DICE

DISCUSSION PAPER

NO 399

Flexible Estimation of Random Coefficient Logit Models of Differentiated Product Demand

Johannes Kandelhardt

May 2023

hhu
Heinrich Heine
Universität Düsseldorf

IMPRINT

DICE DISCUSSION PAPER

Published by:
Heinrich-Heine-University Düsseldorf,
Düsseldorf Institute for Competition Economics (DICE),
Universitätsstraße 1, 40225 Düsseldorf, Germany
www.dice.hhu.de

Editor:
Prof. Dr. Hans-Theo Normann
Düsseldorf Institute for Competition Economics (DICE)
Tel +49 (0) 211-81-15125, E-Mail normann@dice.hhu.de

ISSN 2190-9938 (online) / ISBN 978-3-86304-398-8

Düsseldorf Institute for Competition Economics
(DICE)

Heinrich-Heine-Universität Düsseldorf

Flexible Estimation of Random Coefficient Logit Models of Differentiated Product Demand

Author: Johannes Kandelhardt [1]

May 2023

[1]I thank André Romahn for invaluable advice. I thank Hendrik Döpper, Florian Heiß, Simon Martin, Sebastian Schroers, Joel Stiebale, Frank Verboven and participants of the DICE Discussion Group and Winter School Fügen for helpful comments. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 235577387/GRK1974. Computational infrastructure and support were provided by the Centre for Information and Media Technology (ZIM) at Heinrich Heine University Düsseldorf.

Abstract

The Berry, Levinsohn, and Pakes (1995, BLP) model is widely used to obtain parameter estimates of market forces in differentiated product markets. The results are often used as an input to evaluate economic activity in a structural model of demand and supply. Precise estimation of parameter estimates is therefore crucial to obtain realistic economic predictions. The present paper combines the BLP model and the logit mixed logit model of Train (2016) to estimate the distribution of consumer heterogeneity in a flexible and parsimonious way. A Monte Carlo study yields asymptotically normally distributed and consistent estimates of the structural parameters. With access to micro data, the approach allows for the estimation of highly flexible parametric distributions. The estimator further allows to introduce correlations between tastes, yielding more realistic demand patterns without substantially altering the procedure of estimation, making it relevant for practitioners. The BLP estimator is established to yield biased and inconsistent results when the underlying distributional shape is non-normally distributed. An application shows the estimator to perform well on a real world dataset and provides similar estimates as the BLP estimator with the option of specifying consumer heterogeneity as a function of a polynomial, step function or spline, resulting in a flexible estimation procedure.

1 Introduction

The Berry, Levinsohn, and Pakes (1995, BLP hereafter) model is a powerful tool of empirical industrial organization. Its success is based on several features. The model yields realistic predictions of substitution patterns with minimal data requirements. It also deals with the endogeneity of price that is borne out of the correlation with unobserved demand shocks. The normally distributed random coefficients, based on consumer-product interaction terms, are key to obtaining realistic substitution patterns. Commonly observed markets with differentiated products can be tractably handled by applying the contraction mapping, removing the need to estimate product fixed effects and thus largely reducing the number of parameters to be estimated and improving efficiency. The model runs on aggregate data but is flexible enough to also incorporate micro data and demographics. The BLP estimator is therefore considered the workhorse of contemporary demand estimation.

The BLP estimator is frequently applied to analyze important research questions. Recent examples include Björnerstedt and Verboven (2016) who estimate a BLP demand system to evaluate a merger in the Swedish painkiller market. Duch-Brown et al. (2017) use a BLP model to investigate the impact of online sales on consumers and firms. Fan and Yang (2020) employ a BLP demand system in the U.S. smartphone market to study theoretical ambiguities concerning firms product portfolios when firms merge and product choices are endogenous. Bourreau, Sun, and Verboven (2021) estimate a BLP demand system to infer potentially collusive conduct in the French telecommunications industry.

The normality assumption of the random coefficients is a convenient but restrictive assumption and is usually justified by data availability restrictions and ease of implementation. Whereas mean and variance are often reliably estimated in spite of misspecification, this does not hold for higher moments of the distribution, as discussed by Hess and Axhausen (2005). Hess, Bierlaire, and Polak (2005) discuss problems related to the estimation of random taste distributions. They find distributions mistakenly specified as symmetric to bias point estimates. Results about potential misspecification bias is contributed by Compiani and Smith (2022), who finds that misspecification affects equilibrium profits. Despite this finding, it is not generally clear how the arbitrary distributional assumption of normality affects subsequent analyses. This is somewhat astonishing as the parameter estimates of the BLP demand system are often just an input to a more elaborate structural model. Getting estimates of the demand model right is crucial to derive such important structural measures as elasticities and markups. To provide correct policy advice, it is thus imperative to carefully model the underlying distributions.

The present paper aims at relaxing the normality assumption by using a method of sieves approach as implemented by Train (2016), which is referred to as the logit mixed logit (LML) model. By combining the BLP model and the logit mixed logit model, the random coefficients are parsimoniously estimated in a flexible way. In short, the distributions of the random coefficients are dependent on a parameter vector and random tastes are drawn from a prespecified grid and inserted into the nonlinear interaction term. Nonrandom coefficients can still be concentrated out as in the BLP model. The original estimation procedure is largely unchanged, and practitioners are able to apply all common tools that are used in conjunction with BLP. The model can be further enhanced by introducing correlations between random coefficients or adding micro data to estimate demand parameters more precisely. This in turn allows substitution patterns to be closely mapped to the data without the bias of misspecification.

The paper further touches on several issues that might be of interest to practitioners. In line with the results of Reynaert and Verboven (2014), optimal or approximately optimal instruments are shown to yield more precise estimation results, benefiting post estimation by increasing accuracy of estimated markups or marginal costs. The addition of micro data is shown to make the estimation of sophisticated taste distributions feasible, without the need to increase the sample size of the aggregate data. This is of importance, since changes in the underlying true taste distributions are set out to imply different market equilibria, which will result in biased BLP estimates if not appropriately taken into account. An estimation using solely micro data in the moment conditions confirms the findings in Berry and Haile (2022), that micro data identifies substitution patterns, and instruments are needed to pin down the price coefficient. Lastly, the flexible estimator is applied to a real world dataset. The estimator is subjected to different specifications of the grid and the estimated coefficients are found to be quite robust to the grid specification. This result is reassuring, since the estimator's validity largely hinges on the robustness of the subjectively chosen grid support.

The rest of the paper is structured as follows. Section 2 relates the paper to the literature. In Section 3, the BLP model, the logit mixed logit model and the flexible model are introduced along appropriate notation. Section 4 presents a Monte Carlo study and shows consistency of the estimator. Section 5 enriches the model with micro data and shows how micro data can be used to estimate sophisticated distributions. Section 6 investigates the effects of distributional misspecification on the performance of the BLP estimator. Section 7 applies the flexible estimation procedure to a commonly used dataset. Section 8 concludes.

2 Literature

The literature concerned with demand estimation and estimation of heterogeneous preferences is vast. I focus on the literature of empirical industrial organization that evolved from the discrete choice literature, with BLP as its most prominent member. Narrowing it down to flexible demand estimation with respect to consumer heterogeneity, several strands of the literature can be distinguished. The approaches crucially depend on how much structure is put on the taste distributions underlying the demand functions. I do not provide a complete literature review and other options for organizing the literature are feasible.

A first strand of the literature uses the full set of parameterizations, with random coefficients modeled as a parametric distribution and modeling the error term to be type 1 extreme value distributed, yielding the logit model. BLP belongs to this category, using a normal distribution. Researchers also use discrete type distributions when consumer types are observable, e.g., Berry, Carnall, and Spiller (1996), Berry and Jia (2010) and Doi (2022).

A second strand of the literature relaxes the parametric assumption of the random coefficients, while retaining the logit error term. The approaches rely on micro data and do not easily generalize to a setting with aggregate data only and price endogeneity; the high flexibility allows a good fit to the data and the linearization provides computational ease and fast estimation. Train (2008) retains the logit formula and uses the expectation maximization algorithm. Fosgerau and Mabit (2013) propose flexible specifications of the mixing distributions using power series approximations. Fosgerau and Hess (2008) use Legendre polynomials and mixtures. Fox, Kim, and Yang (2016) linearize the nonlinear logit model and estimate the probability masses directly via regression. Heiss, Hetzenecker, and Osterhaus (2021) discuss potential problems related to this approach such as the shrinkage of a significant amount of probability masses to zero due to high correlation in the covariates. They provide an enhancement of the approach by putting more structure or smoothing on the derived probability masses. Train (2016) uses a semi-nonparametric approach to approximate the mixing distribution in a way that is rather flexible and easy to implement. Train uses micro level data and estimates the model by hierarchical Bayes and maximum likelihood estimation. Other models are more closely related to the BLP model, usually with a semi-nonparametric specification of the random coefficients. The models aim for a direct transformation of the BLP model to relax the distributional assumptions without wavering too far from the BLP framework. Wang (2022) and Lu, Shi, and Tao (2022) employ a semi-nonparametric approach and linear regression to model the random coefficients. The applied sieves in these models can be relatively freely chosen, e.g., according to Gallant and Nychka (1987) or as in Train (2016). These models are convenient as they retain the advantages of BLP but relax the normality assumption.

Lastly, a third strand of the literature transforms the possibly nonlinear underlying functions and drops the logit error term to estimate demand as flexibly as possible. The models can handle complements and are thus able to flexibly structure the relationship between product groups when the relationship is not clear ex ante. Compiani (2022) and Monardo (2021) use linear transformations and Bernstein polynomials to approximate the market shares nonparametrically. To identify parameters, the authors use economically motivated restrictions and distance measures based on the characteristics space. The flexibility comes at the price of high computational cost and data requirements, restricting the use in practice.

This concludes the review of the literature. In the next Section I move on to introduce the notation of the BLP and logit mixed logit model before discussing the implementation of the flexible estimation procedure.

3 Relaxing the Normality Assumption

3.1 The BLP Estimator

The BLP model builds on the discrete choice literature and assumes each consumer to derive utility from consumption based on product attributes and consumer specific tastes. Let $j = 0, ..., J$ index J products, where $j = 0$ denotes the outside option. Let $k = 1, ..., K$ index K characteristics of a product. Characteristic k of product j is denoted by x_{jk} and part of the $1 \times K$ dimensional row vector x_j. Let consumer tastes be denoted by the $1 \times K$ dimensional row vector ν_r for $r = 1, ..., R$ consumers and let ξ_j be a characteristic unobserved by the econometrician. The functional form of demand further depends on the parameter vector θ, which relates tastes and attributes and determines the marginal utility the consumer obtains from consuming an additional unit of attribute x_{jk}. Following BLP, consumer utility is given by

$$u_{rj} = x_j\beta - \alpha p_j + \xi_j - p_j\theta_p\nu_{rp} + \sum_k x_{jk}\theta_k\nu_{rk} + \epsilon_{rj}. \tag{1}$$

Utility can be split into a mean utility component $\delta_j = x_j\beta - \alpha p_j + \xi_j$ and an individual-specific component $\mu_{rj} = -p_j\theta_p\nu_{rp} + \sum_{k=1}^{K} x_{jk}\theta_k\nu_{rk}$. The individual taste deviations from the mean are represented by ν. Finally, ϵ_{rj} is an idiosyncratic preference shock assumed to be type I extreme value distributed, resulting in the familiar logit functional form for the choice probabilities, originally developed by Luce (1959):

$$\text{Prob}_{rj}(\delta;\theta,\nu_r) = \frac{\exp(\delta_j + \mu_{rj})}{1 + \sum_{k=1}^{J} \exp(\delta_k + \mu_{rk})}, \tag{2}$$

The unconditional market shares can be found by integrating over consumer heterogeneity,

$$\text{Prob}_j(\delta;\theta,\nu_r) = \int \text{Prob}_{rj}(\delta;\theta,\nu_r) d\nu_r. \tag{3}$$

To estimate the model parameters, BLP apply a nested fixed point contraction mapping, equating the simulated market shares with the observed market shares. Given an appropriate identification assumption, the parameter vector θ is estimated as

$$\hat{\theta} = \underset{\theta}{\text{argmin}}\, \hat{\xi}'(\hat{\delta};\theta,\nu) Z W Z' \hat{\xi}(\hat{\delta};\theta,\nu), \tag{4}$$

with $\hat{\xi}$ as a vector stacked over markets, Z a matrix of instruments stacked over markets and W being an appropriate weighting matrix, often the conditionally homoskedastic weight matrix. For a more elaborate discussion, please refer to Berry, Levinsohn, and Pakes (1995).

3.2 The Logit Mixed Logit Model

Train (2016) uses the logit model to calculate a weight for each conditional logit probability. The random coefficients logit model is then aggregated by weighting each consumer-specific choice probability by its respective population weight, formally

$$s_j = \sum_{r \in S} \text{Prob}_{rj}(\nu_r) \cdot W(\tilde{\nu}_r|\theta) = \sum_{r \in S} \left(\frac{\exp(x_j \nu_r)}{\sum_{k \in J} \exp(x_k \nu_r)} \right) \cdot \left(\frac{\exp(z(\tilde{\nu}_r)\theta)}{\sum_{s \in S} \exp(z(\tilde{\nu}_s)\theta)} \right). \tag{5}$$

The θ parameters shape the distribution of the random coefficients in combination with a vector valued function $z(\cdot)$ in the exponentials to form the weight $W(\tilde{\nu}_r|\theta)$. This procedure covers a wide range of possible shapes. S refers to the space of all possible random coefficients and will in practice be substituted by random draws of the random coefficients. The length of the grid needs to be specified in advance, which is a drawback of the approach. Train suggests to use an estimate of a parametric specification first (the logit model or random coefficients model with a normal distribution) and then use the estimated mean and variance to determine an adequate grid specification.

Train proposes a Legendre transformation on the random coefficients in the weight function, yielding transformed coefficients $\tilde{\nu}_r$ with a support of $\tilde{\nu}_r \in [-1, 1]$ and additional advantages such as orthogonality and reduced risk of overflow in the exponentials. The transformation is done by calculating $\tilde{\nu}_r = -1 + 2(\nu_r - a)/(b - a)$ with a and b being the lower and upper boundary of respective heterogeneity parameters of the grid. The Legendre transformation is not mandatory but recommended to avoid potential numerical issues during estimation.

The vector valued function $z(\cdot)$ should take as many terms as necessary to approximate the underlying distribution. As discussed by Train (2016), the function should be modeled in such a way that the true distribution can be appropriately cast. Several specifications are possible, e.g., polynomials, step function or splines. Note that even though only a limited number of parameters has to be estimated, the resulting distribution of consumer preferences is very flexible.

3.3 Combining BLP and the Logit Mixed Logit Model

It is now possible to combine the sophistication of the Berry, Levinsohn, and Pakes (1995) model with the flexibility of the logit mixed logit model of Train (2016), by plugging the utility specification given in Equation (1) into Train's model:

$$s_j = \sum_{r \in S} \text{Prob}_{rj}(\delta_j; \theta, \nu_r) \cdot W(\tilde{\nu}_r|\theta) = \sum_{r \in S} \left(\frac{\exp(\delta_j + \mu_{rj})}{1 + \sum\limits_{k \in J} \exp(\delta_k + \mu_{rk})} \right) \cdot \left(\frac{\exp(z(\tilde{\nu}_r)\theta)}{\sum\limits_{s \in S} \exp(z(\tilde{\nu}_s)\theta)} \right) \tag{6}$$

with Prob_{rj} being the conditional (unweighted) logit market share of consumer type r and $W(\tilde{\nu}_r|\theta)$ (abbreviated W_r) the weight or the probability mass associated with this type of consumer. I label this estimator the flexible estimator. The functions $z(\cdot)$ transform the taste parameters according to the specification of the researcher. It may contain $\tilde{\nu}_{rk}$ and its squared value to estimate a more flexible version of the normal distribution with cut off tails. A fourth order polynomial may represent a bimodal distribution. See Train (2016) for an in-depth discussion.

To distinguish between consumer tastes that are homogeneous and heterogeneous, I define x_j to be the vector of attributes of product j holding all the attributes assumed to be associated with homogeneous tastes and denote the attributes associated with heterogeneous tastes as $\tilde{x}_j$. The vector x_j from now on holds only the linearly entering attributes of the products, whereas $\tilde{x}_j$ holds the nonlinearly entering product attributes. Also the price variable is from now on assumed to be contained in either of these sets.

Utility is similar to the specification in BLP with the modification that consumer heterogeneity draws enter the nonlinear interaction term directly, without a split into mean and deviation as in BLP, according to

$$\delta_j = x_j\beta + \xi_j, \tag{7}$$

$$\mu_{rj} = \sum_{k\in K} \tilde{x}_{jk}\nu_{rk}, \tag{8}$$

$$u_{rj} = x_j\beta + \xi_j + \sum_{k\in K} \tilde{x}_{jk}\nu_{rk} + \epsilon_{rj}. \tag{9}$$

The idea is formalized in Equations (7), (8) and (9). The mean utility component of product j is given by Equation (7) which contains linearly entering attributes x_j. The associated parameters to be estimated β are "concentrated out" similar to BLP as is discussed below. Equation (8) shows the nonlinearly entering interaction term. The standard deviations σ that are present in BLP drop out because the variance of the distribution is determined by θ and by the grid specification.

More details on notation: to keep consistency with BLP, the heterogeneity draws are denoted by ν_{rk}, which has β_{rk} as the equivalent in Train (2016). Therefore, the individual taste of individual r towards characteristic k is given by ν_{rk}. These tastes are drawn prior to estimation from the specified grid and can be scaled to be $\tilde{\nu}_{rk} \in [-1, 1]$ prior to estimation, as discussed in Section 3.2. The draws are held fixed during estimation. Given the grid support, the draws and the data, the nonlinear term μ is fixed and δ can be backed out using a contraction mapping similar to BLP, setting the observed market shares equal to the model implied market shares, $S_j = s_j(\delta, \nu; \theta)$, conditional on θ. See Appendix A.1 on how to rewrite the contraction to speed up estimation times, as shown by Brunner et al. (2017).

Defining $w \equiv e^{x_j\beta+\xi_j} = e^{\delta_j}$, the demand shock is then backed out by applying ordinary least squares or two stage least squares depending on the assumption concerning the price coefficient which also yields the linear demand parameters:

$$\log(w) = x_j\beta + \xi_j \tag{10}$$

$$\xi_j = \log(w) - x_j\beta \tag{11}$$

The contraction mapping yields ξ_j directly, if all consumer tastes concerning all attributes and the constant are assumed to be heterogeneous.

Now that ξ is known, a standard identification argument can be used and the estimation proceeds by the generalized method of moments as described by Hansen (1982).

