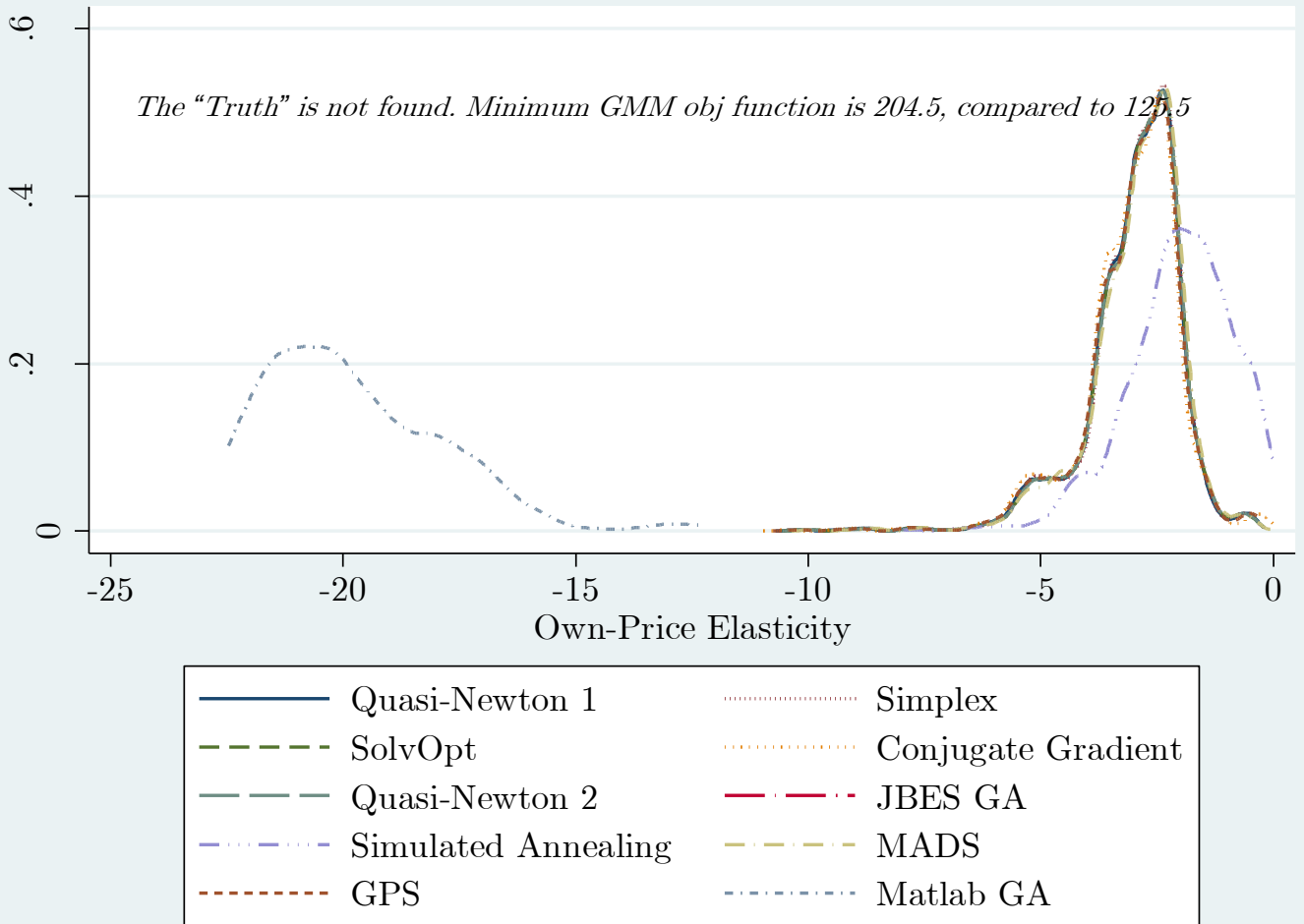


Figure 7: Density estimates of the own-price elasticities across all products in the automobile data, using the “best” set of parameters for each algorithm. This assumes the research tries 20 starting values, instead of 50.

These densities truncate the lower 10% among converged sets of results using the first 20 sets of starting values.

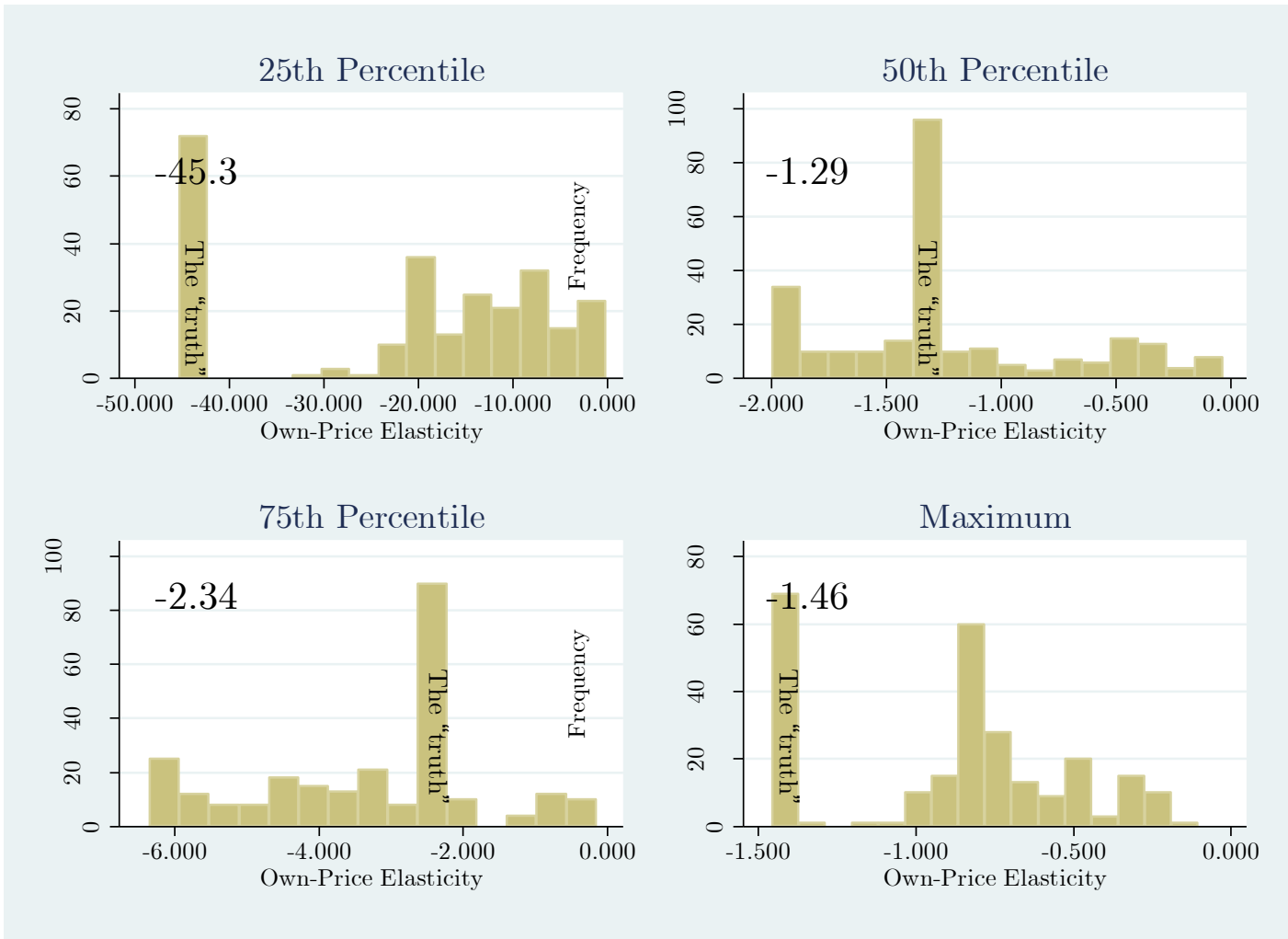


Figure 8: Histogram of candidate set of own-price elasticities for four products for the cereal data. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.

The graphs omit all sets of results that converged but had GMM objective values in the upper decile. The graphs also truncate above at zero and below the 10th percentile.

The average within-product-market standard deviation in the estimated elasticities across all converged results is 7.50; the median is 3.24. The average own-price elasticity is -9.93. Among the “best” sets of parameter values, the mean standard deviation is 4.92, the median is 1.87, and the mean own-price elasticity is -9.96.

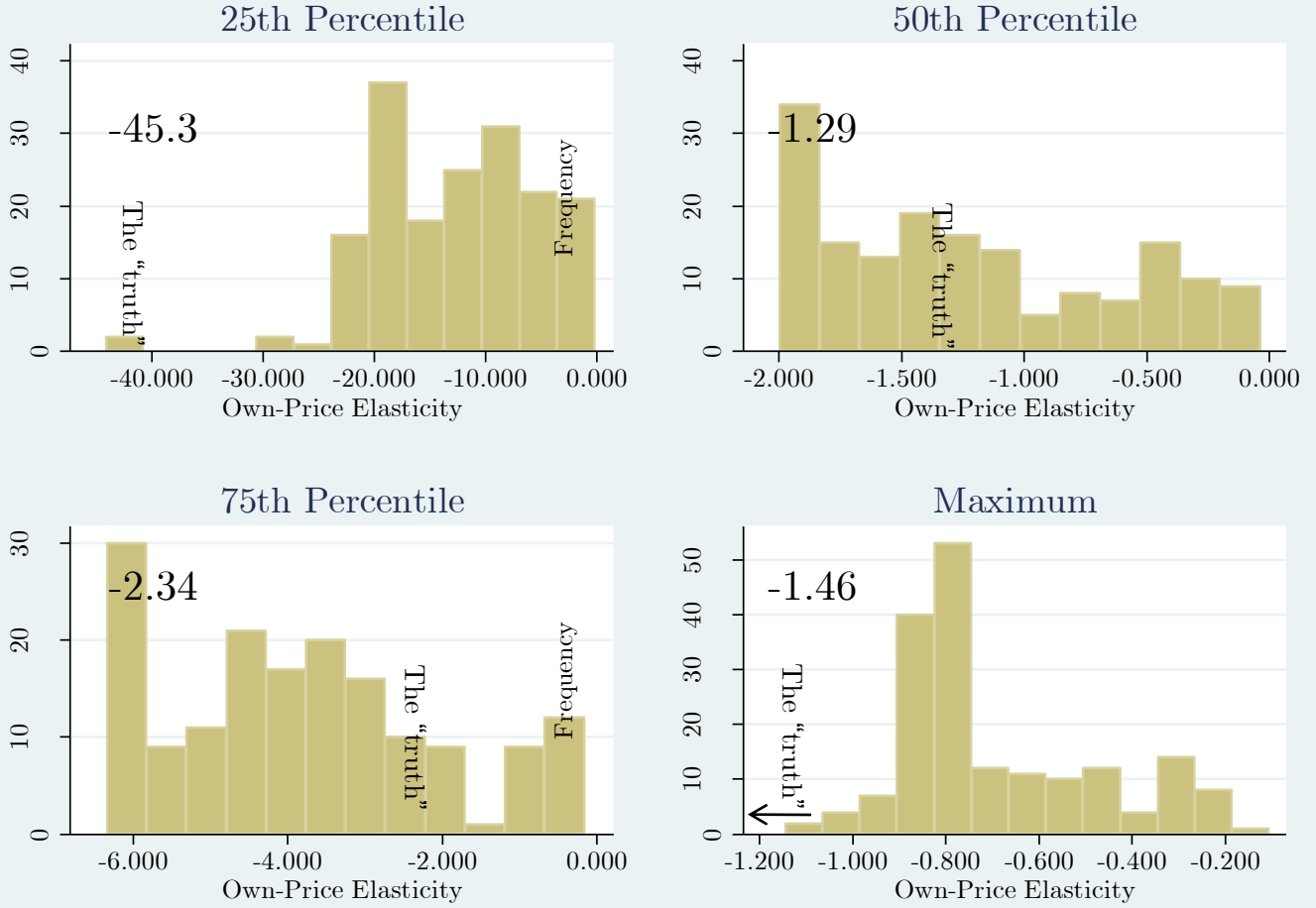


Figure 9: Histogram of candidate set of own-price elasticities for four products for the cereal data without using algorithms 3 and 5. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.

The average within-product-market standard deviation in the estimated elasticities across all converged results is 7.50; the median is 3.24. The average own-price elasticity is -9.93. Among the “best” sets of parameter values, the mean standard deviation is 4.92, the median is 1.87, and the mean own-price elasticity is -9.96.

