

---

---

# Numerical Integration in Random Coefficient Models of Demand

---

---

Inauguraldissertation

zur Erlangung des akademischen Grades eines Doktors

der Wirtschaftswissenschaften

(Dr. rer. pol.)

durch die Wirtschaftswissenschaftliche Fakultät

der Heinrich-Heine-Universität Düsseldorf

---

**von:** Daniel Brunner, M. Sc.  
geboren am 09.06.1989 in Lüdenscheid

**Erstgutachter:** Prof. Dr. Florian Heiss

**Zweitgutachter:** Prof. Dr. Joel Stiebale

**Abgabedatum:** 09.06.2017

**Mündliche Prüfung:** Düsseldorf, den 26.07.2017



# Acknowledgment

This dissertation was carried out during my time as a doctoral researcher at the Chair of Statistics and Econometrics at the Heinrich-Heine-University of Düsseldorf. I am thankful to all people, who supported me on my way during and before writing this thesis.

In particular, I would like to thank Florian Heiss for his helpful and encouraging way of supervising this thesis that allowed me to realize and work on many interesting projects. Since my time in Mainz, his guidance increased my skills and interests in the field of econometrics and provided me with the optimal starting conditions for this dissertation. I also want to thank Joel Stiebale for accepting to be my second supervisor and his helpful comments.

Furthermore, I want to thank my colleagues that characterize this workplace not only as productive but enjoyable to work at. The impact of two colleagues and co-authors deserves special emphasis: Constantin Weiser and André Rohmahn provided me with valuable comments and suggestions about my work and shared a lot of their experience. Their feedback greatly increased the quality of this thesis.

Financially support from the Land Nordrhein-Westfalen and the Konrad-Henkel-Stiftung is gratefully acknowledged.

Finally, I want to thank my friends and family. Without the invaluable efforts of my parents this thesis would not have been possible. I am thankful to my sister Miriam for sharing her outside perspective and the proof reading of countless pages. Finally, I want to thank Angela for believing in me for such a long time.





# Contents

<b>List of Figures</b>	<b>VI</b>
<b>List of Tables</b>	<b>VII</b>
<b>List of Abbreviations</b>	<b>XI</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Simulation Error Causes Weak Identification in Random Coefficient Logit Models Using Aggregate Data</b>	<b>5</b>
2.1 Introduction . . . . .	5
2.2 The BLP Model . . . . .	8
2.2.1 Identification . . . . .	9
2.3 Simulation Error and its Propagation . . . . .	10
2.3.1 The Propagation of Simulation Error in the Sample Moment . . . . .	11
2.4 Estimation Setup . . . . .	12
2.4.1 Model Specification . . . . .	13
2.4.2 Instruments . . . . .	13
2.4.3 Numerical Integration . . . . .	14
2.4.4 Optimization Algorithms and Inner Convergence Threshold . . . . .	15
2.4.5 Benchmark Comparison and Additional Computational Details . . . . .	15
2.5 40,000 BLP Model Estimations . . . . .	16
2.5.1 Weak Identification Caused by Simulation Error . . . . .	17
2.5.2 Estimation Results . . . . .	20
2.6 Conclusion . . . . .	28
Appendix 2.A Computational Details . . . . .	31
2.A.1 Verifying Candidate Minima . . . . .	31
Appendix 2.B Speeding up the Contraction Mapping . . . . .	32
Appendix 2.C Documentation of the R Package BLPestimatorR . . . . .	34
<b>3 Implications of Adaptive Integration Rules for the Performance of Random Coefficient Models of Demand</b>	<b>45</b>
3.1 Introduction . . . . .	45