This specific identification argument is not crucial to the flexible estimator, e.g., covariance restrictions, as in MacKay and Miller (2023), can be used if appropriate. Here, the instruments, z, are assumed to be mean independent of the demand unobservables, formally $E[\xi|z] = 0$ and the parameters identified by setting the objective function as close as possible to zero, with the objective given by

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, \hat{\xi}'(\hat{\delta};\theta,\nu) Z W Z' \hat{\xi}(\hat{\delta};\theta,\nu) \tag{12}$$

and W being an appropriate weighting matrix. For later use, let the objective be abbreviated by $G_1 = \xi' Z W Z' \xi$.

Note that the implementation of the flexible distribution does not change the general procedure of the BLP estimator and can thus be quickly added to the model by practitioners, if necessary, as an additional robustness check and to achieve richer substitution patterns. Adding the flexibility is independent of identification assumptions and does not largely change post estimation, apart from calculating simple derivatives. It is straightforward to add correlations between tastes to further enhance substitution patterns, as discussed by Train.

Now that the flexible estimation procedure is spelled out, its properties are discussed by conducting a Monte Carlo study in the next Section. The derivation of the market share derivatives with respect to price and demand shocks and the gradient are laid out in Appendix A.2 and Appendix A.3, respectively. A derivation for an exemplary specification with a polynomial as z variable is laid out in Appendix A.4.

4 Monte Carlo Study

In order to investigate the properties of the flexible estimator, I conduct a Monte Carlo study. The implementation follows Berry (1994) and Reynaert and Verboven (2014). The procedure assumes consumer heterogeneity to be structured in a certain way by selecting θ to shape the distribution of the random coefficients. Given randomly generated product data, the demand system parameters are estimated. I discuss empirical statistics of the estimator's asymptotic behavior, its distribution and its performance. The next Section lays out the Monte Carlo procedure. Section 4.2 discusses the use of instruments. Section 4.3 presents the results.

4.1 Implementation Details

The setup of the Monte Carlo study can be described as follows. Let product data be generated for $T = \{50,\ 200,\ 500\}$ markets. In each market, four firms produce three goods, so that there are twelve products in total. The number of observations available to the econometrician is therefore $n = \{600, 2{,}400, 6{,}000\}$. Firms compete in prices in a differentiated product market. There is one characteristic for each product in each market that is assumed to be independent over markets and drawn from $x \sim \text{Uniform}(0, 3)$. Consumers are assumed to display heterogeneous preferences concerning the characteristic; I thus denote it with a tilde, presently $\tilde{x}_{jt}$. The additional subscript t denotes market t. Demand further depends on price and on an intercept providing the link to the outside good, captured by $x_{jt} = (-p_{jt}, 1)$. Let $X_{jt} = (x_{jt}, \tilde{x}_{jt})$. The demand side marginal utility parameters of X_{jt} are given by the column vector $\beta_r = (\alpha, \beta_0, \nu_r) = (5, -2, \nu_r)$ with ν_r being the marginal utility parameter for individual r and the price coefficient is assumed to be $\alpha = 5$ and the intercept value -2. No consumer heterogeneity is assumed concerning the price coefficient. The Monte Carlo study is repeated for consumers being heterogeneous concerning the price coefficient, yielding similar results. See Appendix A.5 for details. For an example of a heterogeneous price coefficient applied to real data, see the application in Section 7. Demand follows the functional form given by Equation (6). The product unobservable characteristic ξ_{jt} is drawn from a multivariate normal distribution with a mean of 0 and a standard deviation of 1. Corresponding to each unobserved characteristic a supply shock ω_{jt} is drawn (with covariance of 0.7) inducing strong endogeneity due to the correlation between the unobserved characteristic and price.

Note that ξ_{jt} and $\tilde{x}_{jt}$ are independent. This guarantees that instruments generated from $\tilde{x}_{jt}$ yield adequate properties of the moment conditions with $E[\xi(\delta; \theta_0, \nu)|z] = 0$ and convergence of the objective to zero at the true parameter values (cf. Cameron and Trivedi, 2005; Hansen, 2021). The subscript of zero indicates evaluation of the objective at the true parameter values.

Marginal costs are assumed to be constant and determined by $\tilde{x}_{jt}$ with cost side parameters $\gamma = (2.5, 0.2)$. Marginal costs are further determined by the aforementioned supply shock ω, which is weakened by a factor of $\gamma_c = 0.2$. For clarity, utility is formally given by $u_{rjt} = X_{jt}\beta_r + \xi_{jt} + \epsilon_{rjt} = -5p_{jt} - 2 + \tilde{x}_{jt}\nu_r + \xi_{jt} + \epsilon_{rjt}$. Marginal costs are formally given by $mc_{jt} = 2.5 + 0.2\tilde{x}_{jt} + 0.2\omega_{jt}$.

As far as consumer heterogeneity is concerned, the logit formula is specified with a second order polynomial as z variable, which is similar to a normal distribution. The coefficients are $\theta = (\theta_{21}, \theta_{22}) = (0, -4)$. For a more sophisticated, bimodal shape of the underlying density, see Section 5. The random coefficient lies in an interval of $\nu_r \in [0, 8]$. Integration of the market shares during data generation is done deterministically with 1,000 equally spaced random coefficient values. Random draws are not needed since the integration is one dimensional and a densely packed real line is sufficient to accurately cast the distributional shape. This is convenient as random fluctuations are prevented from entering the objective. During estimation, 200 equally spaced values are used to integrate the market shares.

The general procedure can be described as follows. The data is generated. Based on the parameters, the conduct assumption and the distributional assumption concerning heterogeneity, the market equilibrium is established by iterating over the market participants' reactions until a measure of the price differences is below a threshold. The iterative procedure is necessary to establish the equilibrium, but introduces inaccuracy when generating the data, because no analytical solution to the nonlinear system of equations exists. Accuracy is obtained by setting a low tolerance. The threshold is set to less than 1e−14. As a distance measure, the Euclidean norm is used, formally $\|p^{iter+1} - p^{iter}\| < \epsilon = 1\text{e}{-14}$. The iteration starts with an arbitrary price vector and is based on the first order conditions of the differentiated Nash Bertrand model (cf. Berry, Levinsohn, and Pakes (1995)). This yields product prices and market shares. Then either BLP instruments or an approximation to optimal instruments are calculated and estimation can be performed by minimizing the objective for 10 different starting values, selecting the parameters with the lowest uncovered objective value out of these estimates and calculating summary statistics based on the objective minimizing parameter values. This entire process is repeated 1,000 times for each combination of instruments and market size, yielding $1{,}000 \cdot 2 \cdot 4 \cdot 10 = 80{,}000$ estimations and 8,000 parameter values based on objective minimizing values. Only the demand side is estimated. For computational details, see Appendix A.6.

There is evidence that simultaneous estimation of demand and supply increases accuracy of parameter values if the supply side is correctly specified, as discussed by Reynaert and Verboven (2014). For the Monte Carlo study conducted presently, there is no estimation of supply side parameters and there is no combined objective of demand and supply shocks. I opted for a demand only approach to keep it simple, to avoid additional assumptions and lower computational burden. As the results of Reynaert and Verboven (2014) indicate, given optimal instruments are available, the additional gains from simultaneous estimation of demand and supply are comparatively small. This moves the discussion to the use of appropriate instruments to correct for endogeneity and for efficiency reasons.

4.2 Calculation of Instruments

In this Monte Carlo study, two types of instruments are used. The first type are easy to compute instruments as used in the seminal work of BLP. They are calculated based on sums over characteristics of all other products of the same firm and sums over characteristics of all competing firms, formally for product j, characteristic k and firm f (suppressing time subscript):

$$g_j^*(\tilde{x}) = \left\{ \tilde{x}_{jk}, \sum_{r \neq j, r \in F_f} \tilde{x}_{rk}, \sum_{r \neq j, r \notin F_f} \tilde{x}_{rk} \right\}. \tag{13}$$

Armstrong (2016) investigates properties of instruments based on characteristics and finds instruments to perform worse in situations with few markets and many products. Markets with many products might therefore face an identification problem. Reynaert and Verboven (2014) show that efficiency gains can be realized by calculating approximations to optimal instruments based on Chamberlain (1987). Following the notation of Reynaert and Verboven, the optimal instruments for product j, and the second type of instruments used in this study, are given by

$$g_j^{**}(\tilde{x}) = E\left[\frac{\partial \xi_j(\theta)}{\partial \theta} \middle| \tilde{x}_j\right]' \Omega^{-1} \tag{14}$$

with $\tilde{x}$ being exogenously given values (all characteristics except price) and Ω is a matrix defining the covariance structure of ξ, ω. The inner part of the expectation is known already from the calculation of the gradient. Since only the demand side is estimated, the matrix Ω is set to identity (cf. Reynaert and Verboven, 2014), facilitating calculation of the instruments.

Since all parameter values are assumed to be known, it is possible to back out optimal instruments and survey their performance relative to BLP instruments using the flexible estimation procedure presented here. For simplicity, a first stage to approximate optimal instruments during estimation will not be implemented but instead optimal instruments calculated when generating the market equilibrium and assumed to be known prior to estimation. This is not feasible in real world applications but suitable in the present case since the main focus of the Monte Carlo study is to show the consistency of the flexible estimator.

I calculate an approximation to optimal instruments as presented by Reynaert and Verboven (2014) and Berry, Levinsohn, and Pakes (1999). The approximation "replaces the expected value of the derivatives [...] by the derivatives evaluated at the expected value of the unobservables" (Reynaert and Verboven, 2014, p. 96), which is easy because the expectation of ξ and ω is zero. I will refer to these second kind of instruments as instruments evaluated "at zero" or "approximately optimal". Ease of computation is weighted against a bias as stated by Berry, Levinsohn, and Pakes (1999) and is straightforward to see by Jensen's inequality (cf. Cameron and Trivedi, 2005).

Marginal costs and utility simplify and the market equilibrium can be calculated by simultaneous iteration over the first order derivatives. Then, the approximation to the optimal instruments can be calculated according to Equation (14), which simplifies to the inner derivative of the gradient as derived in Appendix A.3. Integration is done deterministically with 1,000 equally spaced random coefficient values.

As discussed by Reynaert and Verboven (2014), I proceed similarly and enrich instruments g_j^* by calculating the expected price based on a regression. This regression provides the fitted values of price regressed on characteristic $\tilde{x}_{jt}$, its squared value, and the values obtained from g_j^* (except x_j to prevent collinearity). The resulting values are incorporated as an additional instrument.

4.3 Discussion of the Results

4.3.1 Visual Inspection

To start off, an intuitive approach to check for consistency is by graphically inspecting the resulting shapes of the distributions based on the number of markets. Respective parameter values are calculated by taking the arithmetic mean over all estimated values. Figure 1 shows the resulting shapes dependent on the number of markets for the BLP instruments. The solid line represents the true distribution. Note that all values are probability masses based on a support of 100 discrete values (but plotted as a smooth function). The distributions for 25, 50 and 100 markets are overly broad and do have a sign flip on its parameter value, as the distributions are opened upwards. They are poor representations of the true distribution. The distribution for 100 markets is still flipped over, but at the same time not as wide. The distribution for 200 markets seems to grasp the shape of the true distribution with appropriate width and opening downwards. There is still some room left for improvement as the backed out distribution does not fully show the high concentration around the expected value and instead exhibits fat tails. The results of the specification using BLP instruments seem to indicate consistency.

Figure 3 shows the resulting shapes for the specification using approximately optimal instruments. The distribution for 25 markets matches closely to the true distribution with only a little shift to the left, capturing the shape of the true distribution almost exactly. In the case of 50, 100 and 200 markets, the shape of the true distribution is captured so well that it is not visually possible to tell apart the true distribution from the estimated distribution. The results obtained by using approximately optimal instruments clearly indicate consistency.

Figure 1: True and Averaged Estimated Probability Masses of ν_r for a Second Order Polynomial and BLP Instruments

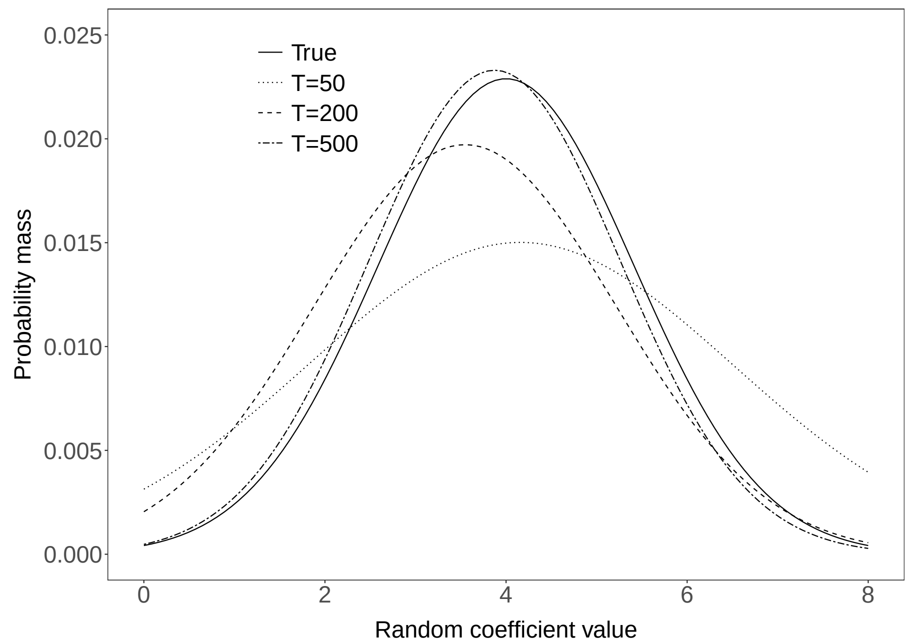


Figure 2: True and Averaged Estimated Probability Masses of ν_r for a Second Order Polynomial and Approximately Optimal Instruments

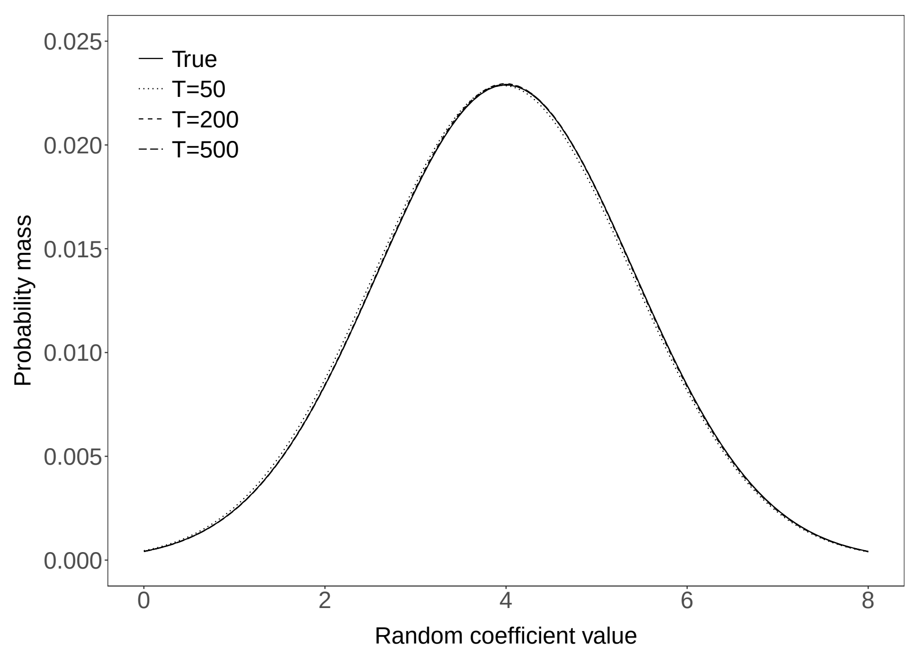


Table 1: Parameter Summary Statistics for a Distribution based on a Second Order Polynomial as z Variable and BLP Instruments

Markets	θ_0	$\bar{\hat{\theta}}$	$\tilde{\hat{\theta}}$	$\theta_{[0.025]}$	$\theta_{[0.975]}$	St. Err.
θ_{21}						
50	0	−0.698	−0.698	−2.842	2.065	5.094
200	0	−0.114	−0.114	−0.866	0.973	0.482
500	0	−0.032	−0.032	−0.562	0.672	0.304
θ_{22}						
50	−4	−0.326	3.674	−10.903	3.961	62.222
200	−4	−3.919	0.081	−5.856	−0.292	1.456
500	−4	−3.987	0.013	−5.196	−2.231	0.781
$\theta_{11} = \alpha$						
50	5	4.256	−0.744	−0.907	11.610	3.120
200	5	4.984	−0.016	2.839	9.400	1.615
500	5	5.012	0.012	3.540	7.315	1.019
$\theta_{12} = \beta_0$						
50	−2	−3.990	−1.990	−18.000	16.448	8.563
200	−2	−2.027	−0.027	−7.870	10.021	4.431
500	−2	−1.963	0.037	−6.003	4.327	2.778

4.3.2 Convergence of Parameter Estimates

Table 1 and Table 2 show statistics conditional on the estimated parameter values and on the number of markets. Reported are the true parameter values θ_0, the average of all estimated parameter values $\bar{\hat{\theta}}$ and the bias $\tilde{\hat{\theta}}$ as defined by $\tilde{\hat{\theta}} = \text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta_0$, with the expectation replaced by the sample moment. The empirical 95% confidence thresholds are reported along with the standard error.

Table 1 shows statistics for the specification using BLP instruments. The table is structured by displaying the nonlinear parameters θ_{21} and θ_{22} first. They shape the distribution of the random coefficient and are followed by the linearly estimated parameters, $\theta_{11} = \alpha$ and $\theta_{12} = \beta_0$. The first subscript indicates linearity (subscript equals 1) or nonlinearity (subscript equals 2), whereas the second subscript enumerates respective values. Next to the column with the true parameter values θ_0 on the left side are the estimated parameter averages. It is straightforward to see that all estimated averages converge towards the true parameter values as the number of markets increases. The empirical confidence intervals clearly tighten around the true values, e.g., the first nonlinear parameter θ_{21} shaping the distribution of $\tilde{x}_{jt}$ going from -6.231 and 3.637 at 25 markets to -0.861 and 0.914 at 200 markets with a true value of 0. The price coefficient α is increasingly precisely estimated, with a value of 3.274 at 25 markets and 4.867 at 200 markets with a true value of 5. This is especially important as it shows that endogeneity is taken into account and instrumental variables are valid.

Table 2 shows statistics for the specification using approximately optimal instruments. The results are similar in the sense that parameters seem to converge to their true counterparts. At the same time, the convergence is stronger, e.g., in the case of θ_{21} the empirical confidence interval is estimated to be -1.476 and 2.244 at 25 markets and closes around -0.536 and 0.714 at 200 markets with a true parameter value of 0. All values are estimated precisely, the estimated average of the price coefficient α is estimated as 5.011 with a true parameter value of 5 and the intercept estimated as -1.968 at 200 markets with a true parameter value of -2. For 100 and 200 markets the precisely estimated parameters mildly fluctuate around the true value. It is explained by the calculation as means and by outliers when looking at the confidence interval actually strongly tightening. Comparing the estimated values at 200 markets to the estimated values using BLP instruments, it can be said that the approximately optimal instruments increase precision by a substantial amount.