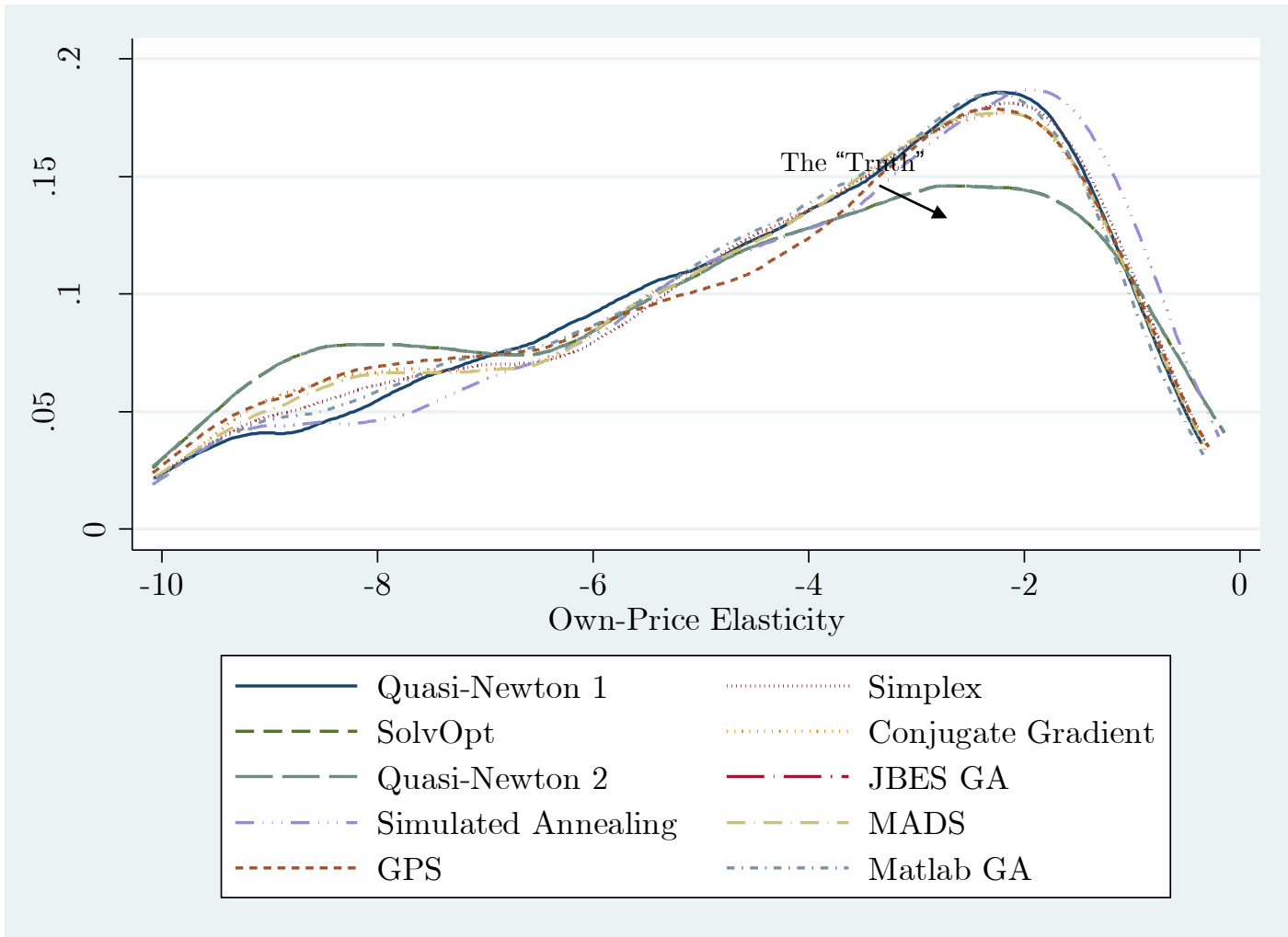


Figure 10: Density estimates of the own-price elasticities across all products in the cereal data, using the “best” set of parameters for each algorithm.

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

These densities truncate the lower 10% among converged sets of results.

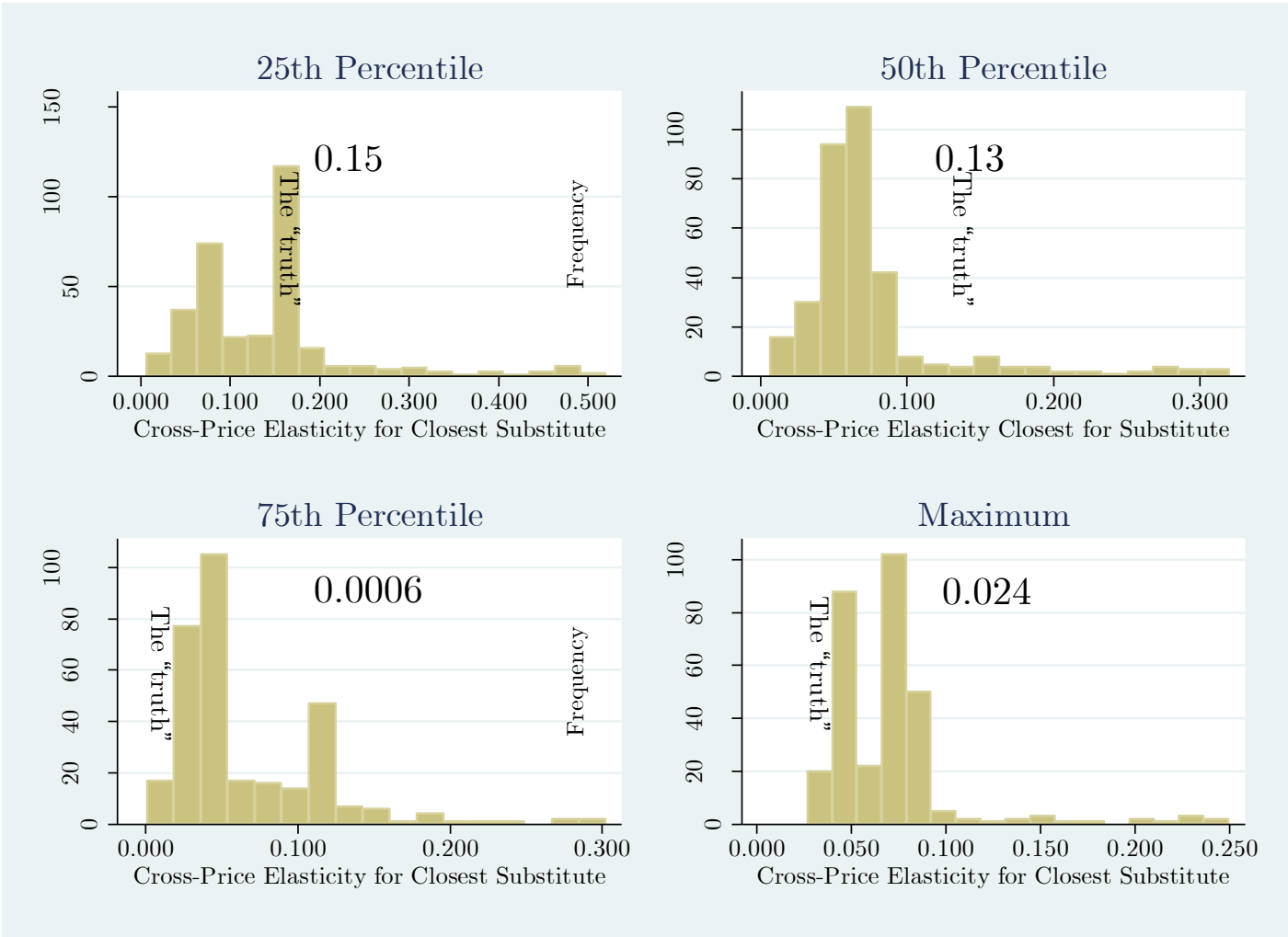


Figure 11: Histogram of candidate set of cross-price elasticities for four products in the automobile data. The cross product is chosen as the product that is the closest substitute across all sets of results that led to convergence. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value. All graphs truncate the upper 10% of values and below zero.

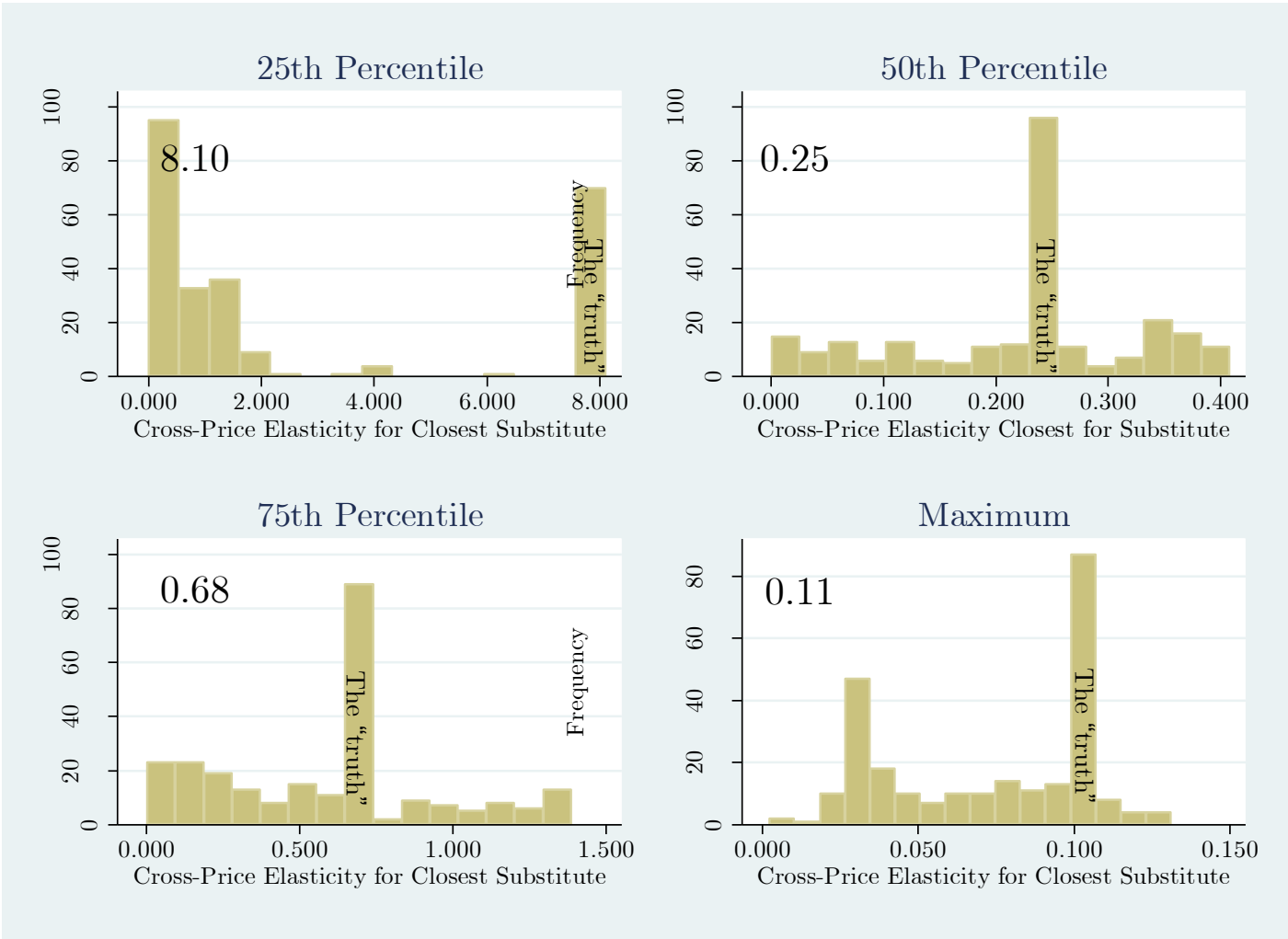


Figure 12: Histogram of candidate set of cross-price elasticities for four products in the cereal data. The cross product is chosen as the product that is the closest substitute. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.

These graphs truncate the largest 5 percent of elasticities and the worst 10% of results in terms of the GMM objective function.