## Contents

3.2	The Model . . . . .	47
3.2.1	Demand Side . . . . .	47
3.2.2	Estimation Algorithm . . . . .	49
3.3	Market Share Integration . . . . .	50
3.3.1	Standard Approaches . . . . .	50
3.3.2	Adaptive Integration . . . . .	52
3.3.3	Integration in the BLP Model . . . . .	53
3.4	Results . . . . .	59
3.4.1	Data Generating Process . . . . .	59
3.4.2	Precision . . . . .	59
3.4.3	Computational Cost . . . . .	62
3.5	An Application to the Cereal Market . . . . .	62
3.6	Conclusion . . . . .	65
	Appendix 3.A Data Generating Process . . . . .	67
	Appendix 3.B Computational Details . . . . .	69
	Appendix 3.C Problems with SG Integration . . . . .	70
	Appendix 3.D Adaptive Integral Approximation on the Market Level . . . . .	71
<b>4</b>	<b>Copulas for Aggregated Demand Models with Partly Observed Heterogeneity</b>	<b>73</b>
4.1	Introduction . . . . .	73
4.2	The Model . . . . .	75
4.2.1	Estimation Algorithm . . . . .	76
4.3	Modeling Demographics' Joint Distribution . . . . .	77
4.4	Empirical Application . . . . .	83
4.4.1	Modeling the Demand of Deposits . . . . .	83
4.4.2	Geographic Market Definitions . . . . .	84
4.4.3	Data . . . . .	85
4.4.4	Results . . . . .	86
4.5	Conclusion . . . . .	91
	Appendix 4.A Variable Description . . . . .	92
	<b>Bibliography</b>	<b>94</b>
	<b>Eidesstattliche Versicherung</b>	<b>98</b>

# List of Figures

2.1	Empirical Distribution of $\hat{\xi}$ for Selected Products . . . . .	18
2.2	Shape of the Objective Function at the Global Minimum Candidate . . . . .	19
2.3	Range of Random Coefficient Estimates and Their Joint Statistical Signifi- cance . . . . .	24
2.4	Change in Consumer Welfare . . . . .	27
2.5	Computational Burden of Solving the Fixed Point . . . . .	33
3.1	Exemplary Integrand in Setting A . . . . .	55
3.2	Exemplary Integrand in Setting B . . . . .	56
3.3	Impact of Inaccurate Integration Rules on the GMM Function . . . . .	57
3.4	Computational Cost with Product- and Market-Wise Integration Rules . . . . .	58
3.5	Total Computation Time in Seconds Across Datasets . . . . .	63
3.6	Convergence of $\hat{\sigma}_{price}$ (Cereal Data) . . . . .	65
3.7	Exemplary Distribution of Market Shares . . . . .	68
3.8	Negative Integral Approximation with SG . . . . .	70
3.9	Market Level Integration with a Low and High Efficiency Loss . . . . .	71
4.1	Dependence Structure of Different Copula Classes . . . . .	80
4.2	Data Quality in Different Market Definitions . . . . .	85
4.3	Visualization of Empirical and Parametric Dependence Structures . . . . .	88



## List of Tables

2.1	Estimated Random Coefficients Using KM's 50 Monte Carlo Draws . . . . .	16
2.2	Objective Function Values Obtained Using Monte Carlo and MLHS Draws	20
2.3	Range and Spread of the Identified Minima . . . . .	21
2.4	Behavior of the Nested Fixed Point Estimator . . . . .	22
2.5	Choice of Optimizer and Convergence Threshold . . . . .	23
2.6	Range of Own-Price Elasticities Using MC and qMC Integration . . . . .	26
2.7	Counterfactual Price and Profit Changes for the Merging Parties following a Chrysler-GM Merger . . . . .	28
3.1	RMSE of $\hat{\sigma}$ for Non-Adaptive Measures (Simulated Data) . . . . .	60
3.2	RMSE of $\hat{\sigma}$ for Adaptive Measures (Simulated Data) . . . . .	61
3.3	RMSE of $\hat{\sigma}$ for Non-Adaptive Measures (Cereal Data) . . . . .	64
3.4	RMSE of $\hat{\sigma}$ for Adaptive Measures (Cereal Data) . . . . .	64
3.5	DGP Parameters . . . . .	68
4.1	Descriptive Statistics of Demographic Data . . . . .	87
4.2	Descriptive Statistics of Product Level Data . . . . .	87
4.3	Estimation Results . . . . .	89
4.4	Demand Estimation Variables . . . . .	92
4.5	Instrumental Variables . . . . .	93



## List of Abbreviations

ACS	American Community Survey
BLP	Berry, Levinsohn and Pakes (1995)
CBSA	core based statistical area
CDF	cumulative distribution function
CR	credit report
DGP	data generating process
FHFA	Federal Housing Finance Agency
GMM	generalized method of moments
IV	instrumental variable
KM	Knittel and Metaxoglou (2014)
MC	Monte Carlo
ML	Maximum Likelihood
NIC	National Information Center
OES	Occupational Employment Statistics
PDF	probability density function
PUMS	Public Use Microdata Sample
qMC	quasi Monte Carlo
RMSE	root mean squared error
SOD	summary of deposits
SG	sparse grids
TIGER	Topologically Integrated Geographic Encoding and Referencing





# 1 Introduction

The causal relation between a product's purchased quantity and observable characteristics of the product itself or its substitutes lies at the heart of many economic questions. Demand models relate these two types of variables and allow, for example, to answer industrial organization related questions, which analyze changes in the environment on market outcomes to derive policy implications. In a business context, the optimization of dynamic pricing strategies by predicting consumers' reactions to price changes gives another interesting and highly relevant application. Due to their fundamental importance, the evolution of demand models has a long history and is still in progress.