Next, I move on to discuss and compare the convergence of a single parameter value based on the empirical distribution. Figure 3 and 4 view histogram and density for one selected parameter value dependent on market size. The dashed vertical line indicates the true parameter value. Figure 3 (Figure 4) captures the convergence of the parameter for BLP instruments (approximately optimal instruments). In both cases consistency is visible. Firstly, the peak of the distribution matches somewhat to the true value in the case of BLP instruments and matches closely to the true value for approximately optimal instruments. Secondly, the width of the distribution narrows down with increasing sample size, indicating a decrease in variance. The approximately optimal instruments display a more pronounced convergence, as the peak is more closely situated at the true parameter value at each sample size and the variance appears to be smaller in each case compared to BLP instruments. This underlines the efficiency of the approximately optimal instruments.

Figure 11 in Appendix A.7 plots the convergence of the price coefficient by market. A convergence to the true parameter value can be observed. The peak of the distribution moves ever more closely to the true value and the variance steadily decreases as the number of observations grows. Apparently, the endogeneity of price induced by the correlation between the price and the unobserved characteristics is fully accounted for as long as the instruments are valid. BLP instruments again perform worse compared to

Table 2: Parameter Summary Statistics for a Distribution based on a Second Order Polynomial as z Variable and Instruments at the Expected Value of the Structural Shocks

Markets	θ_0	$\bar{\hat{\theta}}$	$\tilde{\hat{\theta}}$	$\theta_{[0.025]}$	$\theta_{[0.975]}$	St. Err.
θ_{21}						
50	0	−0.067	−0.067	−1.046	1.668	0.683
200	0	−0.013	−0.013	−0.548	0.669	0.317
500	0	0.001	0.001	−0.341	0.398	0.192
θ_{22}						
50	−4	−3.991	0.009	−6.155	−1.181	1.239
200	−4	−4.021	−0.021	−4.881	−2.941	0.477
500	−4	−4.012	−0.012	−4.581	−3.409	0.310
$\theta_{11} = \alpha$						
50	5	4.931	−0.069	2.789	8.624	1.514
200	5	4.962	−0.038	3.771	6.539	0.704
500	5	5.002	0.002	4.245	6.006	0.453
$\theta_{12} = \beta_0$						
50	−2	−2.174	−0.174	−8.045	7.850	4.120
200	−2	−2.105	−0.105	−5.301	2.114	1.909
500	−2	−1.993	0.007	−4.025	0.694	1.225

approximately optimal instruments. Distributions for BLP instruments are not plotted for brevity.

Apart from consistency, the distribution of the estimator is indicative to be roughly approximated by a normal distribution. Looking at the density plotted along the histogram of the various figures, e.g., for the second nonlinear parameter in Figure 4, the shape of the density seems to be somewhat fitting to a normal distribution. All versions with a higher number of markets seem to roughly fit a normal distribution. This is reassuring as one can be confident that the estimator exhibits the properties of nonlinear generalized method of moments and thus allows inference.

A disclaimer in this regard is the slight skewness to the right that can be observed in the density plots in Appendix A.5 and A.7, which is a possible threat to inference.

Figure 3: Monte Carlo Convergence of Parameter θ_{22} for a Second Order Polynomial as z Variable and BLP Instruments

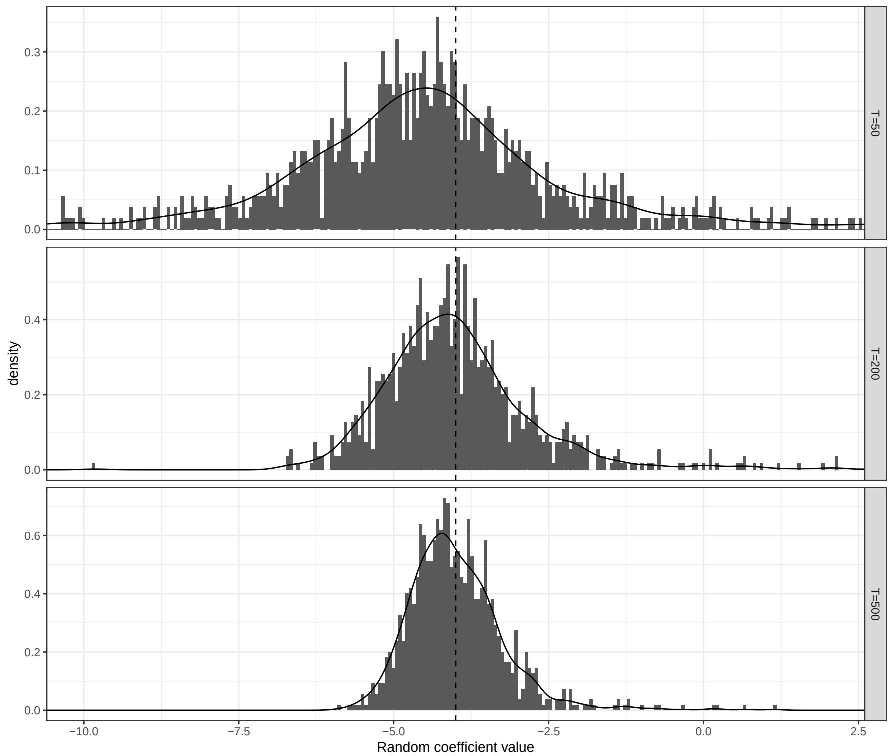
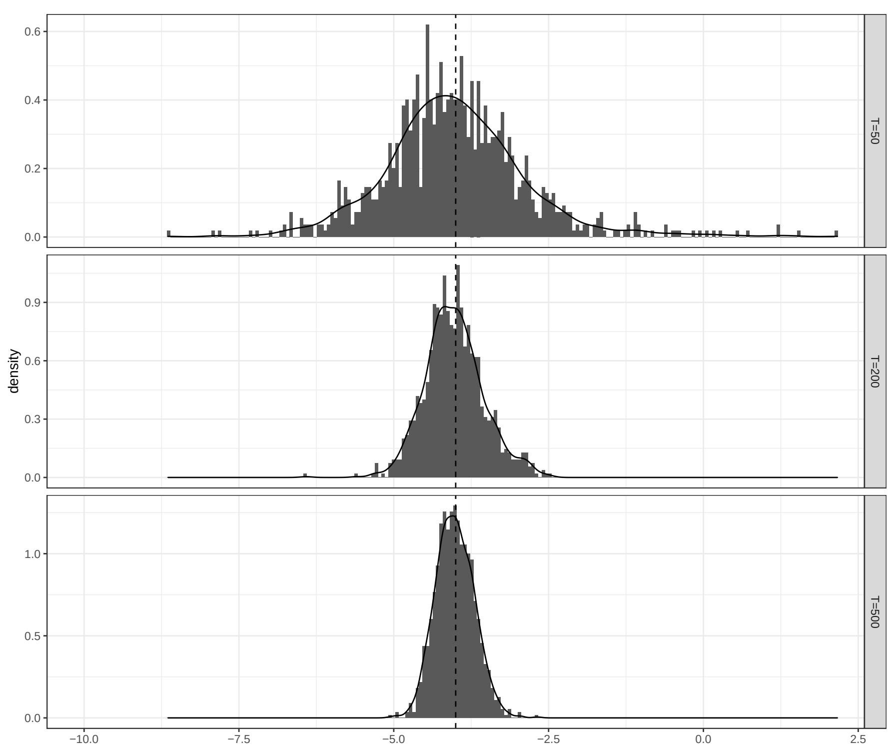


Figure 4: Monte Carlo Convergence of Parameter θ_{22} for a Second Order Polynomial as z Variable and Instruments at the Expected Value of the Structural Shocks



4.3.3 Convergence of Aggregate Values

Table 3 provides detail on the estimation of structural parameters, namely convergence of the demand shocks and the marginal costs. Also, a measure for the convergence of respective distributions is presented. All values are averages over all estimations. The left column shows the average absolute deviation of the estimated shock term from the true shock term $\Delta\xi_j$ of a random product. Similarly, the column labeled Δmc_j shows the average absolute deviation of the estimated marginal cost from the true marginal cost of a random product. Both cases, g_j^* (BLP instruments) and g_j^{**} (approximately optimal instruments), show clear convergence, with the bias being less pronounced for g_j^{**}. To gain a more overall picture, the third column displays the root mean squared error (RMSE) for the vector of structural shocks. The RMSE is defined by $\text{RMSE}(\hat{x}) = \sqrt{E[(\hat{x} - x_0)^2]} = \ddot{\hat{x}}$, with the arithmetic mean replacing the population moment. I rescale the RMSE by a factor of 1000. A falling RMSE generally indicates a better fit. A clear decrease of the bias can be observed. The bias of the specification with approximately optimal instruments is less pronounced for all numbers of markets compared to BLP instruments. The value of the RMSE for the structural demand terms drops for specification g_j^* from 0.597 at 25 markets to 0.231 at 200 markets, whereas the RMSE drops from 0.345 to 0.124 with regard to specification g_j^{**}. Similar results hold for the the convergence of marginal costs; noteworthy is the comparatively rather dramatic decrease in the bias, as soon as additional markets enter the sample. The right column shows the RMSE of the probability masses for the taste distribution of $\tilde{x}_{jt}$, $\ddot{\hat{f}}_{\tilde{x}_{jt}}$, and is an indicator of the convergence with respect to the underlying functional form. Again, the bias decreases in both cases and approximately optimal instruments estimate the underlying parameters more closely.

4.3.4 Additional Comments

Section 4 is now completed with a summary and interpretation of the results. First and most importantly, the study indicates that the proposed estimator is largely consistent. The Monte Carlo study shows a variety of different statistics to converge to their true counterparts. A necessary assumption is that the support of the random coefficient is correctly specified, and its underlying functional form is flexible enough to capture the true shape of the distribution. Additionally, the instruments need to be mean independent of the structural shock terms and sampling and simulation error must be small. The latter is most easily guaranteed if only one parameter for consumer heterogeneity is assumed, as a deterministic approach to integration avoids rather large fluctuations induced by simulation.

Table 3: Summary Statistics of Structural Convergence

Markets	$\Delta\xi_j$	Δmc_j	$\ddot{\hat{\xi}}$	$\ddot{\hat{mc}}$	$\ddot{\hat{f}}_{\tilde{x}_{jt}}$
g_j^*					
25	0.479	1.303	0.597	12.241	3.129
50	0.353	0.945	0.460	2.413	2.261
100	0.264	0.342	0.336	1.275	1.542
200	0.191	0.333	0.231	1.040	1.012
g_j^{**}					
25	0.273	0.445	0.345	1.913	1.690
50	0.206	0.333	0.259	0.840	1.156
100	0.148	0.269	0.181	0.599	0.746
200	0.100	0.281	0.124	0.557	0.514

Secondly, there is evidence that the estimator exhibits the behavior of a normal distribution. Based on the exclusion of sampling and simulation error, the estimator is established to follow the rules of nonlinear generalized method of moments. This can be seen by inspecting the empirical distributions of the estimated values. The results are in line with the theoretical findings in Berry, Linton, and Pakes (2004).

Thirdly, the theoretical results of Fox et al. (2012) discussing nonparametric identification of random coefficient logit models seem to be confirmed by the results of the Monte Carlo study. The parameters of the distribution as well as the parameters of the model appear to be uniquely identified.

Fourthly, it can be reported that estimating more sophisticated shapes with aggregate data only is a difficult enterprise. I estimate a similar specification with a fourth order polynomial and convergence requires more than 10 times the amount of observations. For details, see Section 5. The weakened convergence is straightforward so see by adding 2 additional monomials to the second order polynomial and replicating the Monte Carlo study described in this Section.

Fifthly, a part of the observations of Reynaert and Verboven (2014) can be replicated, namely that the use of Chamberlain's (1987) optimal instruments improves efficiency and reduces small sample bias. Comparing the results using BLP instruments and the results using approximately optimal instruments it is clear that in this set up BLP instruments are almost always inferior in efficiency. It is necessary to stress that the instruments are here assumed to be available at high accuracy, which will not be the case in practice.

5 Monte Carlo Extended: Using Micro Data

By circumventing the normality assumption, estimation of more flexible distributions can pose a challenge. These models often face large data requirements and are subjected to the curse of dimensionality problem, resulting in serious challenges for practitioners. It is thus paramount to investigate ways to estimate models flexibly without being constrained too much by the modeling assumptions and the amount of data required.

The flexible estimator presented here offers these kinds of benefits. Moving from a normality assumption to a flexible structure comes at the cost of having more parameters to estimate consistently, which usually requires more data. Yet it is possible to show how the availability of precise micro data can alleviate estimation and make it possible to estimate highly flexible distributions with only limited data requirements. I therefore investigate the estimator's behavior with a more sophisticated distribution of consumer tastes and discuss an appropriate implementation.

The aim of this Section is to show that the estimator can be used with micro data and consistently estimates arbitrary distributions. I therefore firstly estimate a flexible estimation with a bimodal underlying distribution without micro data and then repeat the estimation with micro data and compare the results. The procedure is similar to Petrin (2002) and Berry, Levinsohn, and Pakes (2004). I compare the performances of estimation with and without micro data and find that using micro data makes it possible to uncover sophisticated distributional shapes that are not easily estimated in the absence of micro data.

I proceed as follows. Section 5.1 presents a simple way to add micro data to the model, whilst Section 5.2 discusses an implementation and the results.

5.1 Specification and Introduction of Micro Moments

Let D_r be consumer specific information, e.g., family size or income, and $\mathcal{P}(\mathrm{D})$ the associated demographic distribution. For simplicity, its mean is assumed to be zero. Let d_j be a dummy such that $d_j = 0$ if $\tilde{x}_{jk} < \underline{\tilde{x}}_{jk}$ and $d_j = 1$ otherwise. The threshold value $\underline{\tilde{x}}_{jk}$ distinguishes segments. The dummy enables different segments of products in a market to be separated and makes it possible to calculate expected demographics conditional on segments. Let ζ be a utility shifter which is estimated. Then write utility as

$$u_{rj} = x_j\beta + \xi_j + \tilde{x}_j(\nu_r + \mathrm{D}_r\zeta d_j) + \epsilon_{rj}. \tag{15}$$

The associated market shares are the same as in Equation (6) with the addition of the demographics in the nonlinear part of utility, i.e., the idiosyncratic shock is tier I extreme value distributed, resulting in the logit model. The nonlinear utility is now given by $\mu_{rj} = \tilde{x}_j(\nu_r + \mathrm{D}_r\zeta d_j) + \epsilon_{rj}$. I refrain from adding several nonlinear attributes or demographics. This keeps notation at a minimum, as the sum notation can be left out and generalization to multiple demographics is straightforward.

Estimation proceeds similarly as before. The distribution $\mathcal{P}(\mathrm{D})$ is either known and random draws can be inserted during estimation or the distribution is estimated along the other parameters if unknown. A derivation of the partial derivatives is presented in Appendix A.2.

A crucial change are the micro moments that are appended to the BLP moments. To form the micro moments it is necessary to calculate the expectations of demographics conditional on a specific product, j, or conditional on a segment, o. This also conditions on the occurrence of a purchase. The model implied expected demographics of a product, j, are then approximated by multiplying each consumer type specific weighted market share by its corresponding demographic value and averaging over consumer types,

$$\mathbb{E}[\mathrm{D}|j] = S_j^{-1} \sum_{r \in S} W_r \cdot s_{rj} \cdot \mathrm{D}_r, \tag{16}$$

with S_j referencing the observed market share of product j. The model implied expected demographic value conditional on a specific segment, o, is approximated similarly by conditioning on the segment and additionally summing up across all products of the segment,

$$\mathbb{E}[\mathrm{D}|\text{type} = \text{o}] = S_o^{-1} \sum_{j \in o} \sum_{r \in S} W_r \cdot s_{rj|o} \cdot \mathrm{D}_r, \tag{17}$$

with S_o referring to the market share of segment o.

The expected demographics based on the micro data are calculated as

$$\bar{\mathrm{D}}_j = R_j^{-1} \sum_{r=1}^{R_j} \mathrm{D}_{r|j}, \tag{18}$$

$$\bar{\mathrm{D}}_o = R_o^{-1} \sum_{r=1}^{R_o} \mathrm{D}_{r|o}, \tag{19}$$

with R_j indicating the number of all consumers observed to purchase product j and R_o indicating all consumers sampled in segment o. The micro moments are then formed by the squared difference (stacked over all products and markets, indicated as bold) and weighted by the inverse of the moment variance

$$G_{22} = \mathbb{E}[\mathbf{D}|\boldsymbol{j}] - \bar{\mathbf{D}}_{\boldsymbol{j}},$$
$$G_{33} = \mathbb{E}[\mathbf{D}|\mathbf{type} = \mathbf{o}] - \bar{\mathbf{D}}_{\boldsymbol{o}},$$
$$G_2 = G'_{22}G_{22}/\sigma^2_{G_{22}},$$
$$G_3 = G'_{33}G_{33}/\sigma^2_{G_{33}}$$

and appended to the BLP moments so that the objective can be written as $G = G_1 + G_2 + G_3$. See Appendix A.8 for a derivation of the gradient.

5.2 Monte Carlo Study: Micro Data

To see the flexible estimator in action and investigate its performance with micro data, I conduct a Monte Carlo study, similarly as in Section 4. Consumer utility depends on price, a constant, a randomly generated characteristic, $\tilde{x}_{jt}$, and on a randomly generated demand shock correlated with a supply side cost shock. As before, consumer utility additionally depends on a consumer, product specific arbitrary shock term (logit error), but now also on consumer demographics. Consumers are heterogeneous with respect to the product characteristic $\tilde{x}_{jt}$, as indicated by tilde. Firms compete Nash in prices. Marginal costs are similarly structured as in Section 4. For details on the specification of the model, see Appendix A.9.

I first run 100 Monte Carlo repetitions for 2000 markets with four firms in each market and each firm providing three products, i.e., there are $24,000$ observations available for estimation. In this first run, I do not utilize micro data during estimation and use only BLP moments and an identification assumption based on exogenous characteristics to estimate the model, labeled 'No Micro'. I secondly run 100 Monte Carlo repetitions, now for 100 markets and the same industry structure, i.e., 1200 observations are available to the econometrician. The model is then estimated with micro data generated for each product as described in Section 5.1. This second micro data based run is labeled 'Micro'.