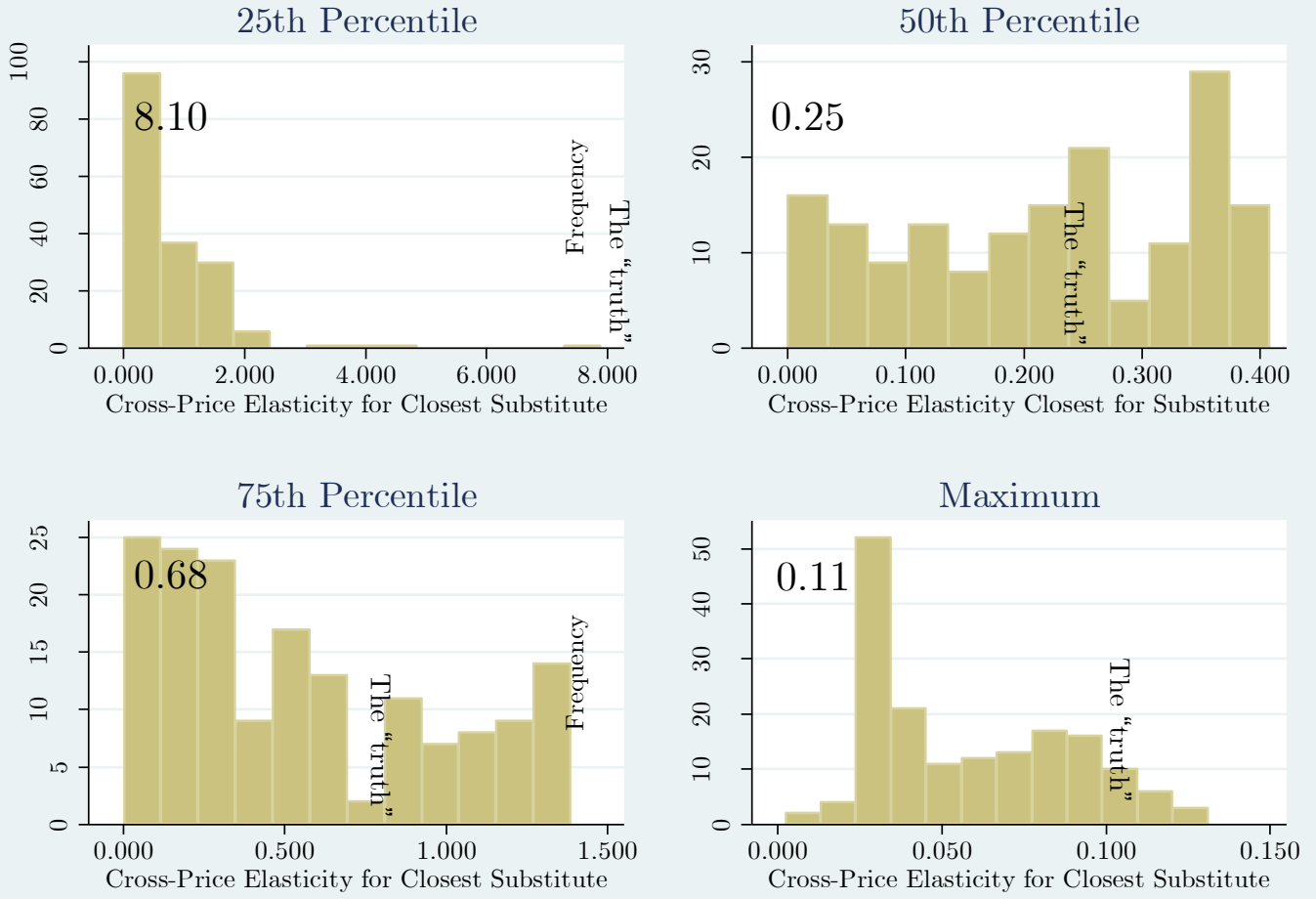


Figure 13: Histogram of candidate set of cross-price elasticities for four products in the cereal data without using algorithms 3 and 5. The cross product is chosen as the product that is the closest substitute. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.

These graphs truncate the largest 5 percent of elasticities and the worst 10% of results in terms of the GMM objective function.

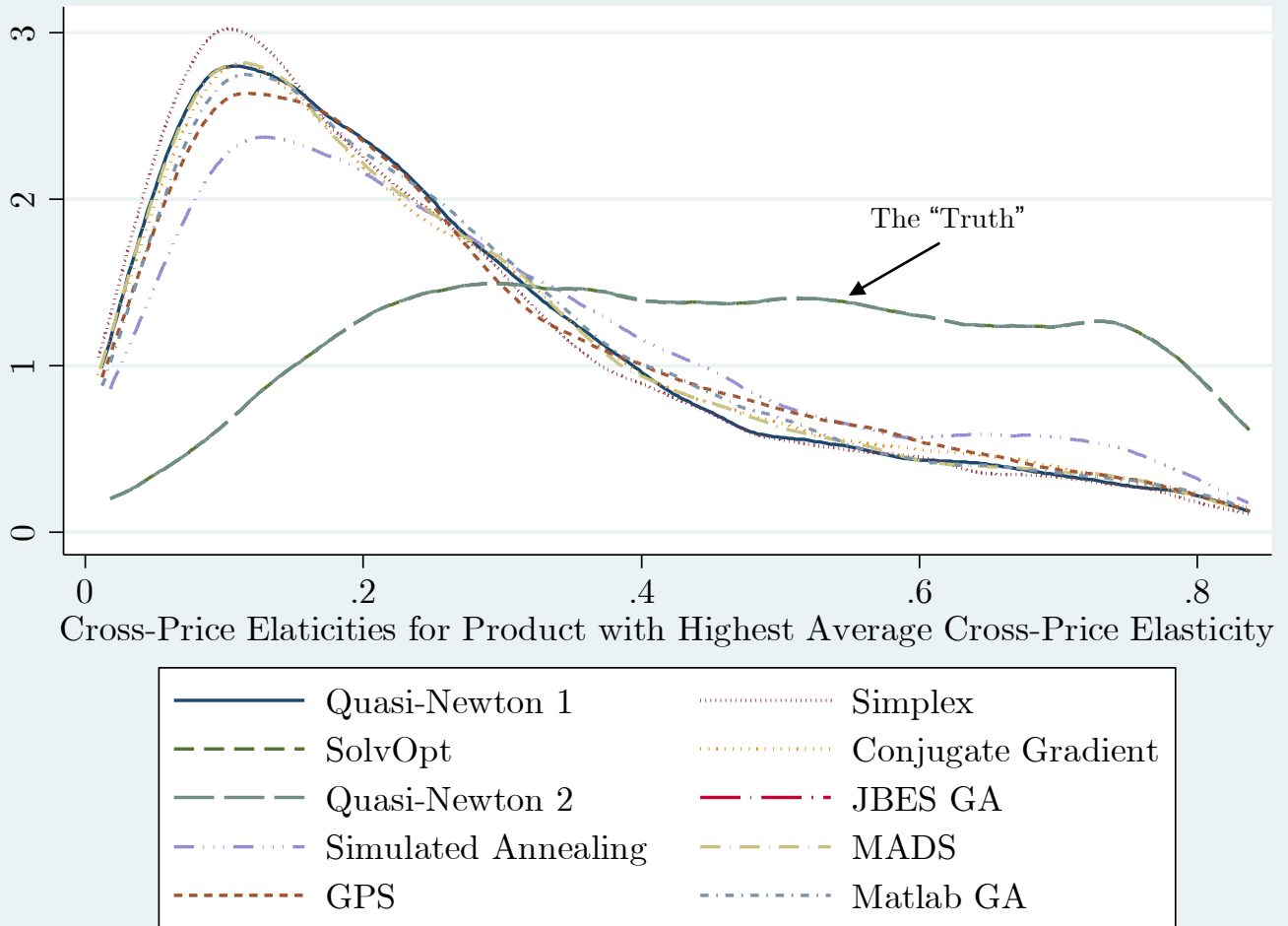


Figure 14: Density estimates of the cross-price elasticities across all products in the cereal data for the product that has the highest average cross-price elasticity. These results use the “best” set of parameters for each algorithm and truncate at zero and above the 75th percentile.

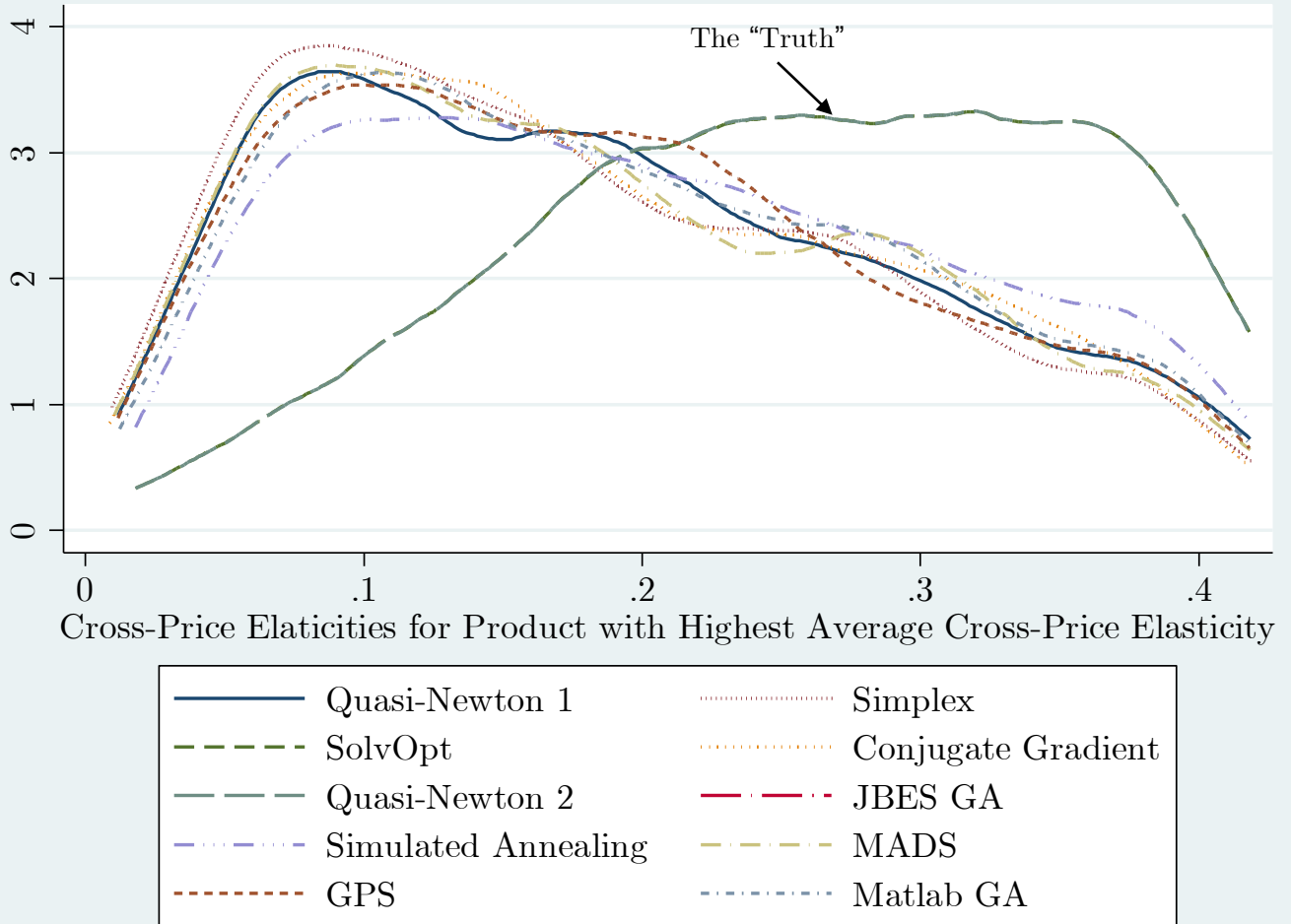


Figure 15: Density estimates of the cross-price elasticities across all products in the cereal data for the product that has the highest average cross-price elasticity focusing on the elasticities ranging from 0 to one half of the 75th percentile. These results use the “best” set of parameters for each algorithm.