The model of Berry, Levinsohn and Pakes (1995, BLP hereafter) provides the most recent and powerful framework to estimate the demand of differentiated products with aggregate data. Their model treats products as a bundle of characteristics and explicitly considers individual consumer reactions to changes of these characteristics without actually observing individual decisions. Moreover, the usual problem of co-movements between observed and unobserved variables is addressed in a standard instrumental variable approach. Identifying the effect of a product's price is a well-known example, where valid instruments allow to rule out the impact of numerous unobservable quality components. By combining these properties with an efficient way of calculating a large amount of parameters, BLP created a tool to make realistic economic predictions. In the presence of consumer heterogeneity, this results in consumers, who rather switch to another product with similar characteristics, if they are confronted with an unfavorable change in the environment (e.g. a price increase).

The economic use of considering consumers' heterogeneity comes at the cost of a more complex and computationally burdensome model. When economists face problems with heterogeneity or other forms of uncertainty, the approximation of integrals without analytic solutions often becomes necessary. In the BLP model, a fundamental step of the estimation algorithm consists of numerous aggregations of individual purchase probabilities for a given product. The aggregation is realized by the following integral:

$$\int \Pr(\text{"purchase product j"}|\beta_i) f(\beta) d\beta .$$

The preference parameter of all individuals  $i$  can be either assumed to follow a certain parametric distribution (unobserved heterogeneity) or to depend on observed demographics (observed heterogeneity). This results, for example, in the possibility to model a lower

## 1 Introduction

price sensitivity for wealthier consumers and contributes to realistic substitution patterns. From a technical point of view, consumer aggregation gives an integral without an analytical solution.

This thesis deals with different challenges related to the numerical integration problem. More precisely, harmful effects of an inaccurate numerical integration in the integral above (chapter 2), adaptive approaches to mitigate this inaccuracies (chapter 3) and the problem of having only limited information about  $f(\beta)$  (chapter 4) are addressed in three different chapters. Each of them is explained in greater detail in the following.

Chapter 2 entitled “Simulation Error Causes Weak Identification in Random Coefficient Logit Models Using Aggregate Data” (co-authored with Florian Heiss, André Romahn and Constantin Weiser) investigates the propagation of integration error in the GMM objective function. It is shown that inaccurate integration impacts the shape of the GMM objective severely and produces wrong point estimates. This has immediate consequences for economic predictions: implied average own price elasticities, for example, based on different stochastic integration rules range from -44 to -2.3 with the lowest integration accuracy in BLPs’ car data. With the most accurate integration rule, this range collapses to -11 to -6.8. Accurate integration does not only contribute to parameter reliability but fixes many other problems reported in the literature (Knittel and Metaxoglou, 2014).

Reducing the integration error in BLPs’ model is further examined in chapter 3 (“Implications of Adaptive Integration Rules for the Performance of Random Coefficient Models of Demand”). Contrary to the strategy of the former chapter, integrals are not accurately approximated by a large amount of function evaluations but fewer and more cleverly placed ones. The idea is to evaluate the integral at an initial “standard” set of points and then, based on information about the function’s shape, perform an adjustment of the initial set. Adaptive integration rules are of particular use, if unfortunate parameter combinations cause very small regions on the integrand’s support with a high contribution to the integral. This is demonstrated with simulated data and Nevo’s cereal data (Nevo, 2000), where the gain of adaptive approaches in integration and parameter precision is impressive. Especially in the BLP model, adaptive approaches cause significant additional computational cost. This issue is addressed by proposing a new integration rule combining the advantages of standard and adaptive rules.