The results can be seen in Table 4. I report the true values, θ_0, along with an estimate of the median, $\hat{\theta}_{[0.5]}$, and the bias, $\tilde{\hat{\theta}}$, as defined by the deviation from of the true value from the median. The last columns show empirical quantiles and the interquartile range (IQR). Each block separated by horizontal lines is a comparison of an estimate between the version with and without micro data. The names of the

Table 4: Comparing Micro to Non Micro Estimates with Bimodal Underlying Distribution

Specification	θ_0	$\hat{\theta}_{[0.5]}$	$\tilde{\hat{\theta}}$	$\hat{\theta}_{[0.025]}$	$\hat{\theta}_{[0.975]}$	IQR
θ_{21}						
No Micro	-1.5	-2.085	-0.585	-13.960	12.487	5.196
Micro	-1.5	-1.506	-0.006	-1.507	-1.375	0.000
θ_{22}						
No Micro	5	4.144	-0.856	-36.414	9.613	4.715
Micro	5	4.997	-0.003	2.743	4.998	0.000
θ_{23}						
No Micro	3	4.651	1.651	-14.282	103.251	16.113
Micro	3	3.016	0.016	1.534	3.017	0.001
θ_{24}						
No Micro	-6	-8.140	-2.140	-85.859	12.565	14.892
Micro	-6	-6.005	-0.005	-6.006	-2.940	0.000
$\theta_{11} = \alpha$						
No Micro	6	5.997	-0.003	5.969	6.034	0.021
Micro	5	4.989	-0.011	4.795	5.297	0.198
$\theta_{12} = \beta_0$						
No Micro	-4	-4.011	-0.011	-4.121	-3.881	0.090
Micro	9	8.979	-0.021	8.371	9.931	0.607

'No Micro' estimated on $24,000$ observations without micro data. 'Micro' estimated on $1,200$ observations with micro data. Columns: θ_0: True parameter values. $\tilde{\hat{\theta}}$: Bias of the estimates. $\hat{\theta}_{[x]}$: x quantile of the estimates. IQR: Interquartile range.

specifications are self explanatory. It is expedient to work with empirical quantiles, as the estimation of deep parameters can sometimes produce large deviations from the true values. Note that this does not imply a poor estimation of the post estimation parameters, e.g., marginal costs or demand shocks.

The results show that, despite a large advantage of the 'No Micro' model in observations over the 'Micro' model, that the model with micro data produces vastly superior estimates. The first two rows show statistics for parameter θ_{21} with a true value of -1.5. The median for 'No Micro' is estimated to be -2.085 compared to an estimated value of -1.506 for 'Micro' with a bias of -0.585 and -0.006, respectively. The other deep parameters, θ_{22}, θ_{23} and θ_{24}, show qualitatively similar results. The empirical quantiles show large diversions from the true values for 'No Micro' and slight deviations for 'Micro'. Due to the semi-nonparametric nature of the model, abnormal estimates are collected on the fringe. To get a better picture of the dispersion, I report the interquartile range (IQR), as defined by IQR= $\hat{\theta}_{[0.75]} - \hat{\theta}_{[0.25]}$. Here it can be seen that there are vast differences in the estimated values. As far as 'No Micro' is concerned, estimates differ widely. The 'Micro' specification is estimated almost exactly, with a slight bias remaining due to simulation inaccuracy. The IQR of the first deep parameter θ_{21} is estimated to be 5.196 for No Micro and 2.569e−4 for 'Micro', which differs by a factor of over $20,000$. This observation holds for all deep parameters. The homogeneous taste parameters are estimated comparably less precisely in the 'Micro' case when compared to 'No Micro', but still precise. The relative imprecision of the homogeneous parameters is due to the small sample size of 'Micro', which increases the variance of the two stage least square regression results.

I also estimate a version in which I only use micro moments in the estimation, obtaining largely similar results. I do not report the results for brevity. In line with the results in Berry and Haile (2022), the reliance on instruments is largely reduced by introducing micro data; instruments are only needed to determine the magnitude of the price coefficient.

All in all, the results show how a rather sophisticated shape can be estimated with quite limited data requirements. Estimating a bimodal distribution is a difficult undertaking, yet the flexible model is able to consistently estimate the parameters with only $1,200$ observations, given that low measurement error micro data is available.

I want to close out with a few remarks: the variances of the market shares and micro data are kept to a minimum. Whether these assumptions can be justified, has to be decided on a case by case basis. Integration of the market shares is assumed to be done at high accuracy, which can be guaranteed if enough computational power is available. The specification presented here is further stylized and might seem simplistic, yet it is similar to specifications often found in contemporary empirical research and is thus expected to be useful in the future.

6 Effects of Distributional Misspecification

The present Section aims at highlighting the potential importance of taking arbitrarily distributed taste heterogeneity into account when estimating demand and simulating mergers. I do so by comparing market equilibria and estimating respective parameters with different underlying taste distributions.

I proceed similarly as in Section 4 and Section 5. The taste parameter concerning the price is assumed to be heterogeneous across consumers. The empirical setup is unchanged with minor changes to the true coefficient values and the grid, ensuring that the equilibrium can be established. Approximately optimal instruments and two additional cost shifters are used as instruments. For additional details, please refer to Appendix A.10.

Section 6.1 compares BLP and flexible estimator true market equilibria, i.e., the estimation parameters are assumed to be known. Additionally, I simulate a merger for each model to compare the results. Section 6.2 moves on and estimates the parameter values using the BLP model, when the true underlying bimodel distribution is generated by the flexible model and assesses the performance of the BLP estimator. Section 6.3 discusses the results.

6.1 Comparing True Market Equilibria

I now investigate whether it is any different to model the market equilibrium based on a normality assumption, or by a logit mixed logit model with a second order polynomial as z variable or based on a fourth order polynomial as z variable, which I assume to be a bimodal distribution. For ease of language, I refer to the models in the following as BLP, Second Order and Fourth Order, respectively. In this Section there is thus no estimation of any model, only calculated true market equilibria are compared. To rule out any differences due to first and second order moments or asymmetries of the distribution, I choose expectation and variance to be equal for all distributions and model the logit mixed logit distributions to be symmetric. This allows a direct comparison of the different underlying distributional assumptions and its implied structural parameters. After generating the data, I simulate a merger between two firms and report respective statistics, namely welfare change, average price changes of the merging parties and the change of the Lerner index (all statistics in percent), as well as average cross and own elasticities prior to the merger.

The results can be seen in Table 5. The first and second columns show means ($\bar{x}_{conc}$) and respective standard errors (SE_{conc}) over 100 markets and 100 Monte Carlo repetitions within a relatively concentrated industry, with three firms offering two products each. The results based on BLP differ quite considerably from the results based on the logit mixed logit model, with the largest difference to the bimodal distribution, and the results of the second order polynomial as z variable in a middle ground. The percentage welfare loss is calculated to be approximately 1% in the BLP model, 0.77% in the Second Order case and 0.26% for the bimodal distribution. The price changes of the merging parties are backed out as 5% for BLP, 3% for the Second Order and

Table 5: Summary Statistics for Market Equilibria of BLP (Normality) and LML (Second and Fourth Order Polynomial) Simulating a Merger

	$\bar{x}_{conc}$	SE_{conc}	$\bar{x}_{comp}$	SE_{comp}
Welfare Change (%)				
BLP	−1.019	(0.134)	−0.386	(0.031)
Second Order	−0.774	(0.007)	−0.358	(0.004)
Fourth Order	−0.257	(0.003)	−0.137	(0.001)
Price Change (%)				
BLP	5.000	(0.000)	2.680	(0.000)
Second Order	3.006	(0.000)	2.158	(0.000)
Fourth Order	1.462	(0.000)	1.312	(0.000)
Lerner Change (%)				
BLP	22.181	(0.109)	10.874	(0.051)
Second Order	17.304	(0.064)	9.935	(0.049)
Fourth Order	10.453	(0.045)	7.005	(0.032)
Cross Elasticities				
BLP	0.723	(0.002)	0.418	(0.001)
Second Order	0.811	(0.002)	0.466	(0.001)
Fourth Order	0.646	(0.002)	0.410	(0.001)
Own Elasticities				
BLP	−7.060	(0.006)	−8.110	(0.004)
Second Order	−7.948	(0.003)	−8.668	(0.003)
Fourth Order	−9.104	(0.004)	−9.618	(0.002)

The subscript "conc" indicates a more concentrated industry, whereas the subscript "comp" indicates a more competitive industry. Generated with 100 Monte Carlo repetitions using 100 markets each. Markets consist of 3 firms offering 2 products each in the more concentrated case, and 4 firms offering 3 products each in the more competitive case. Heterogeneous price coefficient, intercept and consumer characteristic preference assumed to be linear. Aggregation is done by weighting with respective market shares.

1.46% for the bimodal distribution, as well as increases of the Lerner index of 22%, 17% and 10% due to the merger, respectively. These results show considerable differences between distributional assumptions. Whereas the BLP model overestimates the welfare loss with a factor of four compared to the bimodal distribution, BLP overestimates the price increase of the merger and overpredicts the increase in market power indicated by the Lerner index. This picture is reflected in the own elasticities, with BLP estimating the average own elasticity to be -7.06, with the Second Order estimate being -7.95 and the bimodal estimate being -9.10, i.e., BLP may underestimate the quantity effect by roughly 2 percentage points in this setting, if one estimates by BLP, but the underlying distribution is truly bimodal. The cross elasticities show no such hierarchical ordering; the cross elasticities estimated by BLP are intermediate with an average value of 0.72, the Second Order estimate is 0.81 and the bimodal estimate being 0.65.

One may wonder why the results of BLP and Second Order differ at all, given that both distributions have the same mean and variance and they also share the same shape. Welfare and price changes could be expected to be largely equal. The difference can be explained by the support of the random coefficients. Since the grid of the logit mixed logit model limits the support of the random coefficients, rare outliers are impossible in the flexible specification, whereas the support of the normal distribution is unlimited. These rare outliers are amplified when entering the exponential function and spill over into welfare estimates. The normality assumption in BLP can be seen as a special case of the logit mixed logit model by stretching to grid to infinity. Running a specification for Second Order with a stretched grid, whilst keeping the variance constant, produces the same results as in the BLP model. The flexible estimator is thus more general and accommodates all cases in between a single linear taste and an infinite support for tastes.

The two right columns show the results for a more competitive industry, with four firms offering three products each, indicated by "comp". The results are similar from a qualitative perspective. Looking at the magnitude, it turns out that problems are more aggravated in the concentrated industry, which is intuitive, as effects are spread over more market participants and products in the less concentrated industry. Following this argument, the price effect of BLP is $5/1.46 = 3.42$ times the effect of the bimodal distribution in the concentrated industry, compared to a factor of $2.68/1.31 = 2.05$ times in the less concentrated industry.

6.2 Estimating BLP when Consumer Tastes are Non-Normally Distributed

Next, I investigate what happens when the true data is generated by an underlying bimodal taste distribution, but estimated with BLP, thus wrongly modeling the distribution to be normally distributed. I generate 1000 datasets as described in Section 5.1, but change the number of markets to 1000 to minimize small sample bias. As in the previous Section, the distribution of the tastes for the price coefficient is modeled to be bimodal, which is achieved by setting the z variable to a fourth order polynomial. The distribution can be described as symmetric with two pronounced peaks and a large trough in the middle. There are three firms in the market, each offering two products. For more details, please refer to Appendix A.10.

Table 6 shows the estimated standard deviation and the associated linear parameter of the random coefficient, α, along with the parameters for homogeneous tastes, i.e., the parameter for the constant, β_0, and the parameter for the characteristic, β_1. I report mean estimates, medians are essentially equal. The column with the BLP estimates can be directly compared to the true values in the right column. The standard deviation of the random coefficient is estimated to be 0.82, compared to a true value of 1.75. The estimated width of the shape is seemingly compressed in the estimation to fit the model to the data. The mean value of the nonlinear price coefficient is estimated to be too low, with an estimated value of 4.26 compared to the true mean of 5.50, implying that demand is estimated less elastic than it actually is, consistent with the results of Section 6.1. The intercept and the characteristic coefficients are estimated relatively well, with an intercept estimate of 3.71 compared to the true value of 4.00 and β_1 almost exactly estimated.

Table 7 shows aggregated parameters of the market equilibrium before and after a simulated merger. I report the root mean square percentage error (RMSPE) of elasticities, price changes, the Lerner index and model implied welfare. The root mean square percentage error is defined as $\text{RMSPE}(x_0, \hat{x}) = 100\sqrt{E[((\hat{x} - x_0)/x_0)^2]}$ and is a measure of the accuracy of the estimated values compared to the true values. The RMSPE of the own and cross price elasticities are estimated to be 7.41 and 13.98 respectively, indicating that the BLP model misses the true estimates of own and cross price elasticities by roughly 7% and 14%. The RMSPE value of the own and cross

Table 6: BLP Estimates When Truly Underlying Distribution is Bimodal

Variable	BLP	St. Err.	True
Standard deviation	0.82	(0.04)	1.75
Price coefficient (α)	4.26	(0.29)	5.50
Intercept (β_0)	3.71	(0.80)	4.00
Characteristic (β_1)	1.99	(0.06)	2.00

Estimated with 1000 Monte Carlo repetitions and 1000 markets in a concentrated industry (3 firms with 2 products each).

Table 7: Aggregate Statistics Bias When Estimated By BLP and Truly Underlying Distribution is Bimodal

Statistic	RMSPE(x_0,$\hat{x}$)
η_{jj}	7.41
η^a_{jj}	7.21
η_{jk}	13.98
η^a_{jk}	12.97
Δp(merging)	13.62
Δp(fringe)	35.55
Lerner	7.69
Lerner^a	7.84
Welfare	17.14

Aggregate statistics using estimates from Table 6. I report root mean square percentage error (RMSPE) between estimated and true values. All measures are weighted by market share. η_{jk} is the percentage market share response of product j due to a 1% increase in price of product k. The superscript 'a' indicates after merger results. Aggregation is done by weighting with respective market shares.

price elasticities after the merger are marked by a superscript 'a', and have similar magnitudes, showing that the bias carries over to the merger simulation results. The percentage price change of the merging parties is overestimated by 13.62% compared to the true values, and the percentage price change of the fringe parties is overestimated by 35.55% compared to the true values. In both cases, before and after the merger, the Lerner index is overestimated by roughly 8% compared to the true values. The welfare loss is underestimated by 17% compared to the true welfare loss.

6.3 Discussion

The results of the previous Sections show bias on two different scales. In Section 6.1, the biases are quite large, whereas in Section 6.2 the biases are more moderate. When estimating the BLP model on non-normally distributed tastes, the bias is not as pronounced, e.g., the ratio of the price changes calculated with BLP and the Fourth Order in Table 5 imply that price changes are overestimated by more than 100% in BLP, whereas in Table 7, it is implied that merging parties price increases are overestimated by 14%. The BLP model seems to be quite adaptive and minimizes the error induced by misspecification, at the cost of biased parameter estimates.

The BLP model does not estimate the implied welfare loss correctly. The percentage welfare loss of the merger is underestimated by 17%, which is a considerable bias given the misspecification. This is interesting because one may expect the welfare loss to be overestimated, since firms enjoy heightened market power. Yet this effect is counterveiled by the large tails in the true distribution, such that consumers that are situated off-center in the preference distribution incur relatively more losses due to increased prices, inducing a higher welfare loss with regard to the bimodal distribution.

By repeating the estimation with a different parameter values, I find the bias of the structural estimates of BLP to aggravate for skewed true underlying distributions and in more concentrated industries (cf. Section A.12 for more details). Moreover, the point estimates of the BLP standard deviation increasingly fluctuate with non-normally distributed true underlying distributions. Non-normality might thus potentially be a source of the reported numerical instabilities in BLP, albeit that numerical stability appears to be fostered by tight tolerances and accurate integration, as shown on the datasets used by BLP and Nevo (cf. Knittel and Metaxoglou (2014), Nevo (2000b), Dubé, Fox, and Su (2012), Brunner et al. (2017)).

To conclude, the results generally hint at the potential biases induced by a wrong specification of the distribution. The results seem to indicate that market power, price effects and welfare cannot be precisely predicted by BLP when subject to distributional misspecification, i.e., when the underlying true distribution cannot be approximated by a normal distribution. I further establish that the bias in the welfare estimate cannot be ignored when the true underlying distribution is symmetric. A more asymmetric distribution may introduce numerical instabilities and increase the bias of the estimates. More concentrated industries amplify the bias induced by misspecification.

7 Application to BLP car data

7.1 Set up, Specification and Data

Next, I apply the flexible estimator to a commonly used dataset. The dataset provides data on U.S. car sales from 1971 to 1990. Aggregate shares and prices are given along with important characteristics of the cars. Characteristics include horsepower, weight and the measures of a car along with fuel consumption and a dummy variable indicating the presence of an air conditioning system. The dataset contains $2,217$ observations. For more information on the dataset, see Berry, Levinsohn, and Pakes (1995).

The characteristics used to explain consumer utility are largely similar to BLP and Brunner et al. (2017). The ratio of horsepower to weight (Hpwt) serves as a proxy for the power of a car, air conditioning (Air) serves as a proxy for luxury, miles per gallon (Mpg) captures the consciousness of consumers towards energy efficiency and a space variable (Space) serves as a proxy for safety and preference for larger interior space and aesthetics (cf. BLP). The price variable is assumed to be endogenous and consumers price sensitivity is assumed to be heterogeneous. All other variables are treated as linear variables.

I estimate the flexible model with a second order polynomial (Flex 2nd) and a fourth order polynomial (Flex 4th) as z variables for price. The grid of the random coefficient for price is based on some initial experimentation with different specifications using the R package BLPestimatoR and set to $\alpha_r \in [-0.2, 8.2]$, assuming a mean of $\bar{\alpha} = 4$ and a standard deviation of 1.4 and the grid to span 3 standard deviations below and above the mean value. The results are then compared to results obtained from BLPestimatoR, which also treats price as the only heterogeneity term and all other characteristics as linear. BLPestimatoR is available to users of the programming language R, an open source statistical software environment. The estimation in R uses default tolerances and 10,000 modified latin hypercube sampling draws.

It is not clear whether the structural error terms are conditionally homoskedastic. I do not cluster by car model and use a two step GMM procedure and calculate a first consistent estimate using an appropriate identity as a weighting matrix. The weighting matrix is then recalculated based on the initial results according to $\hat{W} = (n^{-1} \sum_{i=1}^{n} \hat{g}_i \hat{g}_i{}' - \bar{g}_n \bar{g}_n{}')^{-1}$, with $\hat{g}_i$ as the individual moments based on an initial consistent estimate and $\bar{g}_n$ as the average of the moments (cf. Hansen, 2021). This is the procedure to obtain the asymptotically efficient weight matrix.

I use BLP instruments in the first step and then calculate the approximation to optimal instruments, $g_j^{**}(x)$, by iterating over the first order conditions until the prices are sufficiently stable and an equilibrium is assumed, similarly as discussed in Section 4.1. Then, in the second step, I use the approximated efficient weight matrix along with BLP instruments and the backed out approximately optimal instruments to estimate the second stage.