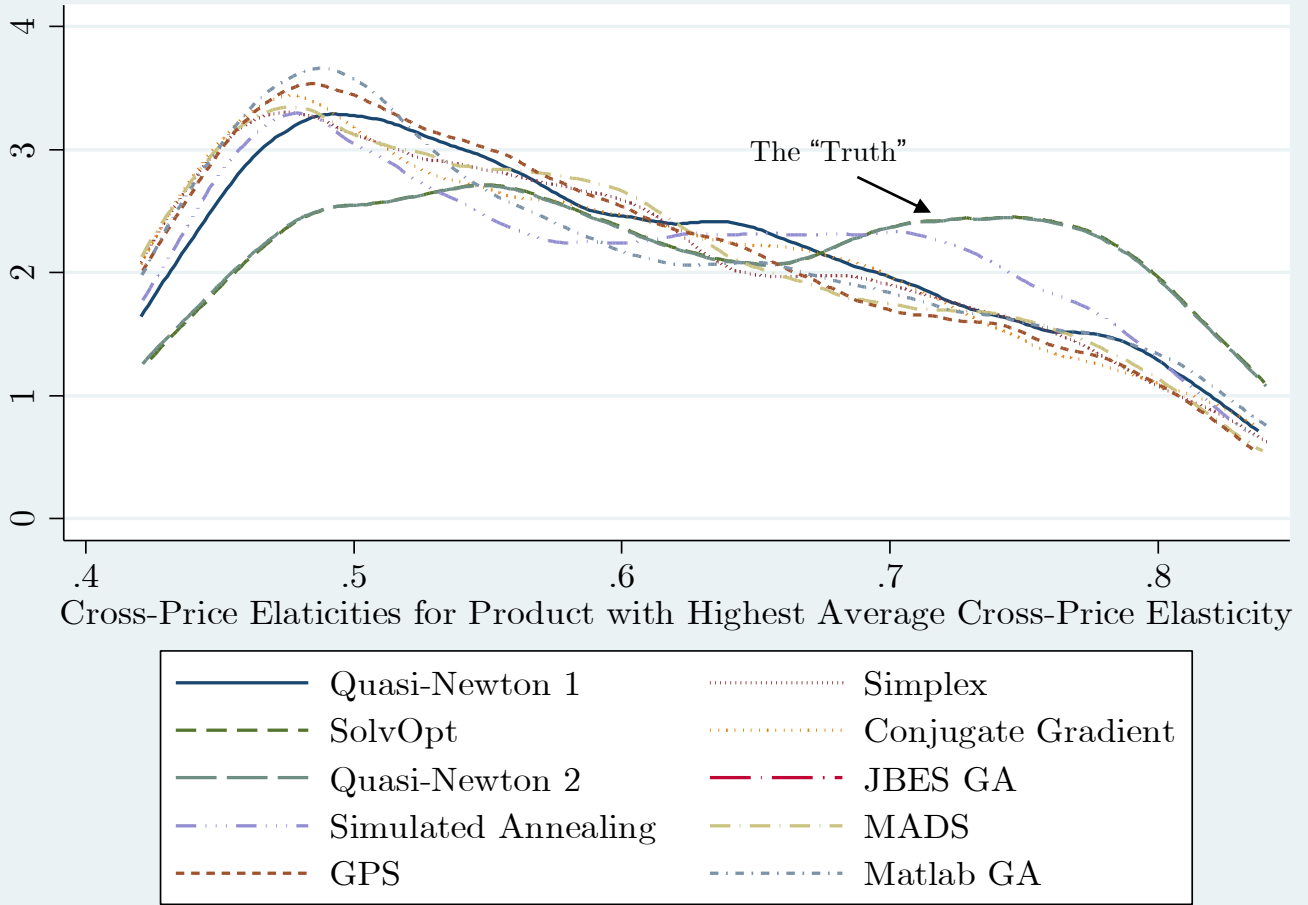


Figure 16: Density estimates of the cross-price elasticities across all products in the cereal data for the product that has the highest average cross-price elasticity focusing on the elasticities ranging from one half of the 75th percentile to the 75th percentile. These results use the “best” set of parameters for each algorithm.

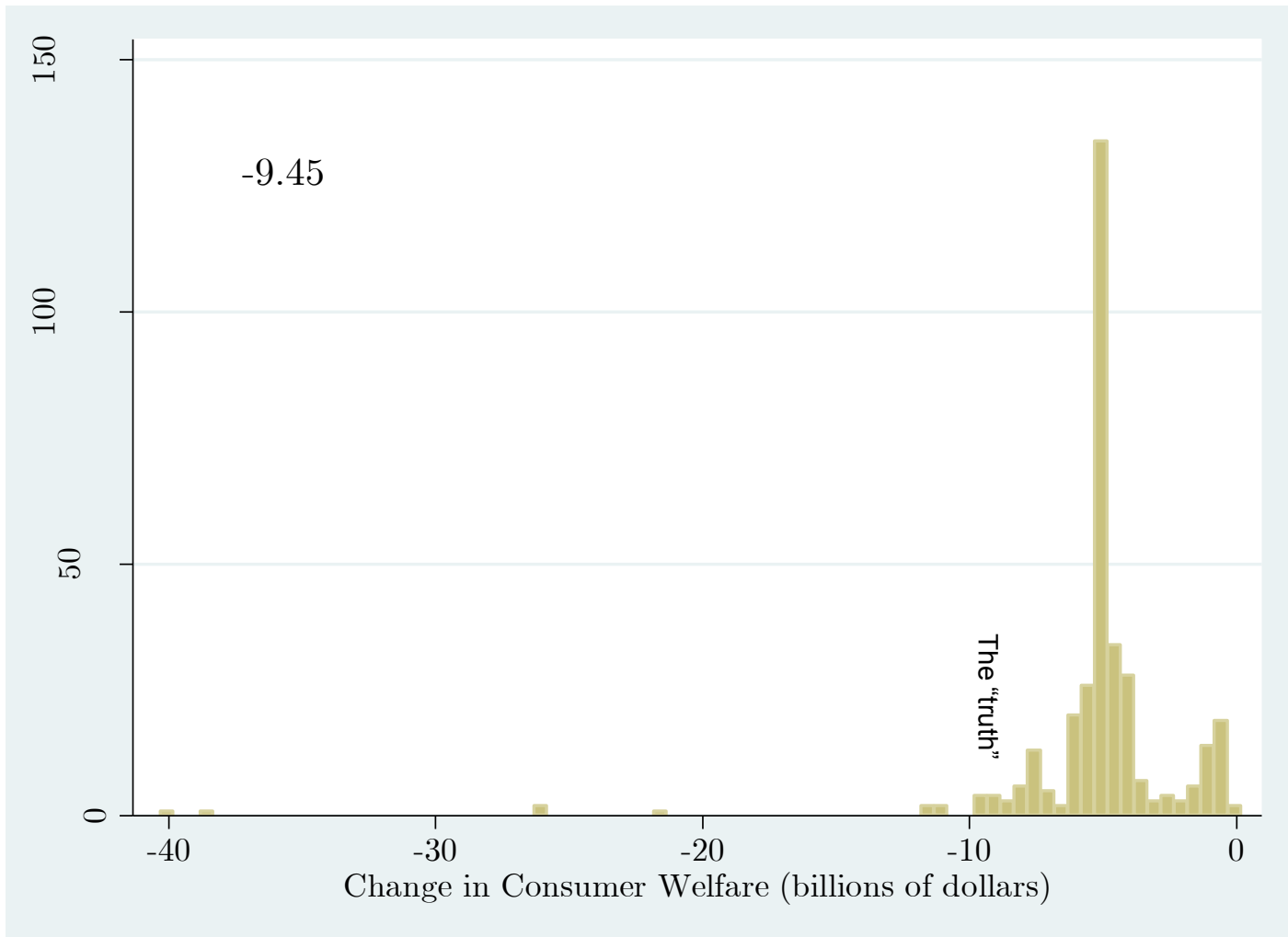


Figure 17: Histogram of estimated change in total welfare from hypothetical merger using the automobile data. This truncates below at the lower 1% and above at 0. The range across eight “best” estimates is -52.76 to +7.05, the upper and lower range among the best estimates are truncated.

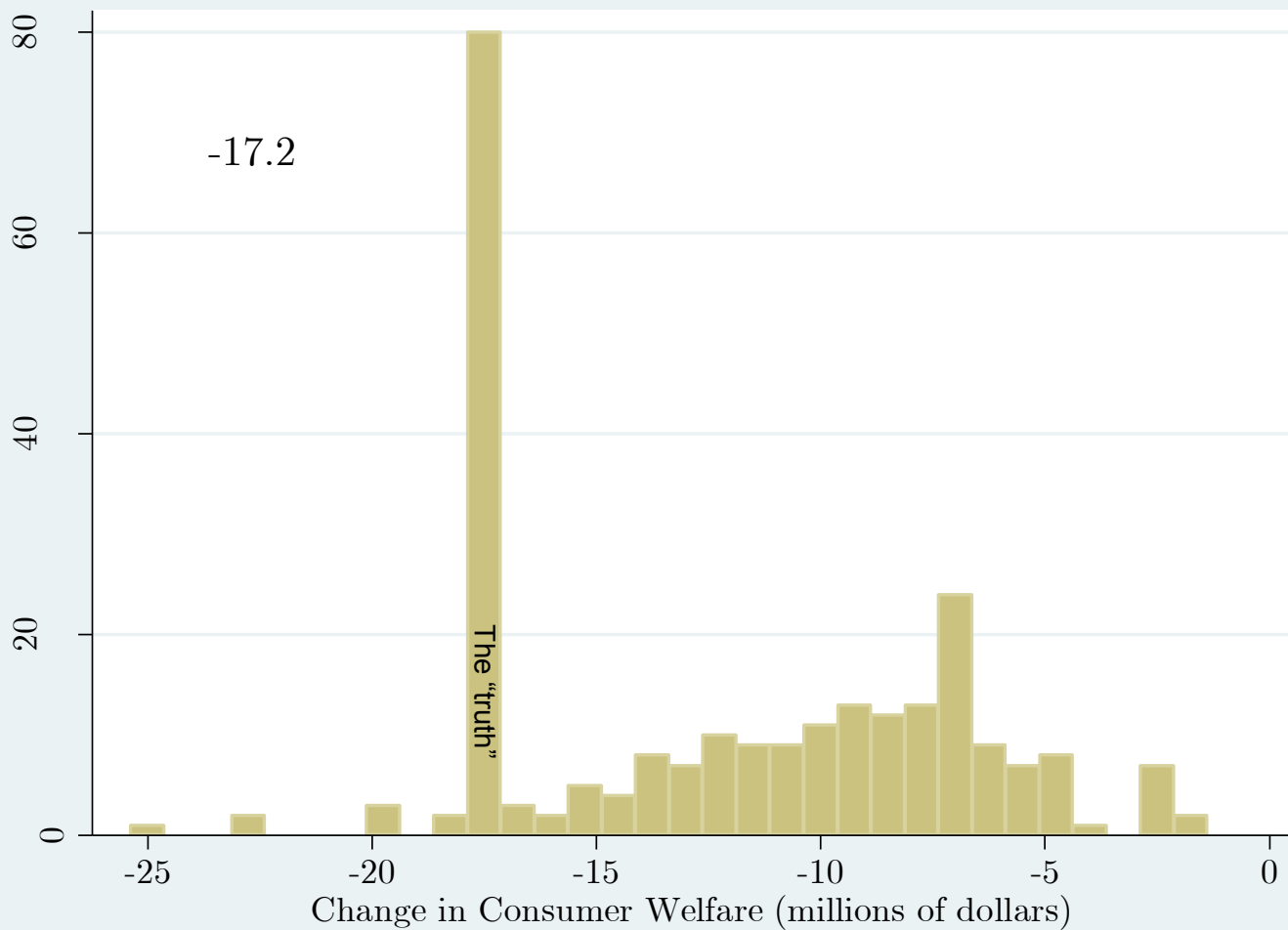


Figure 18: Histogram of estimated change in total welfare from hypothetical merger using the cereal data. The range across nine “best” estimates is -8.66 to -17.2.

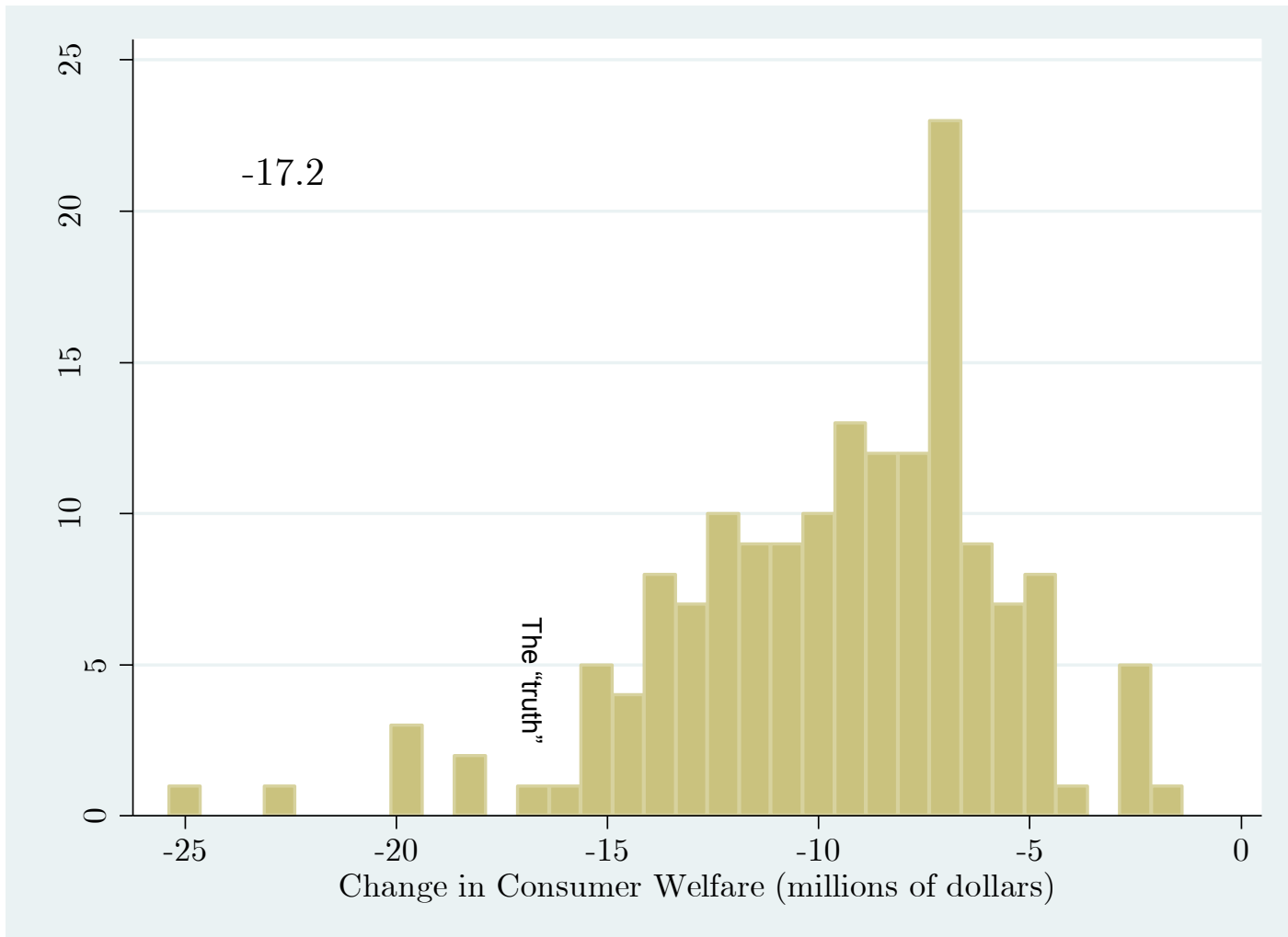


Figure 19: Histogram of estimated change in total welfare from hypothetical merger using the cereal data when algorithms 3 and 5 are not used. The range across eight “best” estimates is -8.66 to -17.2.