Chapter 4 entitled “Copulas for Demand Estimation Models with Partly Observed Heterogeneity” deals with the problem of limited information about  $f(\beta)$  in the presence of observed heterogeneity. BLP incorporate observed consumer heterogeneity by demographic variables and the integration problem processes them by integrating over their joint distribution. The approximation of the integral requires a precise knowledge of this distribution, which is often not available. This problem frequently arises with official statistics like census data that have to protect individuals’ privacy. To deal with some

forms of incomplete information, copula functions are proposed to model the demographics' joint distribution. The power of this approach is illustrated with an application from the banking literature by estimating the demand for deposits in commercial banks. Exclusive information about marginal distributions of demographics, age and income, on a fine geographic level is combined with a dependence structure from a broader geographic level. In this example, copulas enable the researcher to work with incomplete data and predict bank products with similar deposit rates to be good substitutes.

To summarize, this thesis discusses different challenges that can be encountered when making use of the great flexibility in BLPs' demand model and contributes by proposing possible solutions. In applied work, the numerical accuracy of the integral approximations constitutes an important trade-off between computational cost and accurate estimates. The results of chapter 2 indicate that this trade-off should be decided in favor of accuracy, because the integral approximation is a major source of error and easily eliminates the estimator's reliability. Chapter 3 discusses adaptive integration methods to deal with difficult integrands, where an acceptable error level cannot be obtained by standard methods. This is of particular use, if estimation results from very accurate standard integration rules are still impacted by significant integration noise. An accurate approximation is not possible without knowing the density in the integration problem, which gives the starting point of chapter 4. It is motivated by a highly practical problem: the joint distribution for public and official data is often not available on fine geographical levels. In this case, copula functions allow a combination of information from different sources, which enables the researcher to proceed with the demand estimation.



# 2 Simulation Error Causes Weak Identification in Random Coefficient Logit Models Using Aggregate Data

Co-authored with Florian Heiss, André Romahn and Constantin Weiser

## 2.1 Introduction

The seminal contribution of Berry, Levinsohn and Pakes (1995; henceforth BLP) has provided economists with an oligopoly model of differentiated product markets that is capable of producing realistic substitution patterns. The BLP model allows for partially or fully unobserved preference heterogeneity among economic agents and explicitly deals with the endogeneity of product attributes, typically price, and makes it possible to investigate counterfactual market outcomes. Apart from being applied to markets and questions that lie at the heart of Industrial Organization, the model's use has spread to the areas of environmental economics, insurance, voting preferences, and housing markets among others (see Table 1 in Berry and Haile (2015)). The model is parsimonious - compared to the standard logit model only a limited number of additional coefficients must be estimated - and its flexible functional form allows for arbitrary correlations between prices and markups. Products with similar attributes can be closer substitutes than products with very different characteristics.<sup>1</sup>

Consistent identification of the preference parameters depends on the sample moments, which are the product of relevant and valid instrumental variables (IV's) and the BLP model's structural error term. To obtain estimates of the structural error, the observed aggregate market shares have to be inverted. In contrast to the standard logit or nested logit model, where this inversion can be performed analytically (Berry (1994)), in the BLP model it must be computed numerically. BLP prove that this can be done using a contraction mapping and a suitable numerical integration technique for any candidate vector of preference parameters.

Our main finding is that with a crude numerical integration approach the estimate of the

---

<sup>1</sup>In the standard logit model prices and markups are negatively related: high-priced products have lower markups than their low-priced rivals. Moreover, it is highly likely that the best substitute for any other product is the product with the largest market share.

## 2 *Simulation Error Causes Weak Identification*

structural error term is overwhelmed by simulation error, which propagates in the sample moments and thereby in the GMM-IV objective function. Using the U.S. automobile data from BLP, we find that this causes many local minima with widely varying parameter estimates and model-implied economic predictions. We can therefore reaffirm the findings of Knittel and Metaxoglou (2014). However, when we use a large number of simulation draws and thereby reduce the propagation of simulation error in the objective function, we obtain tightly clustered model estimates and economic predictions.