7.2 Discussion of the Results

Table 8 presents the results for the BLP model and the flexible model using second and fourth order polynomials. All estimates are point estimates. Standard errors are not reported. The first column contains the results for the BLP model estimated with BLPestimatoR. The mean marginal disutility of price is estimated to be 3.43. Its standard deviation is estimated to be 1.09 and highly significant, indicating that consumers are indeed heterogeneous with respect to the price to pay. The power proxy Hpwt is estimated to be 2.12. The average consumer thus receives positive utility when confronted with a power increase of the car, ceteris paribus. Air, Space and Mpg are estimated to be 1.14, 2.98 and 0.33, respectively. The signs are intuitive and as expected. The presence of an air conditioning system induces a utility gain as well as an increase in Space and fuel efficiency, represented by Mpg.

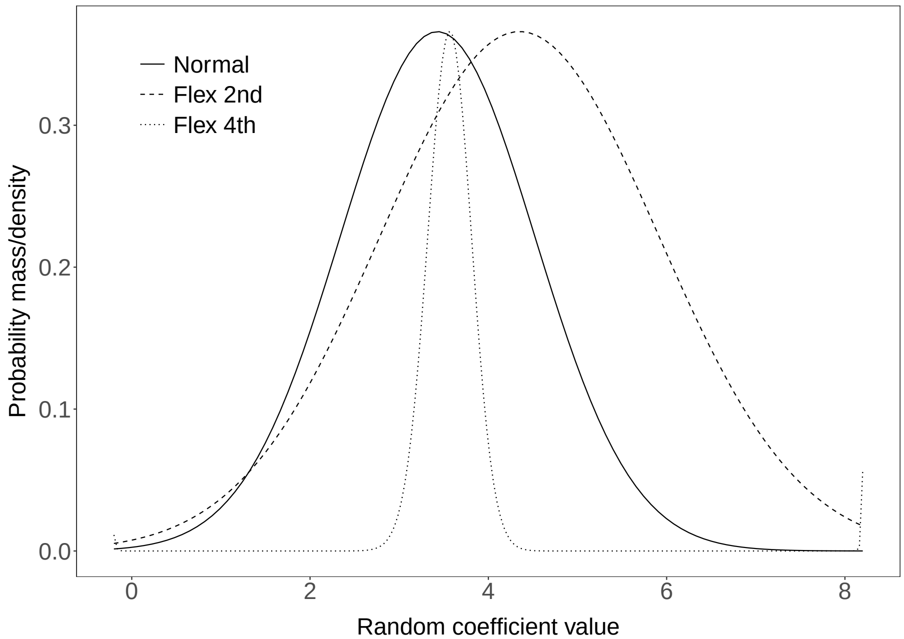
The second column shows the result of the flexible specification using a second order polynomial as z variable. The linear parameters and the mean of the price coefficient are comparable to the BLP specification. The mean of price is estimated to be 4.33 and consumers are therefore estimated to be more price sensitive on average compared to BLP. All other marginal utility parameters are estimated almost exactly the same as in BLP. The utility gain from Hpwt is estimated to be 2.21 and Air is estimated 1.12. Space and Mpg are estimated as 3.05 and 0.34, respectively. The last rows of Table 8 contain the underlying estimates of the z variables, $\theta_{21} = 0.60$ and $\theta_{22} = -3.61$. These parameters do not have a direct economic meaning but rather shape the distribution of the random coefficients. A negative second order term reveals that the shape of the logit distribution is opened downwards, similar to a normal distribution.

In the third column, the results of the specification using a fourth order polynomial as z variable can be seen. The parameter estimates are broadly in line with the other

Table 8: Results with BLP and Flexible Demand Using a Second and Fourth Order Polynomial to Specify Price

Variable	BLP	Flex 2nd	Flex 4th	Misspec 1	Misspec 2
Const	−9.79	−9.40	−9.51	−9.27	−9.41
Price	3.43	4.33	3.64	4.43	4.35
Hpwt	2.12	2.21	2.96	2.56	2.35
Air	1.14	1.12	1.95	1.43	1.14
Space	2.98	3.05	3.08	3.07	3.05
Mpg	0.33	0.34	0.24	0.30	0.33
θ_{21}	–	0.60	−30.58	7.59	−4.42
θ_{22}	–	−3.61	−146.60	−9.64	−8.14
θ_{23}	–	–	31.39	–	–
θ_{24}	–	–	145.51	–	–

Figure 5: Comparison of Shapes BLP Model/Flexible Model with a Support of $\alpha_r \in [-0.2, 8.2]$



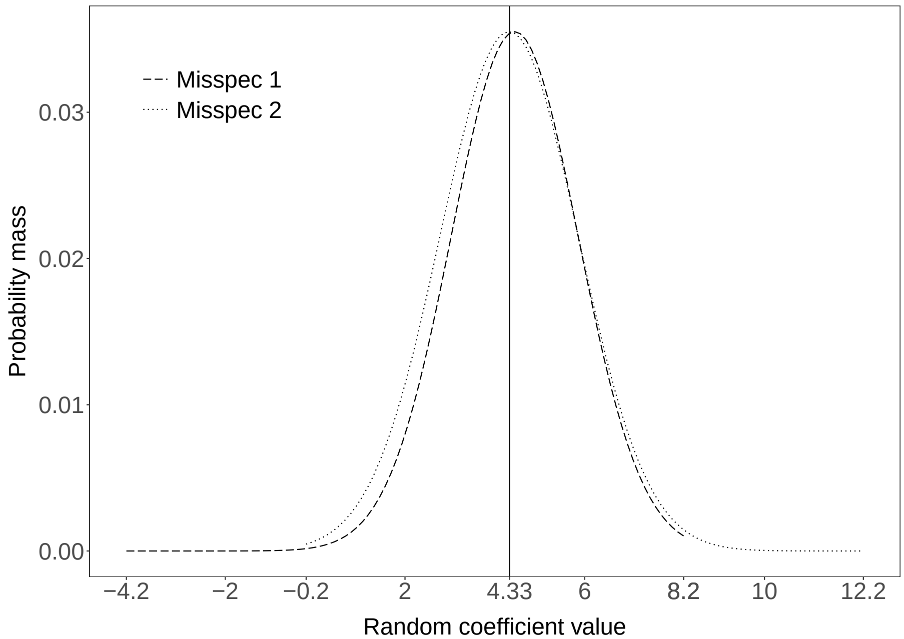
two specifications. Notably, the mean of the price coefficient is estimated to be closer to the mean of the BLP model with 3.64 compared to specification using a second order polynomial. Hpwt and Air are estimated to be comparably high with values 2.96 and 1.95, and Mpg relatively low with a value of 0.24. Again, the last four rows hold the estimated parameter values of the z variables, e.g., $\theta_{21} = -30.58$ or $\theta_{22} = -146.60$. Higher values are usually a good indicator for degenerate distributions or distributions that do not cover the full support as specified.

Figure 5 shows the backed out shapes of the distribution of the price coefficient. The solid line represents the normal distribution with a mean of 3.43 and standard deviation of 1.09 as estimated by the BLP model. The bell shape of the normal distribution is clearly visible. The dashed line shows the distribution calculated based on the flexible model with a second order polynomial as z variable. The shapes differ slightly in terms of mean and width, but the similarity is striking. The dotted line shows the distribution based on the fourth order polynomial. It shows a large and sharp increase in probability around its mean value of 3.64 and quickly falls off to zero at the surrounding area and mostly displays zero probabilities on its support. This is in line with the observation that the parameters of the z variables are estimated as large values. Based on this distribution, one might reject the hypothesis of consumer heterogeneity and conclude that a linear parameter is appropriate.

The robustness observed for the second order polynomial as z variable cannot be said about the fourth order polynomial. The results varied widely based on starting values and grid specification and often showed degenerate distributions. Often two or three large spikes can be identified holding the entire mass of the distribution with all other values having zero probability masses. In other cases, only the smallest and largest random coefficient value hold positive probability masses. Since the second order polynomial is robustly estimated but the fourth order polynomial is not, several interpretations are possible. According to one interpretation, one spike in probability is equivalent to homogeneous consumers, characterized by a linear parameter only. In this case, the nonlinear term and the assumption of consumer heterogeneity are rejected, and the model reverts to the simple version, which is inconsistent with the results from the second order polynomial. Another tempting interpretation thus takes the fragility as a sign that the underlying true distribution in fact can be approximated by a normal distribution but is not estimable due to a lack of data. The failure of the third and fourth order terms of the fourth order specification to converge to zero can then be seen as small sample bias being too strong to let those terms converge to zero and reveal the bell shape. This would be consistent with observations from the Monte Carlo study where $2,400$ observations (200 markets and 12 products) are not enough to consistently estimate the shape of the bimodal distribution. From this perspective, the BLP dataset is just limited in what is possible with $2,217$ entries. A third interpretation is that the underlying distribution might as well be bimodal or follow some skewed distribution. Then the backed out bell shape can be treated as an approximation of the mean and variance of the underlying true distribution. More data is needed to overcome small sample bias and reveal its true shape.

I now want to move on to a more in-depth discussion about the robustness of the estimation procedure. The results obtained for the second order polynomial appear to be robust once the grid is fixed at the specified values. Changing optimization starting values does not result in different parameter estimates. In this respect, the results of the second order polynomial seem to be quite robust. Also changing the grid range by some margin does not result in large changes in parameter estimates, which is reassuring as the specification of the grid is particularly arbitrary and a point of contention. Changing the grid by ± 1 does not result in large shifts of parameter estimates. Interestingly, specifying the grid too low or too high often resulted in zero probabilities of those regions misspecified, further strengthening the impression of validity. In some cases, a grid specified to be far off resulted in a failure to contract the market equilibrium when calculating the instrumental variables approximations. Using only BLP instruments in the second step of generalized method of moments does not largely produce different results.

Figure 6: Shapes of Distributions with Misspecified Grid for a Second Order Polynomial as z Variable



Misspecification 1 refers to a distributional shape backed out based on a support of $\alpha_r \in [-4.2, 8.2]$ and Misspecification 2 refers to a distributional shape backed out based on a support of $\alpha_r \in [-0.2, 12.2]$. The usual caution was taken when optimizing, e.g., the results are robust over different starting values.

The columns Misspec 1 and Misspec 2 of Table 8 refer to different specifications of the grid. I hereby assume that the correct specification of the grid is given by $\alpha_r \in [-0.2, 8.2]$. Misspec 1 refers to a misspecified grid with $\alpha_r \in [-4.2, 8.2]$ and Misspec 2 refers to a misspecified grid with $\alpha_r \in [-0.2, 12.2]$, each specification stretching the grid 4 units on one side. Note that the word misspecification is technically not correct but used for illustrative purposes, as the true grid capturing most or all random coefficients is unknown. Table 8 reveals that the estimated linear values and the expectation of the random coefficient are only marginally influenced by the misspecification, with all values in line with the other specifications. The mean of the price coefficient is unswayed by the grid manipulation, with the grid stretched on the upper side surprisingly resulting in a lower price coefficient of 4.35 than the downward stretched grid with a price coefficient of 4.43. This observation must be taken with care; an earlier version of the paper with a different specification of the estimation indicated that the price coefficient can very well be influenced by the grid specification.

Figure 6 shows the shape of the distributions based on the misspecified grid. Despite the misspecification, the correct shapes are clearly captured. Also, no matter in which direction the misspecification, the correct support of the random coefficient is matched closely, as can be seen in Figure 6 and comparing the shapes to Figure 10. Manipulation of the grid does not force the shapes on a different support. The linear value of the price coefficient backed out by the second order polynomial as z variable is highlighted by a vertical line at 4.33 as reference. Notice that the peaks match closely to the reference specification. Special attention should be given to the lowest and highest values of the random coefficients. At those values that are not in an area where the price coefficient is normally expected to reside, namely at negative values (more specific, $\alpha_r \in [-4.2, 0]$), the probability masses are actually estimated to be mostly zero. At the same time, at the values far to the right, e.g., at values beyond eight, the probability masses are also virtually zero, corresponding to the normal distribution from the BLP model; realization of random coefficient values of higher than eight is highly unlikely. This altogether presents a rather robust picture of the second order specification of the flexible estimation procedure.

This concludes the application section. To summarize, the flexible estimation procedure seems to be able to produce realistic results outside of the artificial environment of a Monte Carlo study. The estimates are comparable to the BLP model and are robust to changes of starting values and grid specification. The dataset does not provide enough information to estimate higher order specifications. This is not surprising as demonstrated in the Monte Carlo study.

8 Conclusion

The Berry, Levinsohn, and Pakes (1995, BLP) model is widely used to obtain parameter estimates of market forces in differentiated product markets. The results are often used as an input to evaluate economic activity in a structural model of demand and supply. Precise estimation of parameter estimates is therefore crucial to obtain realistic economic predictions. The present paper combines the BLP model and the logit mixed logit model of Train (2016) to estimate the distribution of consumer heterogeneity in a flexible and parsimonious way. A Monte Carlo study yields asymptotically normally distributed and consistent estimates of the structural parameters. With access to micro data, the approach allows for the estimation of highly flexible parametric distributions. The estimator further allows to introduce correlations between tastes, yielding more realistic demand patterns without substantially altering the procedure of estimation, making it highly relevant for practitioners. The BLP estimator is established to yield biased and inconsistent results when the underlying distributional shape is non-normally distributed. An application shows the estimator to perform well on a real world dataset and provides similar estimates as the BLP estimator with the option of specifying consumer heterogeneity as a function of a polynomial, step function or spline, resulting in a flexible estimation procedure.

A Appendix

A.1 Rewriting the Contraction

Following the notation of Brunner et al. (2017) (see also Appendix of Nevo (2000a)), the contraction mapping is given by:

$$w_j^{iter+1} = w_j^{iter} \cdot \frac{S_j}{s_j(\theta, \nu)} = w_j^{iter} \cdot \frac{S_j}{R^{-1} \sum\limits_r \frac{w_j^{iter} v_{rj}}{1+\sum\limits_k w_k^{iter} v_{rk}}},$$

with $w \equiv e^{x_j\beta+\xi_j} = e^{\delta_j}$ and $v_{rj} \equiv e^{\mu_{rj}}$.

The contraction is rewritten similar to the Appendix of Brunner et al. (2017) to speed up estimation times. Pull w_j^{iter} into the denominator:

$$w_j^{iter+1} = \frac{S_j}{R^{-1} \sum\limits_r \frac{v_{rj}}{1+\sum\limits_k w_k^{iter} v_{rk}}}$$

and finally applying Train (2016) to the contraction mapping yields the modified contraction mapping:

$$w_j^{iter+1} = \frac{S_j}{\sum\limits_{r\in S} \frac{v_{rj}}{1+\sum\limits_k w_k^{iter} v_{rk}} \cdot W_r}. \tag{20}$$

A.2 Implementation Details: Own- and Cross-Derivatives of the Market Shares

Based on the new formula for the aggregate market shares, entries of Ω cannot be calculated as in the BLP model, but have to be modified. Differentiating s_j in Equation (6) with respect to price and rearranging, the entries of a nonlinear price coefficient become

$$\frac{\partial s_j}{\partial p_j} = -\sum_{r\in S} s_{rj}(1 - s_{rj})(\nu_{rp} + D_r\zeta d_j)W_r \tag{21a}$$

$$\frac{\partial s_j}{\partial p_q} = \sum_{r\in S} s_{rj} s_{rq}(\nu_{rp} + D_r\zeta d_j)W_r \tag{21b}$$

For a model without micro data, set ζ equal to zero. The formulae can take on different signs, depending on how the prices are specified. I specify the price vector to be a vector of negative prices (i.e., the original price vector p times -1) so that the grid is specified to hold positive values.

A.3 Implementation Details: Derivation of the Gradient

Train (2016) stresses the ease of deriving the gradient to speed up estimation times. I now proceed to derive the gradient for the flexible model as discussed in Section 3.3 and 4. The gradient of the objective in Equation (12) is given by

$$\nabla G_1(\theta) = 2\left(\frac{\partial \xi}{\partial \theta}\right)' ZWZ'\xi. \tag{22}$$

Rewriting $\delta = x\beta + \xi$ as $\xi = \delta - x\beta$, the inner derivative can be calculated similarly as described by Nevo (2000a) by totally differentiating the matching condition and subtracting the linear part:

$$\frac{d\xi}{d\theta} = \frac{\partial \delta}{\partial \theta} - \frac{\partial x\beta}{\partial \theta}. \tag{23}$$

The components are given by

$$\frac{\partial \delta}{\partial \theta} = -\left[\frac{\partial s(\delta;\nu,\theta)}{\partial \delta}\right]^{-1}\left[\frac{\partial s(\delta;\nu,\theta)}{\partial \theta}\right] \tag{24}$$

and by taking into account Equation (10) and the β value resulting from a two stage least squares procedure and the matrix formula associated with it, the derivative obtains

$$\frac{\partial x\beta}{\partial \theta} = -x\underbrace{(x'ZWZ'x)^{-1}x'ZWZ'}_{\tau}\underbrace{\left[\frac{\partial s(\delta;\nu,\theta)}{\partial \delta}\right]^{-1}\left[\frac{\partial s(\delta;\nu,\theta)}{\partial \theta}\right]}_{\frac{\partial \delta}{\partial \theta}} = -x\tau\frac{\partial \delta}{\partial \theta} \tag{25}$$

with market shares $s(\delta,\nu;\theta)$ given by Equation (6) and τ referring to the two stage least squares projection matrix and x (cf. Hansen (2021), eqs. 12.30 and 12.31, pg. 345).

Nevo calculates the derivative $\partial\delta/\partial\theta$ by linearizing the system of nonlinear equations, taking the derivative and transforming the system back to the nonlinear system (cf. Simon and Blume (1994) pg. 355). The results are an approximation about a specified point, which should be kept in mind. The gradient is only accurate close to the point evaluated.

Deriving the first term of the right hand side of (24) is straightforward, i.e., the market shares differentiated with respect to the linear part of utility δ, having dimensions $J \times J$, j and q being two products:

$$\frac{\partial s_j(\delta,\nu;\theta)}{\partial \delta_j} = \sum_{r\in S} s_{rj}(1-s_{rj})\cdot W_r \tag{26a}$$

$$\frac{\partial s_j(\delta,\nu;\theta)}{\partial \delta_q} = -\sum_{r\in S} s_{rj}s_{rq}\cdot W_r \tag{26b}$$

Lastly, I calculate the market shares differentiated with respect to the optimization parameters θ. From now on one has to specify a z variable which I assume to be a polynomial of order n. After forming the derivative and simplification, the expression for the market share derivative with respect to nonlinear optimization parameter θ concerning characteristic x_k is given by

$$\frac{\partial s_j(\delta,\nu;\theta)}{\partial\theta^k_n} = \sum_{r\in S} s_{rj}\left(\tilde{\nu}^n_{rk} - \sum_{s\in S} W_s\tilde{\nu}^n_{sk}\right)\cdot W_r, \tag{27}$$

with $\tilde{\nu}^n_{rk}$ referencing the scaled draws as described in Section 3.2 and n indicating which order of the polynomial the derivative refers to. Expression (27) holds for all random coefficients, also for a random coefficient on the constant. Replacing k by p yields the derivative with respect to the price coefficient.