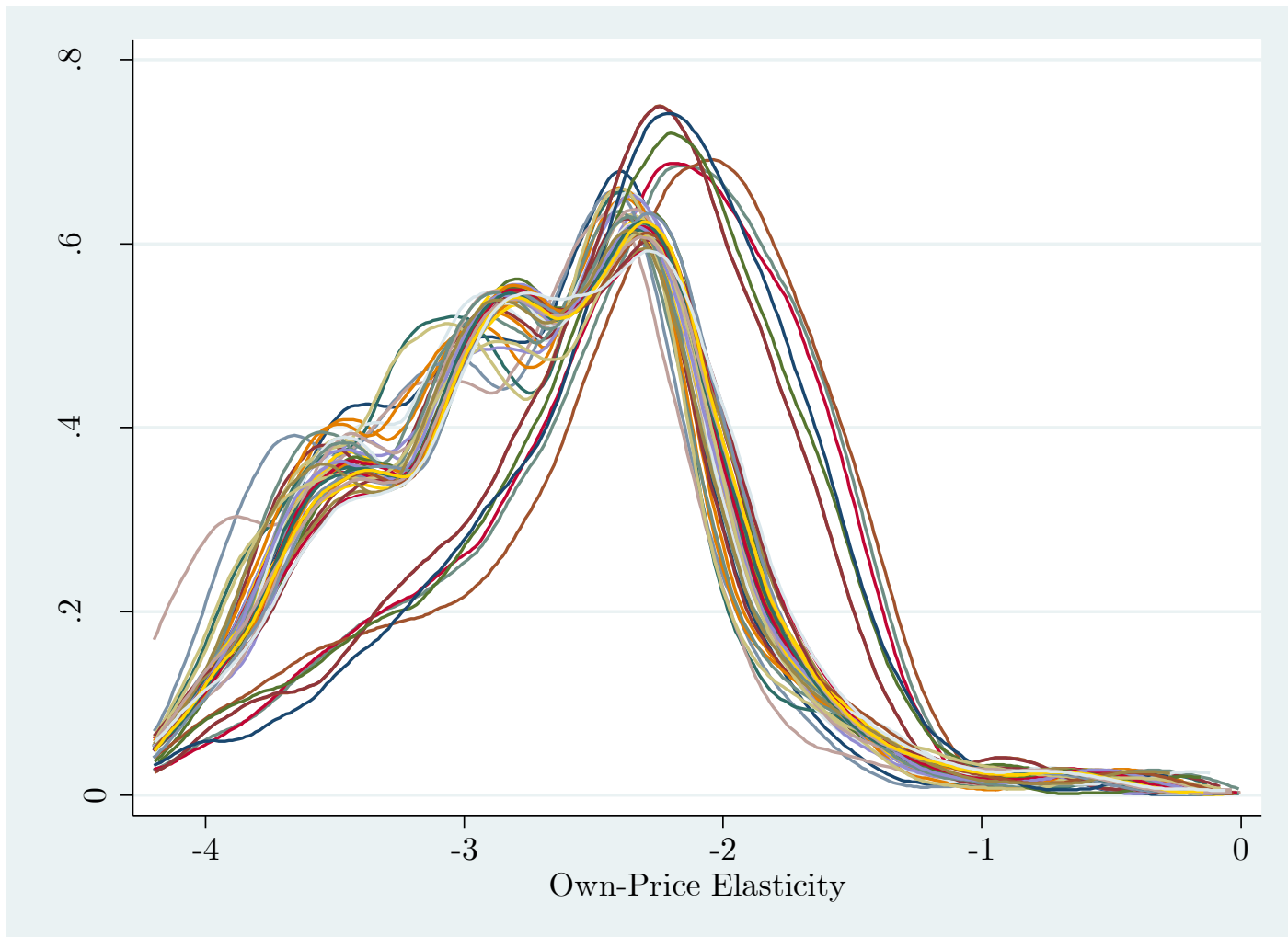


Figure 20: Density estimates of the estimated own-price elasticities for the 60 local minima for the automobile data. This does not include the set of parameters leading to the lowest GMM objective function uncovered in our exercise.

These densities truncate the lower 10% of elasticities and omit minima with objective values above 300.

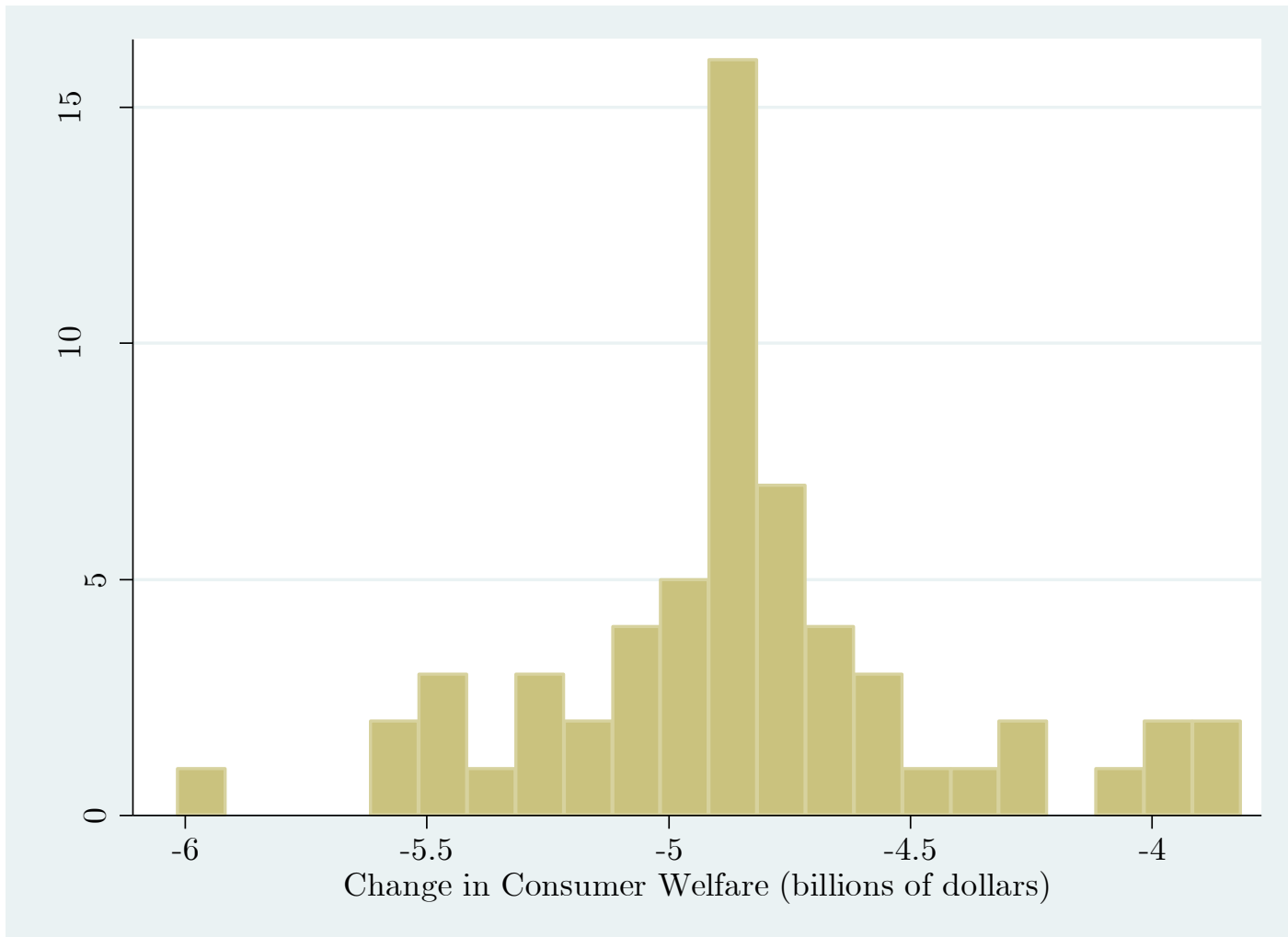


Figure 21: Histogram of estimated change in total welfare from hypothetical merger using the automobile data for the 60 local minima. The range is from -\$3.83 billion to -\$6.01 billion. This does not include the set of parameters leading to the lowest GMM objective function uncovered in our exercise.

The set of parameters leading to lowest GMM objective function resulted in a welfare change of -\$9.45 billion.

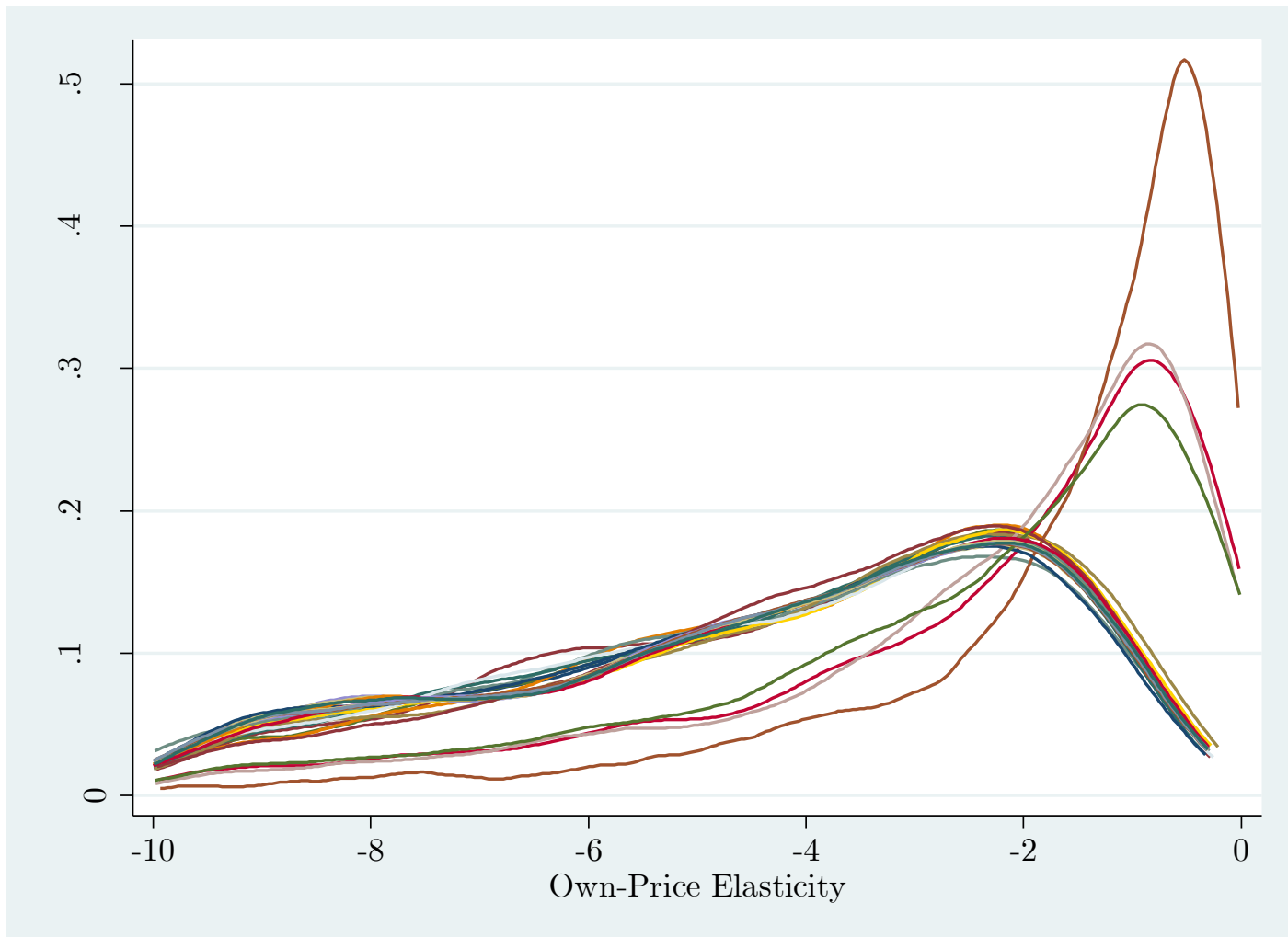


Figure 22: Density estimates of the estimated own-price elasticities for the 26 local minima for the cereal data. These are truncated at -10.