A.4 Implementation Details: Exemplary Derivation of Market Share Derivatives with Respect to Optimization Parameters

Derivatives are calculated exemplary for a logit distribution based on polynomials of order n. Structuring the z variables differently (e.g., step-functions or splines) requires all derivatives to be re-derived. With normalized random coefficients denoted as $\tilde{\nu}_r$, the aggregate market share is given by

$$s_j = \sum_{r\in S}\left(\frac{e^{\delta_j+\mu_{rj}}}{1+\sum\limits_{j\in J} e^{\delta_j+\mu_{rj}}}\right)\cdot\left(\frac{e^{\theta' z(\tilde{\nu}_r)}}{\sum\limits_{s\in S} e^{\theta' z(\tilde{\nu}_s)}}\right)$$

and the logit distribution polynomial of order n in the more elaborate form

$$W_{rk} = W(\tilde{\nu}_k|\theta^k_2) = \frac{e^{\theta^k_{21}\tilde{\nu}_{rk}+\theta^k_{22}\tilde{\nu}^2_{rk}+\ldots+\theta^k_{2n}\tilde{\nu}^n_{rk}}}{\sum\limits_{s\in S} e^{\theta^k_{21}\tilde{\nu}_{sk}+\theta^k_{22}\tilde{\nu}^2_{sk}+\ldots+\theta^k_{2n}\tilde{\nu}^n_{sk}}}.$$

Concerning notation: θ^k_{2n} indicates the nth logit parameter of characteristic k associated with heterogeneity. The subscript '2' indicates a nonlinear optimization parameter. The value $\tilde{\nu}^n_{rk}$ is the nth power of the transformed heterogeneity draw (or taste) of consumer type r towards characteristic k. Then the derivative can be calculated as

$$\frac{\partial s_j}{\partial\theta^k_{2n}} = \sum_{r\in S}\frac{e^{\delta_j+\mu_{rj}}L - e^{\delta_j+\mu_{rj}}\sum\limits_{j\in J} e^{\delta_j+\mu_{rj}}}{L^2} W_r + s_{rj}W_r\left[\tilde{\nu}^n_{rk} - \sum_{s\in S} W_s\tilde{\nu}^n_{sk}\right]$$

with the outer sum covering the whole term and $L = 1 + \sum_{j \in J} e^{\delta_j + \mu_{rj}}$. The term simplifies to:

$$\frac{\partial s_j}{\partial \theta_{2n}^k} = \sum_{r \in S} s_{rj} \left[\tilde{\nu}_{rk}^n - \sum_{s \in S} W_s \tilde{\nu}_{sk}^n \right] W_r. \tag{28}$$

The market share derivatives with respect to the parameters for the price coefficient are derived similarly and calculated as

$$\frac{\partial s_j}{\partial \theta_{2n}^\alpha} = \sum_{r \in S} s_{rj} \left[\tilde{\nu}_{rp}^n - \sum_{s \in S} W_s \tilde{\nu}_{sp}^n \right] W_r. \tag{29}$$

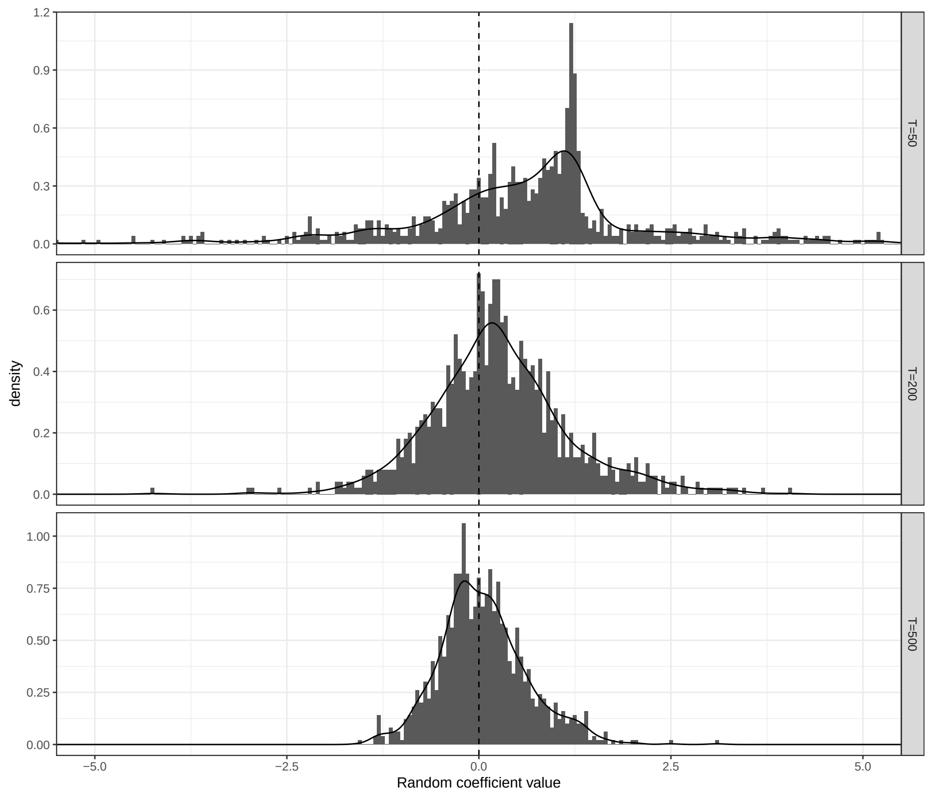
A.5 Convergence of Parameter Values, Consumer Heterogeneity on Price

The following Table 9 and Figures 7, 8, 9 and 10 show convergence of the parameter estimates of the Monte Carlo study in Section 4, with the price coefficient being heterogeneous across consumers, thus denoted by $\tilde{p}$, and the constant and the taste on x being homogeneous across consumers. This changes the model setting only slightly, with $X_{jt} = (\tilde{p}_{jt}, 1, x_{jt})$. The demand side marginal utility parameters of X_{jt} are then given by the column vector $\beta_r = (\alpha_r, \theta_{11}, \theta_{12}) = (\alpha_r, \beta_0, \beta_1) = (\nu_r, 4, 2)$. The grid is set to $\nu_r \in [2, 9]$ and the z variables are set to $\theta = (\theta_{21}, \theta_{22}) = (0, -4)$, modeled by a second order polynomial. The modification is necessary to allow equilibrium convergence when generating the data, which must accommodate a specific combination of parameter values. I modify marginal costs, with two additional cost shifters, c_1 and c_2, excluded from demand. The cost shifters are used as instruments along with the approximately optimal instruments. This is necessary to show clear convergence for 200 markets or $2,400$ observations, as estimation of the distribution for the price coefficient turns out to be more challenging than estimating the distribution on the parameter on characteristic x. The cost shifters (c_s) are generated as uniformly distributed variables or $c_s \sim$ Uniform(0, 0.1). For clarity, utility is formally given by $u_{rjt} = X_{jt}\beta_r + \xi_{jt} + \epsilon_{rjt} = \tilde{p}_{jt}\nu_r + 4 + 2x_{jt} + \xi_{jt} + \epsilon_{rjt}$. Marginal costs are formally given by $mc_{jt} = 2.5 + 0.2\tilde{x}_{jt} + c_1 + c_2 + 0.2\omega_{jt}$. Qualitatively, the results mirror the results of Section 4.

Table 9: Parameter Summary Statistics for a Distribution based on a Second Order Polynomial as z Variable, Heterogeneous Price Coefficient

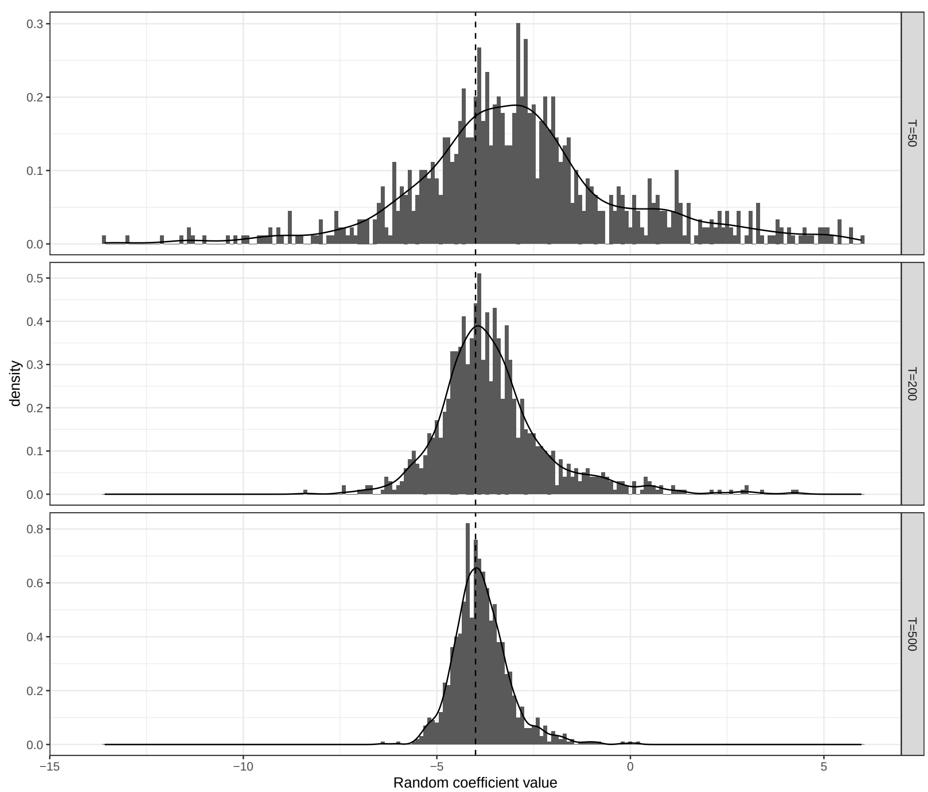
Markets	θ_0	$\bar{\hat{\theta}}$	$\tilde{\hat{\theta}}$	$\hat{\theta}_{[0.025]}$	$\hat{\theta}_{[0.975]}$	St. Err.
θ_{21}						
50	0	0.550	0.550	−3.627	4.321	1.900
200	0	0.269	0.269	−1.406	2.323	0.920
500	0	0.060	0.060	−0.939	1.354	0.576
θ_{22}						
50	−4	34.551	38.551	−8.830	813.005	169.198
200	−4	−3.510	0.490	−5.812	0.385	1.602
500	−4	−3.841	0.159	−5.120	−2.004	0.739
$\theta_{11} = \alpha$						
50	4	3.746	−0.254	0.513	9.751	2.419
200	4	3.994	−0.006	1.844	6.846	1.325
500	4	3.952	−0.048	2.494	5.630	0.812
$\theta_{12} = \beta_0$						
50	2	1.979	−0.021	1.708	2.437	0.190
200	2	1.999	−0.001	1.817	2.223	0.102
500	2	1.996	−0.004	1.883	2.124	0.061

Figure 7: Monte Carlo Convergence of Parameter θ_{21} for a Second Order Polynomial as z Variable and Instruments at the Expected Value of the Structural Shocks



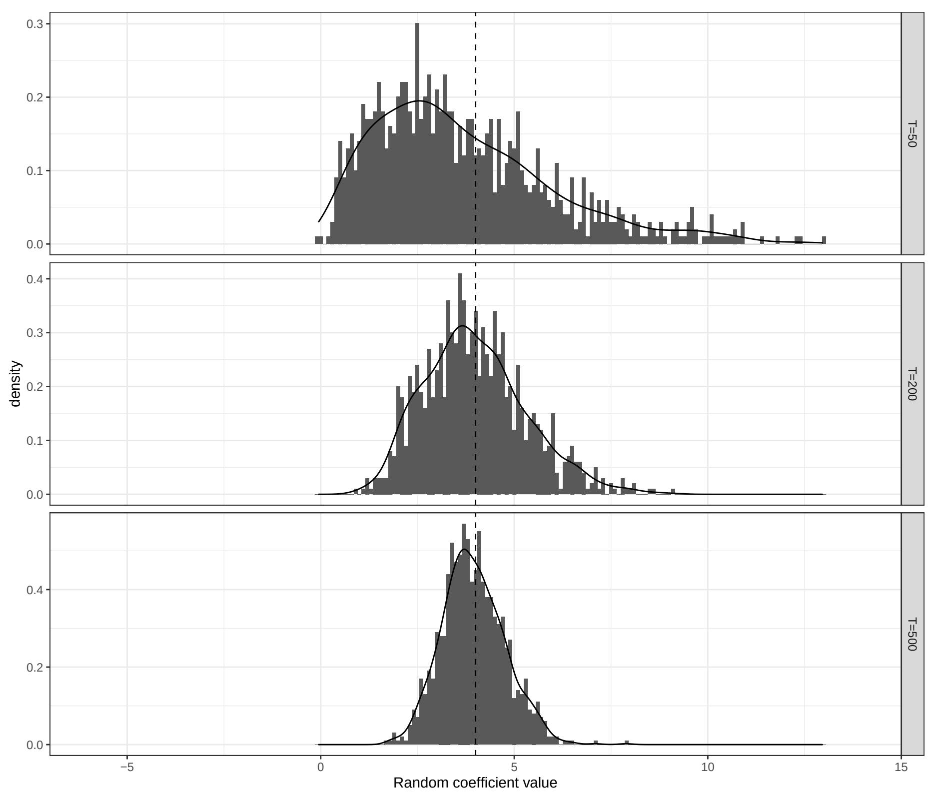
This figure is created with consumers having heterogeneous tastes concerning the price coefficient, $\tilde{p}$. The intercept and the characteristic, x_j, enter the model linearly.

Figure 8: Monte Carlo Convergence of Parameter θ_{22} for a Second Order Polynomial and Instruments at the Expected Value of the Structural Shocks



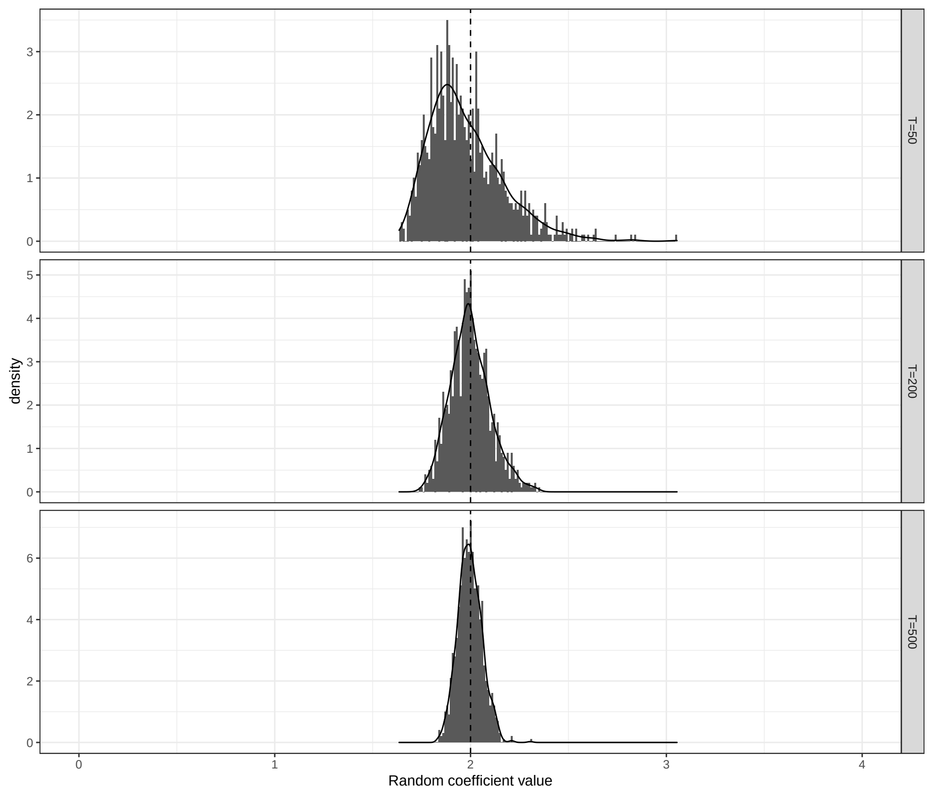
This figure is created with consumers having heterogeneous tastes concerning the price coefficient, $\tilde{p}$. The intercept and the characteristic, x_j, enter the model linearly.

Figure 9: Monte Carlo Convergence of Parameter $\theta_{11} = \beta_0$ for a Second Order Polynomial as z Variable and Instruments at the Expected Value of the Structural Shocks



This figure is created with consumers having heterogeneous tastes concerning the price coefficient, $\tilde{p}$. The intercept and the characteristic, x_j, enter the model linearly.

Figure 10: Monte Carlo Convergence of Parameter $\theta_{12} = \beta_1$ for a Second Order Polynomial and Instruments at the Expected Value of the Structural Shocks



This figure is created with consumers having heterogeneous tastes concerning the price coefficient, $\tilde{p}$. The intercept and the characteristic, x_j, enter the model linearly.

A.6 Monte Carlo Study: Computational Details

All programming is done using the programming language Python, mainly relying on the library NumPy for data generation and data manipulation and SciPy for minimization of the objective. The SciPy function minimize is used along with the BFGS algorithm and an analytical gradient. I scale the objective given by Equation (12) by dividing through the squared number of products in the sample. The tolerance of the optimizer gradient norm is set to $1e-12$. Objective minimizing parameter values are accepted as an estimate only when the optimizer converged; otherwise, the current draws are dropped and new random draws generated. The inner tolerance of the contraction mapping is set to $1e-16$. It is important to keep the inner tolerance tight to prevent inaccuracies from spreading into the system, as discussed by several authors (cf. Dubé, Fox, and Su (2012), Brunner et al. (2017), Conlon and Gortmaker (2020)). The gradient is also affected by inaccurate contraction, aggravating optimization. Also, the maximum number of iterations before premature termination of the contraction should not be set to low and checked. I set it to 2500, which is far in abundance of what is actually needed. The tolerance when initially generating the market equilibrium in Nash Bertrand fashion by looping over the first order conditions is set to 1e-12. The starting values are selected randomly for each optimization based on a standard normal distribution.

A failure of optimization can occur when particularly large values are entered into the exponentials of the logit probability formula, leading to frequent overflow. A solution is presented by Conlon and Gortmaker (2020) who use the following log-sum-exp expression: $\text{lse}(x) = \log \sum_k e^{x_k} = a + \log \sum_k e^{x_k - a}$ for a = $\max\{0, \max_k x_k\}$. I use this formula to rewrite the logit probabilities of the weights in Equation (6), $W(\nu_r|\theta)$, as $\log(e^{x-a}/\sum_k e^{x_k-a}) = x - a - \log \sum_k e^{x_k-a}$. Exponentiating the last term feeds back the logit probabilities. After implementing this expression, problems related to overflow are greatly reduced. Note that I do not apply this formula to the consumer-specific choice probabilities, $\text{Prob}_{rj}(\delta_j; \theta, \nu_r)$.

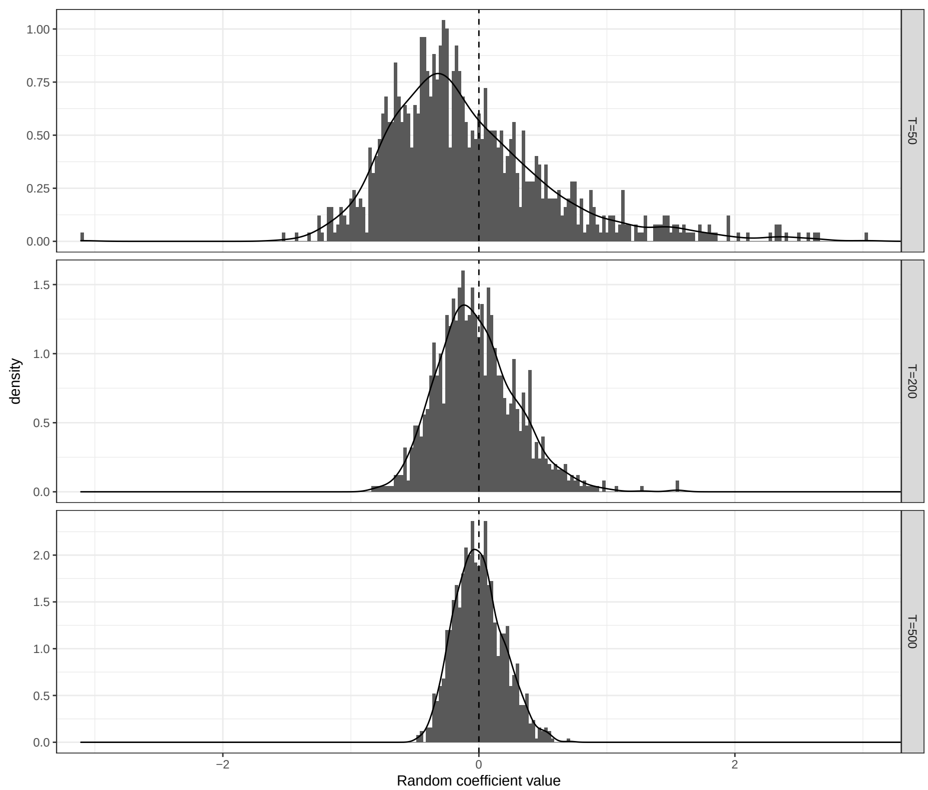
One more point to stress is that in some rare cases the estimated values fall far off the average value of all other parameters, more specific values being $1,000$ times as large as the average of all other values. This happened 17 times out of 4000 parameter values in the case of BLP instruments (of which 16 times when estimating 25 or 50 markets) and only 4 times out of 4000 parameter values for the approximately optimal instruments.

There are several reasons why some values might be far off. It might be the case because the optimizer has trouble finding the true minimum. Another reason might be the inherent randomness of the data, sometimes leading to unusual results. To better evaluate the results, I exclude those extreme values from the discussion of the results and when calculating the summary statistics. Note that I only exclude 21 out of 8000 estimated values. A set of estimated values is dropped if the positive of a value is larger than $1,000$.

A.7 Convergence of Parameter Values, Consumer Heterogeneity on the Characteristic

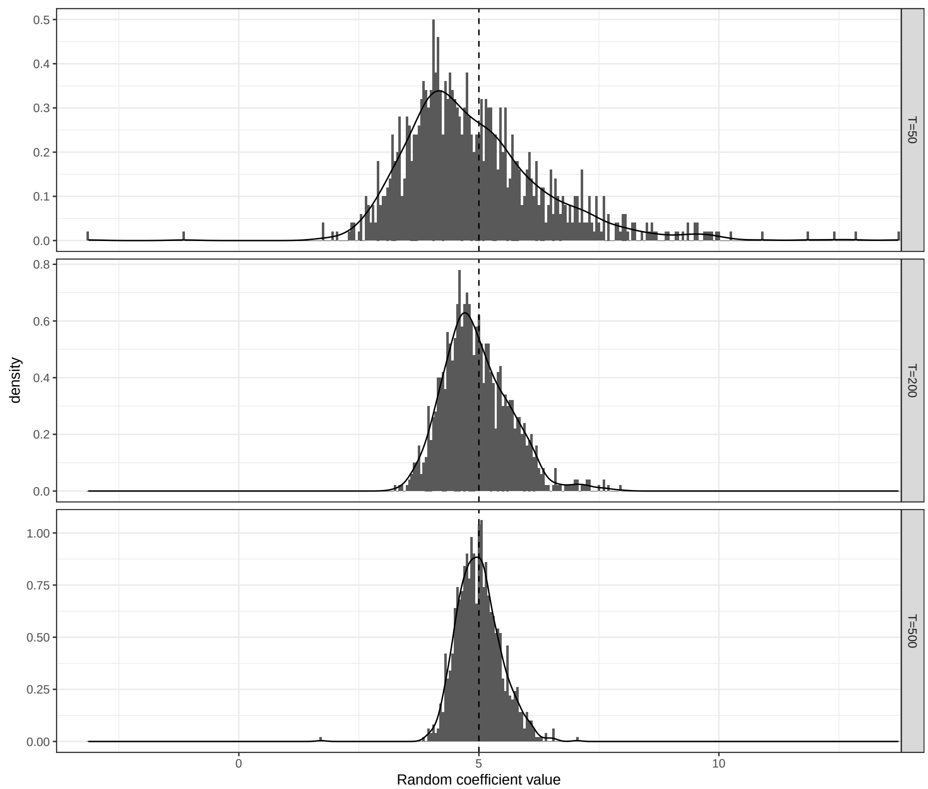
This Appendix shows the convergence of the other parameters of the Monte Carlo study conducted in Section 4. Figure 11 shows convergence of the parameter of the first z variable, responsible for shifting the distribution to the sides (similar to a parabola). Figure 12 shows convergence of the price coefficient. Figure 13 shows convergence of the intercept. For all parameters, the bias is reduced with additional markets entering the sample and all parameters appear to be consistently estimated given an appropriate sample size.

Figure 11: Monte Carlo Convergence of Parameter θ_{21} for a Second Order Polynomial as z Variable and Instruments at the Expected Value of the Structural Shocks



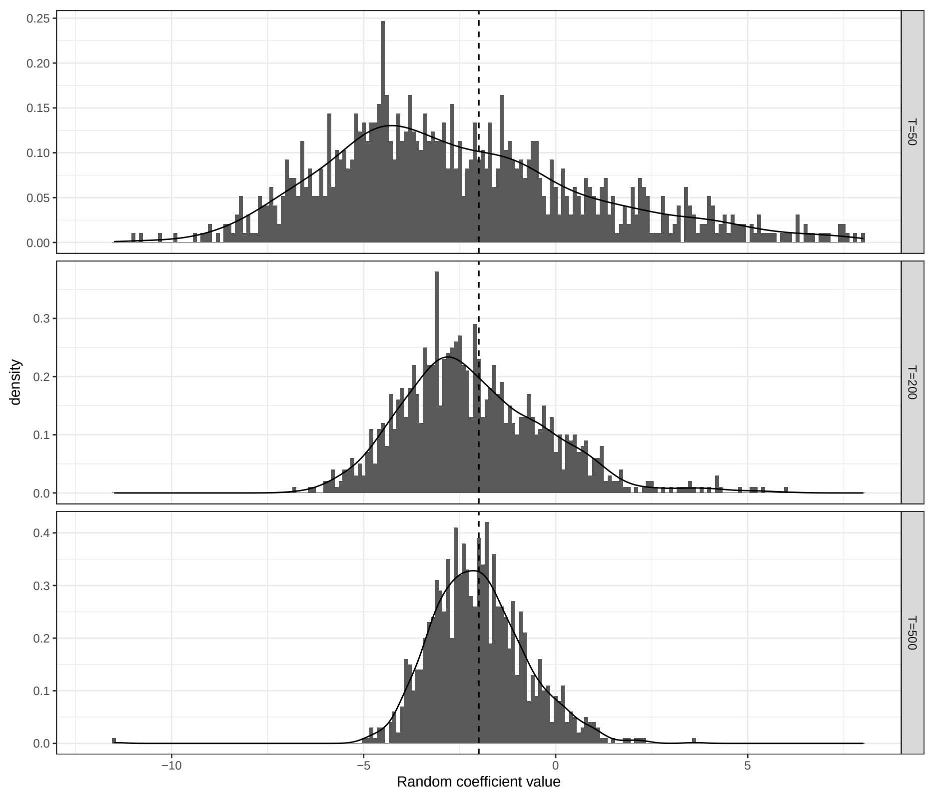
This figure is created with consumers having heterogeneous tastes concerning the characteristic, $\tilde{x}$. The intercept and the price, p, enter the model linearly.

Figure 12: Monte Carlo Convergence of the Price Coefficient $\theta_{11} = \alpha$ for a Second Order Polynomial as z Variable and Instruments at the Expected Value of the Structural Shocks



This figure is created with consumers having heterogeneous tastes concerning the characteristic, $\tilde{x}$. The intercept and the price, p, enter the model linearly.

Figure 13: Monte Carlo Convergence of Parameter $\theta_{12} = \beta_0$ for a Second Order Polynomial and Instruments at the Expected Value of the Structural Shocks



This figure is created with consumers having heterogeneous tastes concerning the characteristic, $\tilde{x}$. The intercept and the price, p, enter the model linearly.

A.8 Implementation Details: Derivation of the Gradient with Micro Data

The market shares are now given by

$$s_j = \sum_{r \in S} W_r \cdot s_{rj} \tag{30}$$

and

$$s_{rj} = \frac{\exp(x_j\beta + \xi_j + \tilde{x}_j(\nu_r + \mathrm{D}_r\zeta d_j))}{1 + \sum_j \exp(x_j\beta + \xi_j + \tilde{x}_j(\nu_r + \mathrm{D}_r\zeta d_j))}. \tag{31}$$

The derivatives of the micro moments are calculated similarly. The inner derivative of the linear part of utility has to be taken into account. Since the linear part, δ, is the outcome of the contraction mapping and depends on the parameter vector, thus $\delta(\theta)$, the derivative is given by:

$$\frac{\partial s_j}{\partial \theta_n^k} = \sum_{r \in S} W_r \left[\underbrace{\tilde{\nu}_r^n - \sum_{i \in S} W_i \tilde{\nu}_i^n}_{\kappa_1} + \underbrace{\frac{\partial \delta_j}{\partial \theta_n^k} - \sum_{j=1}^{J} s_{ij} \frac{\partial \delta_j}{\partial \theta_n^k}}_{\kappa_2} \right] s_{rj}. \tag{32}$$

The inner derivative of W_r is given by κ_1 and the derivative of the linear term given by κ_2. I calculate κ_2 in an initial step according to Equation (27). Then the gradient of the micro moments are

$$\frac{\partial \mathrm{G}_2}{\partial \theta} + \frac{\partial \mathrm{G}_3}{\partial \theta} = 2\mathrm{G}_{22} \frac{\partial\, \mathbb{E}[\mathbf{D}|\mathbf{j}]}{\partial \theta} / \sigma^2_{\hat{D}|j} + 2\mathrm{G}_{33} \frac{\partial\, \mathbb{E}[\mathbf{D}|\mathbf{type{=}o}]}{\partial \theta} / \sigma^2_{\hat{D}|o} \tag{33}$$

with $\mathbb{E}[\mathrm{D}|j]$ and $\mathbb{E}[\mathrm{D}|\text{type=o}]$ defined in Section 5.1 and σ^2 referring to the moment variance. Finally, the derivatives of the expected demographics are

$$\frac{\partial\, \mathbb{E}[\mathrm{D}|j]}{\partial \theta} = \sum_{r \in S} W_r(\kappa_1 + \kappa_2) s_{rj|P} \cdot \mathrm{D}_r \tag{34}$$

$$\frac{\partial\, \mathbb{E}[\mathrm{D}|\text{type} = \text{o}]}{\partial \theta} = \sum_{j \in o} \sum_{r \in S} W_r(\kappa_1 + \kappa_2) s_{rj|P} \cdot \mathrm{D}_r \tag{35}$$

A.9 Micro Data: Implementation Details

This Appendix spells out the modeling details of Section 5.2. In order to study the behavior of the estimator with micro data, I proceed conceptually similar to Section 4. Consumer utility depends on the endogenous price, a constant, a randomly generated characteristic, a randomly generated demand shock correlated with a supply side cost shock and on a consumer, product specific arbitrary shock term. Consumer utility additionally depends on consumer demographics. The characteristic is given by $\tilde{x}_{jt}$, with $\tilde{x}_{jt} \sim \text{Uniform}(0, 3)$ and the demand and cost shocks are multivariate standard normal with covariance 0.7. Products belong to two different segments; they are either in a first segment if $\tilde{x}_{jt} < \underline{\tilde{x}}_{jk} = 1.5$ and otherwise belong to the second segment. Firms compete in prices and consumers display homogeneous preferences concerning intercept and price. The distribution of consumer tastes on the coefficient of $\tilde{x}_{jt}$ is modeled as a fourth order polynomial as z variable with parameter values $\theta = (\theta_{21}, \theta_{22}, \theta_{23}, \theta_{24}) = (-1.5, 5, 3, -6)$. This yields a bimodal distribution with two pronounced peaks and a deep trough. The parameters of the distribution are then estimated by generalized method of moments. The weight function contained in Equation (6) is now a joint distribution of taste and the demographic variable. Micro data is introduced which follows a simple distribution modeled by the logit formula as in Train (2016) with dummy variables as z variables yielding a step function. The distribution has a discrete support of $\mathrm{D}_r \in \{2, 3, 4, 5\}$ with probabilities $p = (0.3, 0.35, 0.2, 0.15)$, respectively. This is reminiscent of Petrin (2002) and can be thought of as family size. During estimation, only demographic deviations from its expected value are considered, i.e., the mean is set to zero. As far as the logit formula modeling the probability distribution is concerned, the probability that an individual is associated with family size i is given by $p_i = \exp(\phi_i) / \sum_{k=1}^{n} \exp(\phi_k)$, which is a step function. Since I assume the probabilities to be known, it is possible to calculate the corresponding ϕ's according to $\phi_i = \ln\left(p_i / (1 - \sum_{i=1}^{n-1} p_i)\right)$ and the last element is normalized to $\phi_n = 0$. See Appendix A.11 for a derivation and Train (2016) for more details.

The specifications 'No Micro' and 'Micro' differ in their exact utility functions. The modification is necessary to allow equilibrium convergence when generating the data, which must accommodate a specific combination of parameter values. For clarity, consumer utility for specification 'No Micro' is given by $u_{rjt} = -6p_{jt} - 4 + \tilde{x}_{jt}(\nu_r + D_r d_j) + \xi_{jt} + \epsilon_{rjt}$. Consumer utility for specification 'Micro' is given by $u_{rjt} = -5p_{jt} + 9 + \tilde{x}_{jt}(\nu_r + D_r d_j) + \xi_{jt} + \epsilon_{rjt}$. The utility shifter ζ is set to unity, and d_j is a segment dummy as described above. Marginal costs are assumed to be constant. Costs depend on a constant and $\tilde{x}_{jt}$ with cost side parameters $\gamma = (2.5, 0.2)$, respectively. Marginal costs are determined by a cost shifter $c^s \sim \text{Uniform}(0, 0.1)$ (only in the case of 'No Micro') and are further determined by a supply shock ω, which is weakened by a factor of $\gamma_c = 0.2$. Marginal cost is given by $mc_{jt} = 2.5 + 0.2x_{jt} + c^s_{jt} + 0.2\omega_{jt}$. The grid of the 'No Micro' specification is set to $\nu_r \in [1, 8]$. The grid of the 'Micro' specification is set to $\nu_r \in [0, 2]$.

The estimator is affected by several sources of variance as discussed in BLP and Berry, Levinsohn, and Pakes (2004). A first source of variance is introduced by the randomness of the product characteristics. A second source of variance stems from the inaccuracy of observed market shares. A third source of variance is introduced by integration of the market shares. Finally, a fourth source of variance follows from the sampled micro data.

I keep the variances at a minimum with the following measures. I use low tolerances when generating the equilibrium to obtain highly precise market shares. Integration is executed with high precision and micro data derived values are assigned with a low variance of $1e - 14$, i.e., the micro moments are basically assumed to be exact. The low variance results in high weights of the the micro moments in the objective function. This is done intentionally to proxy highly accurate micro data. Covariances between the moments are set to zero for simplicity.

A.10 Effects of Distributional Misspecification: Computational Details

This Appendix spells out the modeling details of Section 6. Firms compete Nash in prices and marginal costs are assumed to be constant. Costs are determined by a constant and x_{jt} with cost side parameters $\gamma = (2.5, 0.2)$, respectively. Marginal costs are determined by two cost shifters $c^s \sim \text{Uniform}(0, 0.1)$ and further by a supply shock ω, which is weakened by a factor of $\gamma_c = 0.2$. Marginal costs are formally given by $mc_{jt} = 2.5 + 0.2x_{jt} + c^s_{jt} + 0.2\omega_{jt}$. Utility of a consumer of type r and product j in market t is formally given by $u_{rjt} = -\nu_{rp}\tilde{p}_{jt} + 4 + 2x_{jt} + \xi_{jt} + \epsilon_{rjt}$. Consumer heterogeneity is modeled as follows. The grid is set to $\nu_{rp} \in [2, 9]$. The parameters of the second order polynomial as z variable are $\theta = (\theta_{21}, \theta_{22}) = (0, -1)$ and the parameters of the fourth order polynomial as z variable are $\theta = (\theta_{21}, \theta_{22}, \theta_{23}, \theta_{24}) = (0, 6, 0, -10)$. Tolerances are tightened sufficiently, and integration is done at high accuracy using modified latin hypercube sampling (Hess, Train, and Polak, 2006) and unique draws for each market (cf. Freyberger, 2015).