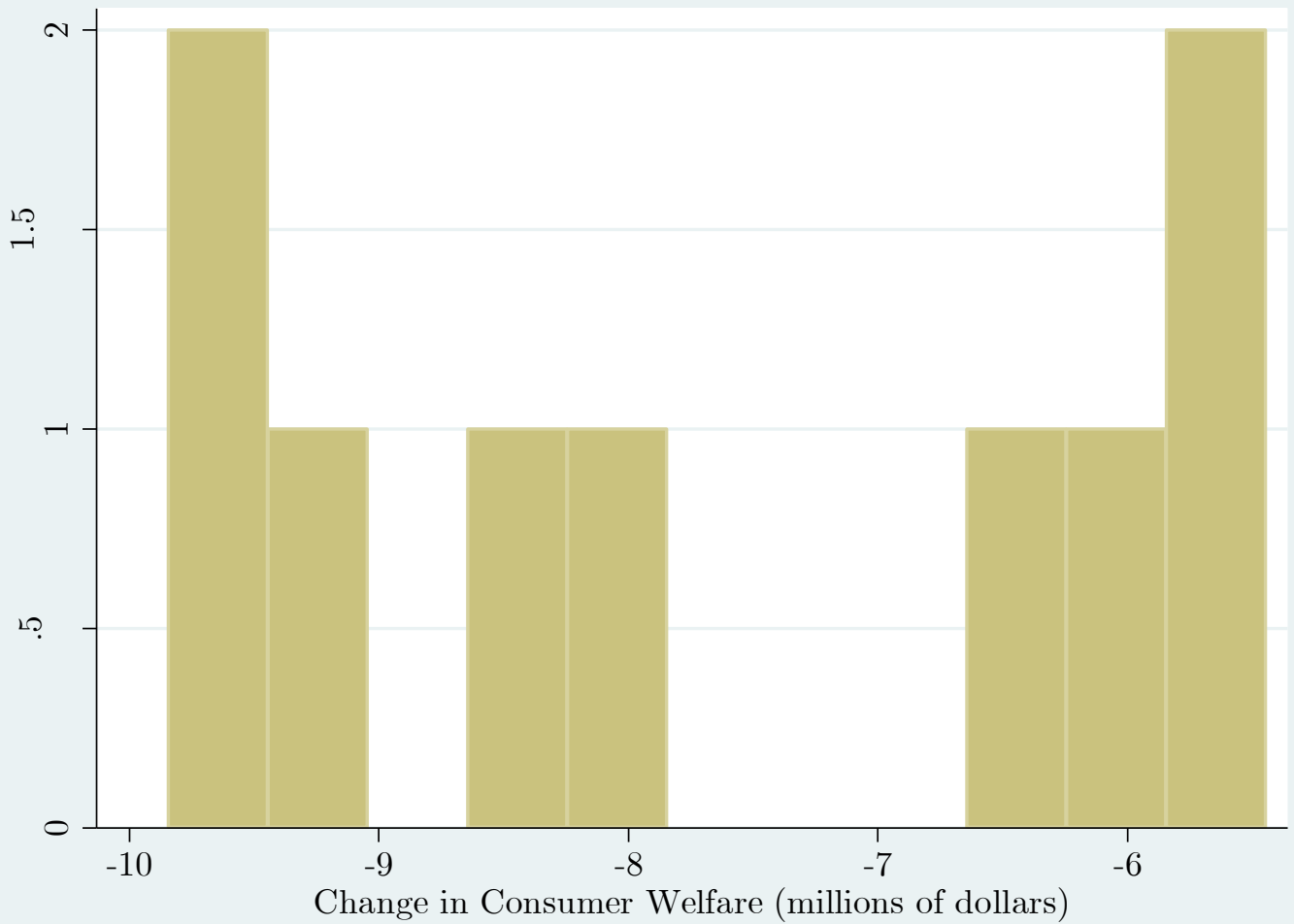


Figure 23: Histogram of estimated change in total welfare from hypothetical merger using the cereal data for the 26 local minima. These results truncate above 0.

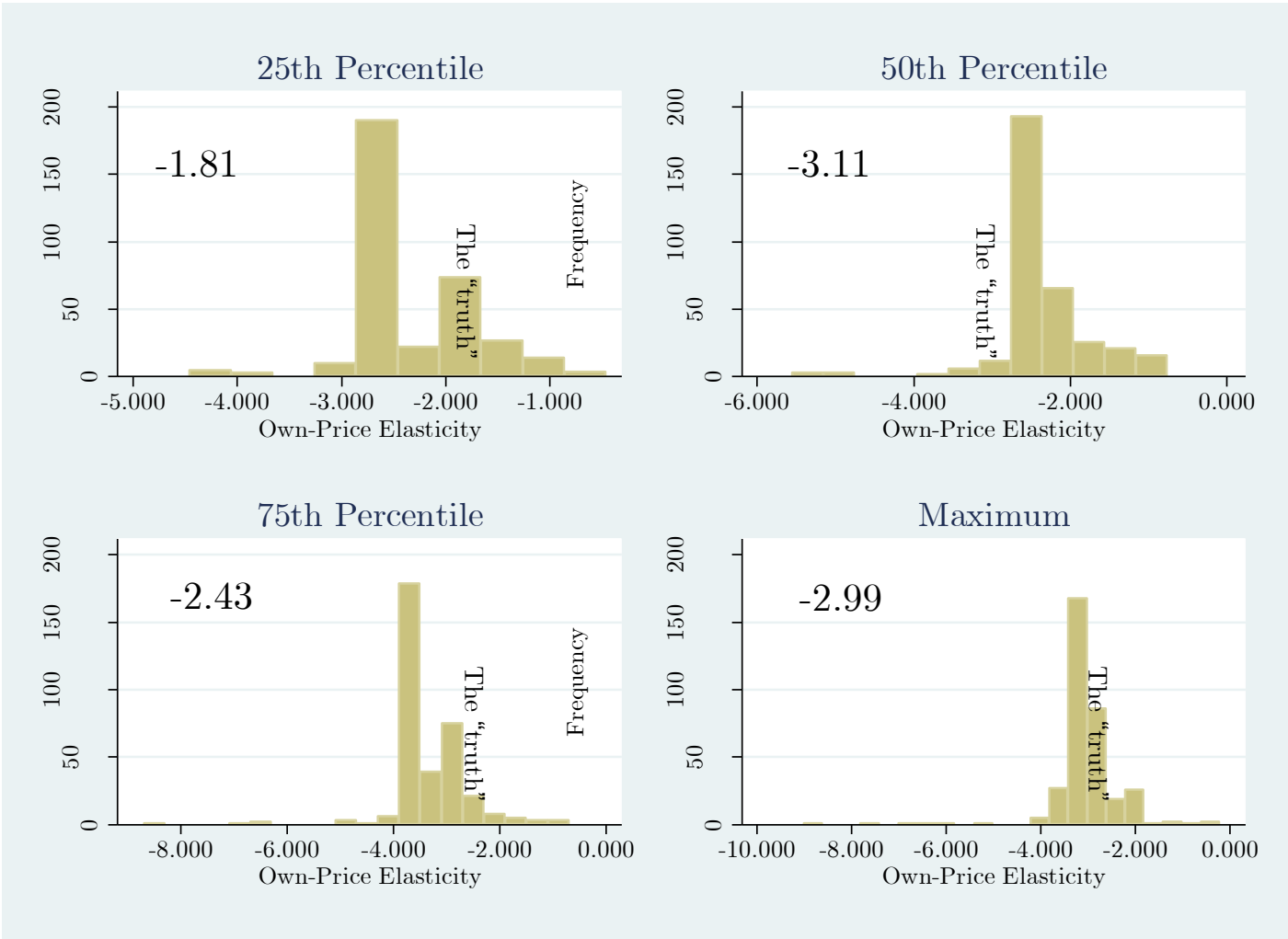


Figure 24: Histogram of candidate set of own-price elasticities for four products for the automobile data using a contraction mapping tolerance of 10^{-16} . The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.

These use roughly the same bin width and horizontal scale as in Figure 5.

The average within-product-market standard deviation in the estimated elasticities across all converged results is 11.9; the median is 8.88. For comparison using 10^{-8} as the tolerance, the mean and median are 15.13 and 9.65, respectively.

The true elasticities using 10^{-8} as the tolerance are -3.00, -3.63, -3.33, and -3.40 for products 1, 2, 3 and 4, respectively.

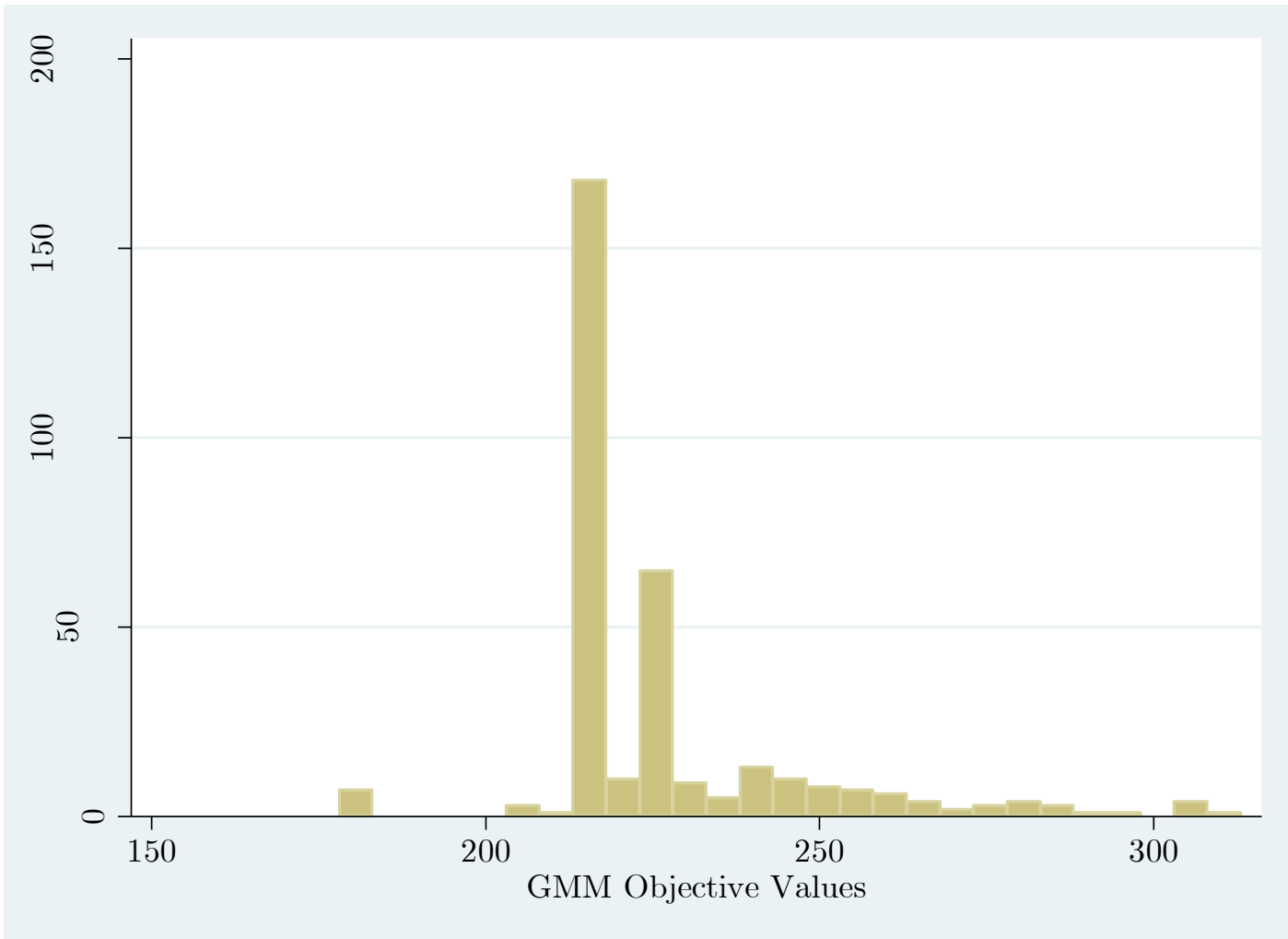


Figure 25: A histogram of GMM objective values for converged algorithms using the automobile data using a contraction mapping tolerance of 10^{-16} . This truncates the upper 10% of the converged GMM objective values. The parameters that lead to the lowest objective value uncovered (133.8) did not achieve convergence; therefore, this objective value is not reflected in the histogram.