A.11 Derivation of Step Function Logit Probabilities for Demographics

The step function discussed in Train (2016) is given by

$$W(\nu_i) = \frac{\exp[\sum_g \phi_g I(\nu_i \in H_g)]}{\sum_{s\in S} \exp[\sum_g \phi_g I(\nu_s \in H_g)]},$$

with ν_i representing a random coefficient value in the finite support of the random coefficients, S, and H_g representing a subset of S. $I(\cdot)$ is the indicator function with $I = 1$ if $\nu_i \in H_g$ and $I = 0$ otherwise. This approach is used to derive a simple step function for consumer demographics. Let $p_i \in (0,1)$ for $i = 1, ..., n$ and $\sum_{i=1}^{n} p_i = 1$ and $1 - \sum_{i=1}^{n-1} p_i = L$ and $D_i \in H_i$ for $i = 1, ..., n$. Then the probability that any interval H_i is associated with the consumer value D_i can be derived from the associated coefficients $\phi = \{\phi_1, \phi_2, ..., \phi_{n-1}, 0\}$ of the indicator function, i.e., the coefficients are n scalars defining the probabilities of respective intervals. If $D_i \notin H_i$, the corresponding value is set to 0 by the indicator function; intervals can be overlapping (for more details, cf. Train (2016)). The coefficients can be derived as follows. Let $p_i = \exp\phi_i / \sum_{i=1}^{n} \exp\phi_i$ and after swapping sides $\sum_{i=1}^{n} \exp\phi_i = \exp\phi_i/p_i = \exp\phi_n/p_n = \exp\phi_n/L$. Then, by taking the natural logarithm of $\exp\phi_i/p_i = \exp\phi_n/L$ and setting $\phi_n = 0$ yields $\phi_i = \ln(p_i/L)$ for $i = 1, ..., n-1$ and $\phi_n = 0$.

A.12 Effects of Distributional Misspecification: Asymmetric Distribution

This Appendix repeats the estimation presented in Section 6 with a differently shaped distribution. The procedure is unchanged with the minor modification that the vector to estimate is now given by $\theta = (\theta_{21}, \theta_{22}, \theta_{23}, \theta_{24}) = (-2, -0.5, 3, -5)$, yielding a strongly skewed distribution. The distribution is not bimodal. Table 10 and Table 11 report the results.

From the Tables, two things can be seen. Firstly, Table 10 shows the estimates to be more closely aligned to the true values as compared to the case with a symmetric bimodal distribution in Section 6, e.g., the standard deviation of the price coefficient is estimated to be 1.28 on average with a true value of 1.20 and the mean price coefficient is estimated to be 6.78 on average with a true value of 6.98.

Secondly, Table 11 shows an increased bias in estimated merger price changes and an increased bias in the welfare estimate (e.g., a bias in the welfare estimate of 23.12% in Table 11 compared to a bias of 17.14% in Table 7), with only minor changes to the bias of elasticities and the Lerner index. This indicates that with a skewed distribution,

Table 10: BLP Estimates When Truly Underlying Distribution is Skewed

Variable	BLP	St. Err.	True
Standard deviation	1.28	(0.08)	1.20
Price coefficient (α)	6.78	(0.41)	6.98
Intercept (β_0)	3.95	(0.90)	4.00
Characteristic (β_1)	2.00	(0.07)	2.00

Estimated with 1000 Monte Carlo repetitions and 1000 markets in a concentrated industry (3 firms with 2 products each).

the variance of the parameter estimates is larger as compared to the bimodal case, which is captured by the empirical standard error of the estimates (in parenthesis). A standard error of 0.04 in Table 6 compares to a standard error of 0.08 in Table 10 for the standard deviation, which is a doubling of the standard error. Whereas the skewed shape captures the true value better on average, frequent deviations from the true distributional shape increase the root mean square percentage error of the estimates by overshooting the true price differentials in both directions, increasing the overall bias of the structural estimates.

Table 11: Aggregate Statistics Bias When Estimated By BLP and Truly Underlying Distribution is Skewed

Statistic	RMSPE(x_0,$\hat{x}$)
η_{jj}	7.27
η_{jj}^a	7.26
η_{jk}	16.58
η_{jk}^a	12.97
Δp(merging)	21.47
Δp(fringe)	46.23
Lerner	7.53
Lerner^a	7.77
Welfare	23.12

Aggregate statistics using estimates from Table 10. I report root mean square percentage error (RMSPE) between estimated and true values. All measures are weighted by market share. η_{jk} is the percentage market share response of product j due to a 1% increase in price of product k. The superscript 'a' indicates after merger results. Aggregation is done by weighting with respective market shares.

References

Armstrong, Timothy B. (2016). “Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply”. In: *Econometrica* 84.5, pp. 1961–1980.

Berry, Steven T. (1994). “Estimating Discrete-Choice Models of Product Differentiation”. In: *The RAND Journal of Economics* 25.2, pp. 242–262.

Berry, Steven T. and Philip A. Haile (2022). “Nonparametric Identification of Differentiated Products Demand Using Micro Data”. In: *NBER Working Paper 27704*.

Berry, Steven, Michael Carnall, and Pablo T. Spiller (May 1996). *Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity*. Tech. rep. 5561.

Berry, Steven and Panle Jia (2010). “Tracing the Woes: An Empirical Analysis of the Airline Industry”. In: *American Economic Journal: Microeconomics* 2.3, pp. 1–43.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995). “Automobile Prices in Market Equilibrium”. In: *Econometrica* 63.4, pp. 841–90.

— (1999). “Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy”. In: *American Economic Review* 89.3, pp. 400–430.

— (Feb. 2004). “Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market”. In: *Journal of Political Economy* 112.1, pp. 68–105.

Berry, Steven, Oliver B. Linton, and Ariel Pakes (2004). “Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems”. In: *The Review of Economic Studies* 71.3, pp. 613–654.

Bourreau, Marc, Yutec Sun, and Frank Verboven (Nov. 2021). “Market Entry, Fighting Brands, and Tacit Collusion: Evidence from the French Mobile Telecommunications Market”. In: *American Economic Review* 111.11, pp. 3459–3499.

Brunner, Daniel, Florian Heiss, Andre Romahn, and Constantin Weiser (2017). *Reliable Estimation of Random Coefficient Logit Demand Models*. DICE Discussion Paper 267, pp. 1–39.

Cameron, A. Colin and Pravin K. Trivedi (2005). *Microeconometrics Methods and Applications*. Cambridge University Press.

Chamberlain, Gary (1987). “Asymptotic Efficiency in Estimation with Conditional Moment Restrictions”. In: *Journal of Econometrics* 34.3, pp. 305–334.

Compiani, Giovanni (2022). “Market Counterfactuals and the Specification of Multiproduct Demand: A Nonparametric Approach”. In: *Quantitative Economics* 13.2, pp. 545–591.

Compiani, Giovanni and Adam N. Smith (Feb. 2022). *Boundaries of Differentiated Product Markets and Retailer Pricing*. SSRN Scholarly Paper ID 3916922.

Conlon, Christopher and Jeff Gortmaker (2020). “Best Practices for Differentiated Products Demand Estimation with PyBLP”. In: *The RAND Journal of Economics* 51.4, pp. 1–77.

Doi, Naoshi (2022). “A Simple Method to Estimate Discrete-type Random Coefficients Logit Models”. In: *International Journal of Industrial Organization* 81.

Dubé, Jean-Pierre, Jeremy T. Fox, and Che-Lin Su (2012). “Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation”. In: *Econometrica* 80.5, pp. 2231–2267.

Duch-Brown, Néstor, Lukasz Grzybowski, André Romahn, and Frank Verboven (2017). “The Impact of Online Sales on Consumers and Firms. Evidence from Consumer Electronics”. In: *International Journal of Industrial Organization* 52, pp. 30–62.

Fan, Ying and Chenyu Yang (2020). “Competition, Product Proliferation, and Welfare: A Study of the US Smartphone Market”. In: *American Economic Journal: Microeconomics* 12.2, pp. 99–134.

Fosgerau, Mogens and Stephane Hess (2008). *Competing methods for representing random taste heterogeneity in discrete choice models*. MPRA Paper.

Fosgerau, Mogens and Stefan Mabit (2013). *Easy and Flexible Mixture Distributions*. Tech. rep. University Library of Munich, Germany, pp. 1–12.

Fox, Jeremy T., Kyoo il Kim, Stephen P. Ryan, and Patrick Bajari (2012). “The Random Coefficients Logit Model is Identified”. In: *Journal of Econometrics* 166.2, pp. 204–212.

Fox, Jeremy T., Kyoo il Kim, and Chenyu Yang (Dec. 2016). “A Simple Nonparametric Approach to Estimating the Distribution of Random Coefficients in Structural Models”. In: *Journal of Econometrics* 195.2, pp. 236–254.

Freyberger, Joachim (2015). “Asymptotic Theory for Differentiated Products Demand Models with many Markets”. In: *Journal of Econometrics* 185.1, pp. 162–181.

Gallant, A. Ronald and Douglas W. Nychka (1987). “Semi-Nonparametric Maximum Likelihood Estimation”. In: *Econometrica* 55.2, pp. 363–390.

Hansen, Bruce E. (2021). “Econometrics”. Wisconsin.

Hansen, Lars Peter (1982). “Large Sample Properties of Generalized Method of Moments Estimators”. In: *Econometrica* 50.4, pp. 1029–1054.

Heiss, Florian, Stephan Hetzenecker, and Maximilian Osterhaus (Feb. 2021). “Nonparametric estimation of the random coefficients model: An elastic net approach”. In: *Journal of Econometrics*.

Hess, S. and K. Axhausen (2005). “Distributional Assumptions in the Representation of Random Taste Heterogeneity”. In:

Hess, Stephane, Michel Bierlaire, and John W. Polak (2005). “Estimation of Value of Travel-Time Savings Using Mixed Logit Models”. In: *Transportation Research Part A: Policy and Practice* 39.2, pp. 221–236.

Hess, Stephane, Kenneth Train, and John W. Polak (2006). “On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit Model for vehicle choice”. In: *Transportation Research Part B: Methodological* 40.2, pp. 147–163.

Knittel, Christopher and Konstantinos Metaxoglou (Mar. 2014). “Estimation of Random-Coefficient Demand Models: Two Empiricists’ Perspective”. In: *Review of Economics and Statistics* 96, pp. 34–59.

Lu, Zhentong, Xiaoxia Shi, and Jing Tao (Oct. 2022). *Semi-Nonparametric Estimation of Random Coefficient Logit Model for Aggregate Demand*. SSRN Scholarly Paper.

Luce, R. Duncan (1959). *Individual Choice Behavior*. Tech. rep. Wiley, New York, pp. 1–51.

MacKay, Alexander and Nathan H Miller (2023). “Estimating Models of Supply and Demand: Instruments and Covariance Restrictions”. In: *Harvard Business School, Working Paper* 19-051.

Monardo, Julien (Nov. 2021). *Measuring Substitution Patterns with a Flexible Demand Model*. SSRN Scholarly Paper. DOI: 10.2139/ssrn.3921601.

Nevo, Aviv (2000a). “A Practitioner’s Guide to Estimation of Random-Coefficients Logit Models of Demand”. In: *Journal of Economics & Management Strategy* 9.4, pp. 513–548.

— (2000b). “Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry”. In: *The RAND Journal of Economics* 31.3, pp. 395–421.

Petrin, Amil (2002). “Quantifying the Benefits of New Products: The Case of the Minivan”. In: *Journal of Political Economy* 110.4, pp. 705–729.

Reynaert, Mathias and Frank Verboven (2014). “Improving the Performance of Random Coefficients Demand Models: The Role of Optimal Instruments”. In: *Journal of Econometrics* 179.1, pp. 83–98.

Simon, Carl P. and Lawrence E. Blume (Apr. 1994). *Mathematics for Economists*. New York London: W. W. Norton & Company.

Train, Kenneth (2016). “Mixed Logit with a Flexible Mixing Distribution”. In: *Journal of Choice Modelling* 19, pp. 40–53.

Train, Kenneth E. (Jan. 2008). “EM Algorithms for Nonparametric Estimation of Mixing Distributions”. In: *Journal of Choice Modelling* 1.1, pp. 40–69.

Wang, Ao (May 2022). “Sieve BLP: A semi-nonparametric model of demand for differentiated products”. In: *Journal of Econometrics*.

PREVIOUS DISCUSSION PAPERS

399 Kandelhardt, Johannes, Flexible Estimation of Random Coefficient Logit Models of Differentiated Product Demand, May 2023.

398 Hermes, Henning, Lergetporer, Philipp, Mierisch, Fabian, Peter, Frauke and Wiederhold, Simon, Discrimination on the Child Care Market: A Nationwide Field Experiment, April 2023.

397 Schmal, W. Benedikt, Haucap, Justus and Knoke, Leon, The Role of Gender and Coauthors in Academic Publication Behavior, March 2023.

396 Magin, Jana Anjali, Neyer, Ulrike and Stempel, Daniel, The Macroeconomic Effects of Different CBDC Regimes in an Economy with a Heterogeneous Household Sector, March 2023.

395 Dertwinkel-Kalt, Markus and Wey, Christian, Resale Price Maintenance in a Successive Monopoly Model, February 2023.
Forthcoming in: Journal of Industrial Economics.

394 Hermes, Henning, Krauß, Marina, Lergetporer, Philipp, Peter, Frauke and Wiederhold, Simon, Early Child Care and Labor Supply of Lower-SES Mothers: A Randomized Controlled Trial, December 2022.

393 Chowdbury, Shyamal, Schildberg-Hörisch, Hannah, Schneider, Sebastian O., and Sutter, Matthias, Information Provision Over the Phone Saves Lives: An RCT to Contain COVID-19 in Rural Bangladesh at the Pandemic's Onset, November 2022.

392 Normann, Hans-Theo and Sternberg, Martin, Human-Algorithm Interaction: Algorithmic Pricing in Hybrid Laboratory Markets, October 2022.
Published in: European Economic Review, 152 (2023), 104347.

391 Hunold, Matthias and Petrishcheva, Vasilisa, Foreclosure and Tunneling with Partial Vertical Ownership, September 2022.

390 Haucap, Justus and Heldman, Christina, The Sociology of Cartels, August 2022.

389 Döpper, Hendrik, Sapi, Geza and Wey, Christian, A Bargaining Perspective on Vertical Integration, May 2022.

388 Bachmann, Ronald, Gonschor, Myrielle, Lewandowski, Piotr and Madoń, Karol, The Impact of Robots on Labour Market Transitions in Europe, May 2022.

387 Fremerey, Melinda, Hörnig, Lukas and Schaffner, Sandra, Becoming Neighbors with Refugees and Voting for the Far-Right? The Impact of Refugee Inflows at the Small-Scale Level, April 2022.

386 Fischer, Kai, Alcohol Prohibition and Pricing at the Pump, March 2022.

385 Cobb-Clark, Deborah A., Dahmann, Sarah C., Kamhöfer, Daniel A. and Schildberg-Hörisch, Hannah, The Determinants of Population Self-Control, March 2022.

384 Sulka, Tomasz, Planning and Saving for Retirement, March 2022.

383 Cattan, Sarah, Kamhöfer, Daniel A., Karlsson, Martin and Nilsson, Therese, The Long-term Effects of Student Absence: Evidence from Sweden, March 2022.
Published in: Economic Journal, 133 (2023), pp. 888-903.

382 Martin, Simon and Rasch, Alexander, Collusion by Algorithm: The Role of Unobserved Actions, March 2022.

381 Haucap, Justus, Nedic, Radivoje and Şimşek, Talha, An Empirical Analysis of German Casino Locations, March 2022.
Published in: European Journal of Law and Economics, 55 (2023) pp. 291-311.

380 Haucap, Justus, Heldman, Christina and Rau, Holger A., Gender and Collusion, March 2022.

379 Schain, Jan Philip, Foreign Institutional Investors and the Great Productivity Slowdown, November 2022 (First Version February 2022).

378 Neyer, Ulrike and Stempel, Daniel, How Should Central Banks React to Household Inflation Heterogeneity?, January 2022.

377 Döpper, Hendrik and Rasch, Alexander, Combinable Products, Price Discrimination, and Collusion, January 2022.

376 Cobb-Clark, Deborah A., Dahmann, Sarah C., Kamhöfer, Daniel A. and Schildberg-Hörisch, Hannah, Self-Control and Unhealthy Body Weight: The Role of Impulsivity and Restraint, January 2022.

375 Martin, Simon and Schmal, W. Benedikt, Collusive Compensation Schemes Aided by Algorithms, December 2021.

374 Kouli, Yaman and König, Jörg, Measuring European Economic Integration 1880 - 1913 – A New Approach, December 2021.

373 Hunold, Matthias and Schad, Jannika, Single Monopoly Profits, Vertical Mergers, and Downstream Entry Deterrence, December 2021.

372 Werner, Tobias, Algorithmic and Human Collusion, November 2021.

371 Jovanovic, Dragan, Wey, Christian and Zhang, Mengxi, On the Social Welfare Effects of Runner-up Mergers in Concentrated Markets, September 2021.
Published in: Economics Bulletin, (2021), Article 201.

370 Jeschonneck, Malte, Collusion among Autonomous Pricing Algorithms Utilizing Function Approximation Methods, August 2021.

369 Gösser, Niklas, Gürer, Kaan, Haucap, Justus, Meyring, Bernd, Michailidou, Asimina, Schallbruch, Martin, Seeliger, Daniela and Thorwarth, Susanne, Total Consumer Time – A New Approach to Identifying Digital Gatekeepers, August 2021.

368 Fischer, Kai, Reade, J. James and Schmal, W. Benedikt, The Long Shadow of an Infection: COVID-19 and Performance at Work, August 2021.
Published in: Labour Economics, 79 (2022), 102281.

367 Suedekum, Jens, Place-Based Policies – How to Do Them and Why, August 2021.

366 Heiss, Florian, Ornaghi, Carmine and Tonin, Mirco, Inattention vs Switching Costs: An Analysis of Consumers’ Inaction in Choosing a Water Tariff, July 2021.

365 Cobb-Clark, Deborah A., Dahmann, Sarah C., Kamhöfer, Daniel A. and Schildberg-Hörisch, Hannah, Sophistication about Self-Control, July 2021.

364 Bie, Xiaodong and Ciani, Andrea, Born Similar, Develop Apart: Evidence on Chinese Hybrid Exporters, July 2021.

363 Ali, Nesma and Stiebale, Joel, Foreign Direct Investment, Prices and Efficiency: Evidence from India, July 2021.

362 Banerjee, Ritwik, Ibanez, Marcela, Riener, Gerhard and Sahoo, Soham, Affirmative Action and Application Strategies: Evidence from Field Experiments in Columbia, April 2021.

361 Wellmann, Nicolas and Czarnowske, Daniel, What Would Households Pay for a Reduction of Automobile Traffic? Evidence From Nine German Cities, March 2021.

360 Haucap, Justus, Moshgbar, Nima and Schmal, Wolfgang Benedikt, The Impact of the German "DEAL" on Competition in the Academic Publishing Market, March 2021.
Published in: Managerial and Decision Economics, 42 (2021), pp. 2027-2049.

359 Korff, Alex, Competition in the Fast Lane – The Price Structure of Homogeneous Retail Gasoline Stations, January 2021.

Older discussion papers can be found online at:
http://ideas.repec.org/s/zbw/dicedp.html

Heinrich-Heine-Universität Düsseldorf

Düsseldorfer Institut für Wettbewerbsökonomie (DICE)

Universitätsstraße 1, 40225 Düsseldorf

ISSN 2190-992X (online)
ISBN 978-3-86304-398-8

www.dice.hhu.de