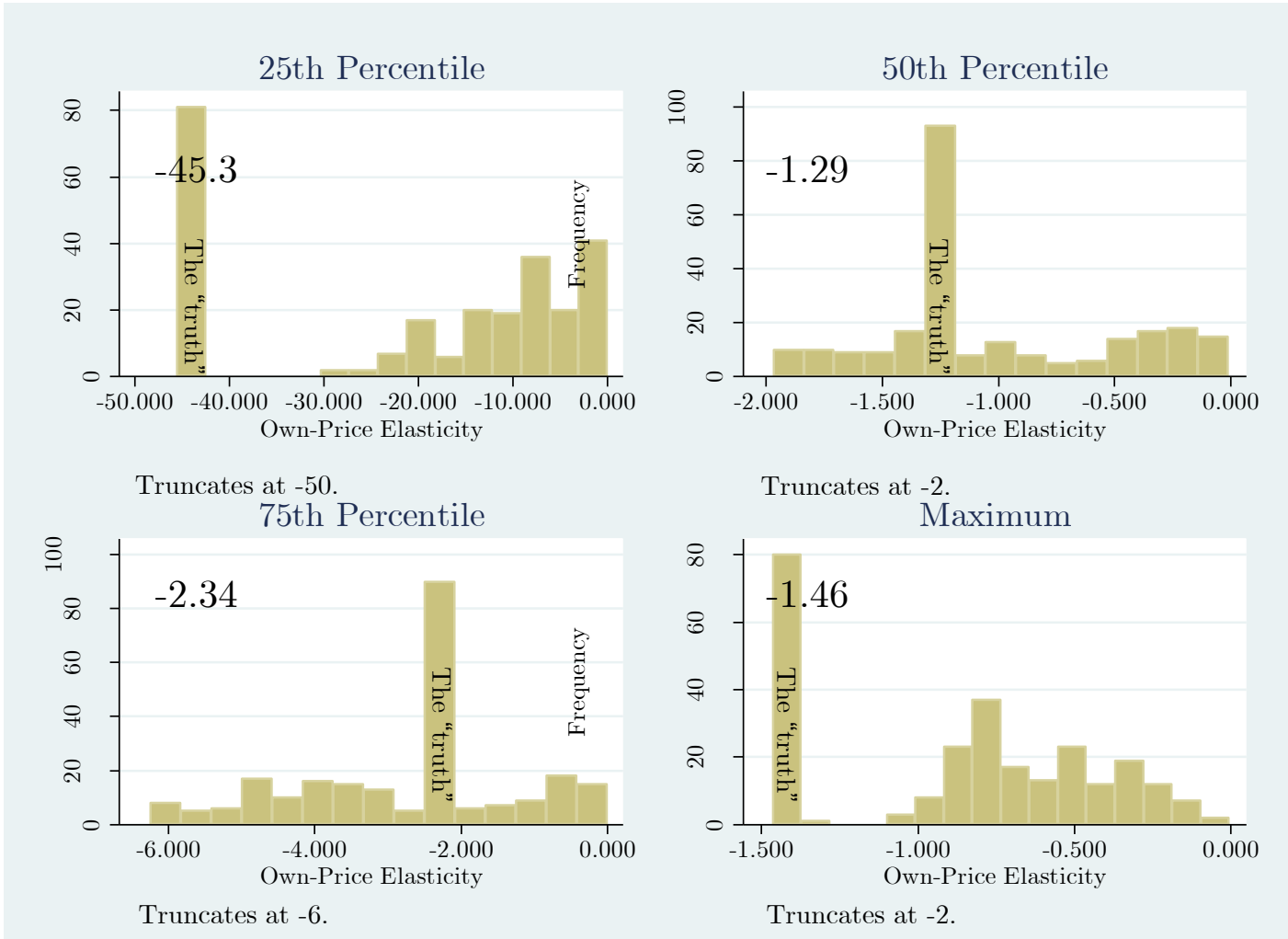


Figure 26: Histogram of candidate set of own-price elasticities for four products for the cereal data. The truth is defined as the set of parameter estimates that lead to the lowest GMM objective value.

These use roughly the same bin width and horizontal scale as in Figure 6.

The average within-product-market standard deviation in the estimated elasticities across all converged results is 7.78; the median is 3.37. For comparison using 10^{-08} as the tolerance, the mean and median are 7.58 and 3.24, respectively.

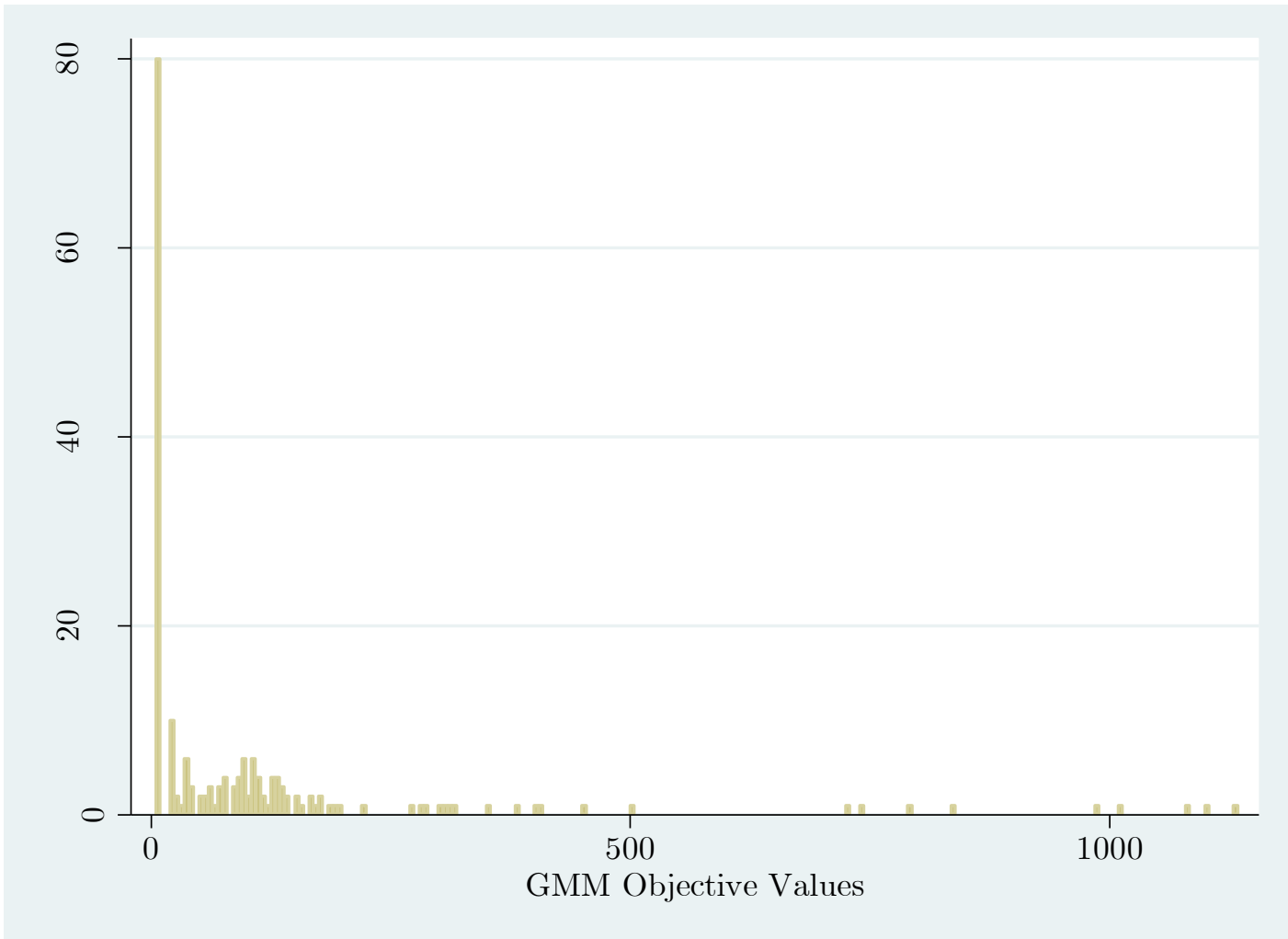


Figure 27: A histogram of GMM objective values for converged algorithms using the cereal data using a contraction mapping tolerance of 10^{-12} . This truncates the upper 25% of the converged GMM objective values.

	QNewton1	Simplex	Solvopt	Conigrad	QNewton2	S A	MADS	GPS	Matlab GA
Price	-0.458	-0.458	-0.409	-0.462	-0.554	-0.551	-0.342	-0.527	-0.215
Constant	-9.753	-9.731	-5.672	-9.685	-14.969	-7.472	-6.804	-16.791	-8.129
HP/Weight	3.063	3.018	-2.610	3.053	2.106	-15.812	-1.795	5.267	-0.197
Air Conditioning	1.265	1.266	0.694	1.286	-5.197	1.094	0.664	1.185	0.128
MPG	-0.610	-0.642	-4.539	-0.635	0.058	-1.088	-3.451	-0.077	-1.321
Size	2.762	2.749	1.738	2.761	3.124	3.820	1.968	3.175	2.592
Sigma_price	0.183	0.183	0.156	0.184	-0.293	-0.275	0.122	-0.225	0.191
Sigma_C	-2.718	-2.767	5.179	-2.710	6.522	-7.201	4.178	7.479	-2.769
Sigma_HP/Weight	1.015	1.164	-7.193	1.039	-2.886	-14.672	-5.882	-0.736	1.266
Sigma_AC	0.541	0.560	1.639	0.476	-7.823	1.832	1.411	-2.468	0.405
Sigma_MPG	0.686	0.703	2.913	0.699	0.048	1.300	2.342	-0.012	0.730
GMM Obj	215.1	215.1	178.1	215.1	207.5	125.5	187.2	196.7	215.6

Table 1: Parameter estimates and GMM objective values for the 9 “best” set of results. The JBES GA results are omitted since they were unreliable. 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.

	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			

	QNewton1	Simplex	Solvopt	Congrad	QNewton2	S A	MADS	GPS	Matlab GA
Price	-0.458 (0.189)	-0.458 (0.187)	-0.409 (0.064)	-0.462 (0.208)	-0.554 (0.119)	-0.551 (0.099)	-0.342 (0.055)	-0.527 (0.087)	-0.215 (0.056)
Constant	-9.753 (1.152)	-9.731 (1.124)	-5.672 (1.17)	-9.685 (1.235)	-14.969 (1.428)	-7.472 (0.96)	-6.804 (1.055)	-16.791 (0.955)	-8.129 (0.822)
HP/Weight	3.063 (0.841)	3.018 (0.827)	-2.610 (1.15)	3.053 (0.96)	2.106 (2.451)	-15.812 (3.388)	-1.795 (0.966)	5.267 (1.02)	-0.197 (0.715)
Air Conditioning	1.265 (1.122)	1.266 (1.214)	0.694 (0.355)	1.286 (0.975)	-5.197 (3.663)	1.094 (0.691)	0.664 (0.311)	1.185 (1.168)	0.128 (0.227)
MPG	-0.610 (0.761)	-0.642 (0.723)	-4.539 (0.983)	-0.635 (0.851)	0.058 (0.186)	-1.088 (0.371)	-3.451 (0.893)	-0.077 (0.166)	-1.321 (0.432)
Size	2.762 (0.287)	2.749 (0.272)	1.738 (0.374)	2.761 (0.308)	3.124 (0.355)	3.820 (0.398)	1.968 (0.317)	3.175 (0.285)	2.592 (0.204)
Sigma_price	0.183 (0.109)	0.183 (0.11)	0.156 (0.031)	0.184 (0.117)	-0.293 (0.09)	-0.275 (0.078)	0.122 (0.028)	-0.225 (0.049)	0.191 (0.045)
Sigma_C	-2.718 (0.816)	-2.767 (0.857)	5.179 (0.907)	-2.710 (0.826)	6.522 (1.647)	-7.201 (1.787)	4.178 (0.801)	7.479 (0.985)	-2.769 (0.454)
Sigma_HP/Weight	1.015 (1.213)	1.164 (1.186)	-7.193 (0.862)	1.039 (1.252)	-2.886 (2.805)	-14.672 (2.669)	-5.882 (0.751)	-0.736 (0.847)	1.266 (1.107)
Sigma_AC	0.541 (4.998)	0.560 (5.186)	1.639 (0.444)	0.476 (5.493)	-7.823 (3.193)	1.832 (2.196)	1.411 (0.405)	-2.468 (1.872)	0.405 (0.708)
Sigma_MPG	0.686 (0.446)	0.703 (0.421)	2.913 (0.538)	0.699 (0.498)	0.048 (0.403)	1.300 (0.279)	2.342 (0.488)	-0.012 (0.25)	0.730 (0.218)
GMM Obj	215.1	215.1	178.1	215.1	207.5	125.5	187.2	196.7	215.6

Table 1b: Parameter estimates and GMM objective values for the 9 “best” set of results with standard errors. The JBES GA results are omitted since they were unreliable. The standard errors are calculated ignoring multiple extrema.

	QNewton1	Simplex	Solvopt	Conjgrad	QNewton2	MADS	GPS	Matlab GA	Reported	SEs
Price	-30.228	-30.270	-62.731	-31.279	-62.718	-34.986	-32.821	-114.110	-32.433	7.743
constant_sigma	0.344	0.343	0.558	0.314	0.558	0.400	0.407	0.273	0.377	0.129
price_sigma	1.379	1.657	3.313	1.489	3.314	1.827	1.060	2.127	1.846	1.075
sugar_sigma	-0.008	-0.003	-0.006	-0.006	-0.006	-0.007	-0.096	0.010	0.004	0.012
mushy_sigma	0.032	0.051	0.093	0.057	0.093	0.103	0.014	0.325	0.081	0.205
C*inc	3.797	3.838	2.292	3.366	2.298	3.520	3.208	2.437	3.089	1.213
P*inc	-0.557	10.239	588.340	0.052	588.160	-6.448	2.369	0.348	16.598	172.334
S*inc	-0.202	-0.199	-0.385	-0.199	-0.385	-0.215	-0.213	-0.121	-0.193	0.005
M*Inc	1.846	1.430	0.748	1.832	0.750	1.775	2.223	1.225	1.468	0.697
P*inc2	-0.051	-0.601	-30.193	0.067	-30.184	0.498	-0.008	0.066	-0.659	8.955
C*Age	0.920	0.739	1.284	1.026	1.281	1.250	0.631	0.645	1.186	1.016
S*Age	0.030	0.031	0.052	0.032	0.052	0.036	0.004	0.018	0.029	0.036
M*Age	-1.783	-1.283	-1.353	-1.869	-1.350	-1.790	-1.865	-0.950	-1.514	1.103
P*Child	1.017	4.366	11.053	2.906	11.042	11.725	-3.395	0.377	11.625	5.207
GMM Obj	19.55	17.22	4.56	18.56	4.56	16.59	50.99	34.24		14.9

Table 2: Parameter estimates and GMM objective values for the 9 “best” set of results. The JBES GA results are omitted since they were unreliable. We include the results from Nevo (2000) for comparison; here the models are identical.

We present the MADS results that lead to the lowest GMM objective function. However, the algorithm did not converge at this point.

	QNewton1	Simplex	Solvopt	Conjgrad	QNewton2	S A	MAIDS	GPS	Matlab GA
Price	-30.228	-30.270	-62.731	-31.279	-62.718	-36.748	-31.796	-32.821	-114.110
constant_sigma	0.344 (0.156)	0.343 (0.132)	0.558 (0.163)	0.314 (0.157)	0.558 (0.162)	0.085 (0.405)	0.352 (0.134)	0.407 (0.229)	0.273 (4.548)
price_sigma	1.379 (1.511)	1.657 (1.177)	3.313 (1.34)	1.489 (1.374)	3.314 (1.339)	1.848 (4.708)	1.926 (1.126)	1.060 (1.352)	2.127 (56.881)
sugar_sigma	-0.008 (0.016)	-0.003 (0.013)	-0.006 (0.014)	-0.006 (0.016)	-0.006 (0.014)	-0.011 (0.036)	-0.004 (0.013)	-0.096 (0.028)	0.010 (0.662)
musby_sigma	0.032 (0.256)	0.051 (0.197)	0.093 (0.185)	0.057 (0.247)	0.093 (0.185)	-0.886 (0.639)	0.060 (0.21)	0.014 (0.447)	0.325 (3.689)
C*inc	3.797 (1.263)	3.838 (1.085)	2.292 (1.209)	3.366 (1.296)	2.298 (1.207)	0.916 (2.093)	3.152 (1.194)	3.208 (2.186)	2.437 (18.664)
P*inc	-0.557 (149.78)	10.239 (196.15)	588.340 (270.45)	0.052 (135.61)	588.160 (270.36)	0.507 (337.82)	5.822 (168.55)	2.369 (94.469)	0.348 (2205.2)
S*inc	-0.202 (0.047)	-0.199 (0.047)	-0.385 (0.122)	-0.199 (0.044)	-0.385 (0.122)	-0.023 (0.062)	-0.187 (0.046)	-0.213 (0.052)	-0.121 (1.215)
M*Inc	1.846 (1.051)	1.451 (0.754)	0.749 (0.802)	1.740 (0.974)	0.748 (0.801)	0.466 (0.659)	1.692 (0.747)	2.223 (1.016)	1.224 (18.218)
P*inc2	-0.051 (7.794)	-0.601 (10.194)	-30.193 (14.102)	0.067 (7.05)	-30.184 (14.097)	0.507 (17.077)	-0.137 (8.766)	-0.008 (4.7)	0.066 (112.14)
C*Age	0.920 (1.527)	0.739 (0.867)	1.284 (0.631)	1.026 (1.484)	1.281 (0.63)	-0.315 (0.761)	1.023 (1.093)	0.631 (1.859)	0.645 (26.252)
S*Age	0.030 (0.051)	0.031 (0.038)	0.052 (0.026)	0.032 (0.048)	0.052 (0.026)	0.131 (0.149)	0.025 (0.041)	0.004 (0.067)	0.018 (5.887)
M*Age	-1.783 (1.787)	-1.283 (1.058)	-1.353 (0.667)	-1.869 (1.708)	-1.350 (0.665)	0.866 (1.388)	-1.446 (1.216)	-1.865 (1.953)	-0.950 (13.7)
P*Child	1.017 (4.963)	4.366 (4.314)	11.053 (4.122)	2.906 (5.044)	11.042 (4.119)	0.056 (6.951)	8.116 (5.042)	-3.395 (8.713)	0.377 (323.62)
GMM Obj	19.55	17.22	4.56	18.56	4.56	131.32	15.79	50.99	34.24

Table 2b: Parameter estimates and GMM objective values for the 9 “best” set of results with standard errors. The JBES GA results are omitted since they were unreliable. The standard errors are calculated ignoring multiple extrema.

Algorithm	25th			Median			75th			Largest		
	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max
Quasi-Newton 1	-2.86	-2.70	-1.19	-2.61	-2.57	-0.78	-3.70	-3.55	-2.37	-3.85	-3.41	-1.66
Simplex	-2.88	-2.72	-0.91	-2.67	-2.59	-0.70	-3.77	-3.57	-1.69	-4.04	-3.42	-1.34
SolvOpt	-5.07	-2.17	-1.71	-4.20	-1.07	-1.07	-5.19	-3.87	-2.76	-3.41	-2.73	1.84
Conj. Grad.	-3.35	-2.73	-1.09	-2.86	-2.59	-1.09	-4.09	-3.54	-1.92	-3.77	-3.41	-0.49
Quasi-Newton 2	-10.64	-3.72	-1.35	-3.20	-3.20	-0.85	-10.35	-4.46	-2.73	-3.52	1.08	1.08
JBES GA	--	--	--	--	--	--	--	--	--	--	--	--
Simulated Annealing	-5.81	-3.00	-1.04	-7.11	-3.63	-1.14	-12.00	-3.33	-0.93	-11.90	-3.40	-0.91
MADS	-3.65	-1.57	-1.57	-4.25	-1.02	0.06	-7.61	-3.33	-2.71	-3.65	-2.42	-1.08
GPS	-40.47	-3.41	-0.79	-49.43	-3.76	-0.72	-91.42	-5.00	-1.79	-87.04	-1.78	-1.78
Matlab GA	-23.55	-1.58	-1.43	-29.85	-1.01	-1.01	-53.72	-0.53	-0.53	-53.16	-0.41	-0.41
Total	-40.47	-3.00	-0.79	-49.43	-3.63	0.06	-91.42	-3.33	-0.53	-87.04	-3.40	1.84

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 25th representing the product with the market share equal to the 25th percentile, etc.

These results truncate the worst 10% 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 16.83; the median is 11.12. The average own-price elasticity is -6.63. Among the “best” sets of parameter values, the mean standard deviation is 8.42, the median is 6.05, and the mean own-price elasticity is -5.96.

Algorithm	25th			Median			75th			Largest		
	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max
Quasi-Newton 1	-26.45	-14.04	-0.38	-2.02	-1.75	-0.17	-6.73	-3.73	-0.19	-1.02	-0.76	-0.23
Simplex	-21.60	-18.58	-0.35	-2.41	-1.92	-0.09	-5.79	-5.71	-0.16	-1.14	-0.75	-0.11
SolveOpt	-45.34	-45.30	-45.30	-1.29	-1.29	-1.29	-2.34	-2.34	-2.34	-1.46	-1.46	-1.46
Conj. Grad.	-22.53	-20.68	-6.91	-2.08	-1.91	-0.73	-6.60	-5.98	-2.30	-0.91	-0.82	-0.63
Quasi-Newton 2	-47.15	-45.35	8.65	-1.93	-1.29	-1.22	-3.10	-2.34	0.24	-1.47	-1.46	-1.03
JBES GA	—	--	--	—	--	--	--	--	--	--	--	--
Simulated Annealing	—	--	--	—	--	--	--	--	--	--	--	--
MADS	—	--	--	—	--	--	--	--	--	--	--	--
GPS	-43.24	-23.66	4.33	-2.10	-1.55	-0.40	-11.22	-6.72	-0.17	-0.98	-0.76	-0.31
Matlab GA	-19.94	-12.55	-2.33	-2.14	-1.76	-0.60	-5.98	-4.61	-2.05	-1.07	-0.77	-0.48
Total	-47.15	-45.30	8.65	-2.41	-1.29	-0.09	-11.22	-2.34	0.24	-1.47	-1.46	-0.11

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. The JBES GA and Simulated Annealing algorithms never converge, so their results are omitted. MADS converges, but only at points where the GMM objective function value is above the 75th percentile.

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

Algorithm	25th (x10)			Median (x10)			75th (x10)			Largest (x10)		
	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max
Quasi-Newton 1	0.37	1.75	1.76	0.23	0.74	0.74	0.14	0.40	0.93	0.27	0.79	0.84
Simplex	0.06	1.79	1.87	0.06	0.77	0.77	0.12	0.40	1.55	0.16	0.80	1.01
SolvOpt	0.68	1.74	4.78	0.55	0.55	1.61	0.16	0.98	2.16	-0.27	0.42	0.79
Conj. Grad.	0.14	1.78	2.65	0.19	0.76	1.08	0.11	0.36	1.28	0.10	0.79	0.88
Quasi-Newton 2	0.65	3.12	4.29	0.39	1.07	3.20	0.16	1.86	1.86	-0.15	-0.15	2.03
JBES GA	—	—	—	—	—	—	—	—	—	—	—	—
Simulated Annealing	0.18	1.51	1.62	0.22	1.28	1.69	-0.50	0.01	6.52	0.11	0.24	2.19
MADS	0.49	1.00	3.17	0.26	0.49	2.04	-0.08	0.75	1.94	0.40	0.45	1.01
GPS	0.10	2.48	27.42	0.10	1.51	17.07	0.10	1.09	46.14	0.21	0.91	39.27
Matlab GA	0.77	1.19	14.61	0.29	0.36	9.51	-0.40	-0.40	11.94	0.07	0.07	14.12
Total	0.06	1.51	27.42	0.06	1.28	17.07	-0.50	0.01	46.14	-0.27	0.24	39.27

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 25th representing the product with the market share equal to the 25th percentile, etc.

These results truncate the worst 10% 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 0.23; the median is 0.04. The average cross-price elasticity is 0.01. Among the “best” sets of parameter values, the mean standard deviation is 0.02, the median is 0.003, and the mean cross-price elasticity is 0.01.

Algorithm	25th (x10)			Median (x10)			75th (x10)			Largest (x10)		
	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max	Min	'Best'	Max
Quasi-Newton 1	0.01	4.77	22.43	0.23	2.63	4.17	0.01	0.74	5.11	0.09	0.15	0.47
Simplex	0.00	10.22	34.90	0.04	3.42	5.71	0.00	2.53	2.79	0.01	0.13	2.03
SolvOpt	80.95	80.98	81.12	2.51	2.52	2.52	0.15	0.15	0.15	1.81	1.82	1.82
Conj. Grad.	2.56	14.47	16.85	1.17	3.70	4.31	0.31	3.22	4.80	0.11	0.12	0.55
Quasi-Newton 2	-35.22	81.13	87.19	2.27	2.51	9.41	0.00	0.15	0.38	-0.05	1.82	1.94
JBES GA	--	--	--	--	--	--	--	--	--	--	--	--
Simulated Annealing	--	--	--	--	--	--	--	--	--	--	--	--
MADS	--	--	--	--	--	--	--	--	--	--	--	--
GPS	-3.27	20.68	40.68	0.59	2.54	4.94	0.00	3.77	10.61	-0.04	0.16	0.38
Matlab GA	0.13	5.22	17.20	0.91	3.40	4.28	0.19	1.34	2.77	0.09	0.17	0.55
Total	-35.22	80.98	87.19	0.04	2.52	9.41	0.00	0.15	10.61	-0.05	1.82	2.03

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 25th representing the product with the market share equal to the 25th percentile, etc.

The JBES GA and Simulated Annealing algorithms never converge, so their results are omitted. MADS converges, but only at points where the GMM objective function value is above the 75th percentile.