Broadly, our findings are related to weak identification in nonlinear GMM-IV estimation as described in Stock, Wright and Yogo (2002). Here, the weak identification is not caused by weak instruments, however, but by random simulation error in the estimates of the structural error terms, which introduces many local minima in the estimator's GMM-IV objective function. Specifically, our findings are in line with the results from Berry, Linton and Pakes (2004), who derive the properties of the BLP nested fixed point estimator when the number of products becomes large. They prove that simulation error in the estimates is bounded if and only if the number of simulation draws grows proportionally with the square of the number of products. The impact of simulation error is therefore more pronounced in samples with many products. In the U.S. automobile data markets have between 72 and 150 products. We find the estimator no longer fails to converge to a local minimum for any of our random starting guesses if we use at least 5,000 Monte Carlo simulation draws to compute the aggregate market share inversion. With our least precise numerical integration approach, which as in Knittel and Metaxoglou (2014) uses 50 Monte Carlo draws, we obtain convergence to a local minimum in less than 63 percent of the estimations and a coefficient of variation among the objective values of the identified minima of more than 30 percent. In contrast, with our most accurate integration approach, which uses 10,000 modified latin hypercube sampling draws (MLHS draws, Hess et al. (2006)) we obtain convergence to a local minimum for 100 percent of the estimations and a coefficient of variation of less than 3 percent. This tight clustering of the identified minima also carries over to the parameter estimates and the model-implied economic predictions. To illustrate, with 50 Monte Carlo simulation draws the 95 percent confidence interval for the average own-price elasticity across all observations in the automobile data ranges from roughly -24 to -3. When 10,000 MLHS draws are used instead, the confidence interval tightens to the range from roughly -9 to -8. Our results are based on a total of 40,000 BLP model estimations using the U.S. automobile data. We use two numerical integration techniques, standard Monte Carlo and MLHS draws. For each of these approaches we consider eight different numbers of simulation draws that range from 50 to 10,000. For each number of draws, 50 independently sampled sets are generated and for each of these sets the BLP model is estimated 50 times using the same specification and random starting guesses as in Knittel and Metaxoglou (2014).

The importance of simulation error has largely been abstracted from in the existing

literature. Dubé, Fox and Su (2012) also identify the contraction mapping as the major source of numerical instabilities, but focus on the convergence threshold of the contraction that is set by the researcher and explicitly shut down the effect of simulation error.<sup>2</sup> A loose threshold speeds up the estimation, but also introduces approximation error in the objective function. In qualitative terms, we can confirm their findings, but in our setting with real world data and therefore simulation error we find the impact of the convergence threshold to be of second order (see the bottom panel of Table 2.5).<sup>3</sup> Reynaert and Verboven (2014) show that approximately optimal instruments can substantially reduce weak identification in BLP model estimation that is caused by weak instruments in the first stage. Again, the impact of simulation error is shut down, because a highly accurate numerical integration approach is used throughout the simulations and only markets with at most 20 products are considered. Similarly, in ongoing work Gandhi and Houde (2016) develop an alternative approach to approximately optimal instruments that avoids the two-stage estimation procedure. Also here the effect of simulation error is shut down, because the simulated market data has only 15 products and an accurate quadrature method is used to integrate the aggregate market shares. Knittel and Metaxoglou (2014), whose replication files we use to set up our own study, use the same set of 50 Monte Carlo draws throughout all of their estimations. This allows for substantial simulation error, but this aspect and the resulting weakly identified preference parameters are not investigated. Our results show that the numerical instabilities that have been identified by Knittel and Metaxoglou (2014) are driven by simulation error and these vanish once the numerical integration of aggregate market shares is computed accurately with many simulation draws.

The Knittel-Metaxoglou critique has urged researchers to more carefully implement their estimations and to more transparently report implementation details and estimation results (see the supplemental appendix of Goldberg and Hellerstein (2013) for an example). This development has fostered the credibility of structural estimation using the BLP demand model. We believe that our results further contribute to substantially simplifying the reliable implementation and communication of BLP model estimations using the nested fixed point algorithm.

The remainder of the paper is organized as follows. Section 2.2 briefly reviews the BLP model and its identification using the nested fixed point algorithm. Section 2.3 theoretically shows how simulation error propagates in the GMM-IV sample moment and objective function using results from Berry, Linton and Pakes (2004). Section 2.4 presents the setup for our large-scale study of the BLP estimator using the U.S. automobile data.

---

<sup>2</sup>See p. 2263 in Appendix A of Dubé, Fox and Su (2012): "...Because our focus is not on numerical integration error, we use the same sample of 1000 draws to compute the market shares in the data-generation and estimation phases."

<sup>3</sup>We consider convergence thresholds of  $10^{-16}$ ,  $10^{-9}$  and  $10^{-4}$ . Only the latter, extremely lax criterion, yields a noticeably wider spread in the model estimates.

## 2 Simulation Error Causes Weak Identification

We trace out the effects of simulation error in our 40,000 BLP model estimations in Section 2.5. In our Conclusion, we provide some takeaways for empirical practice and discuss the trade-off between numerical precision and thereby reliability on the one hand and computational burden on the other.

### 2.2 The BLP Model

We briefly present the BLP model and its estimation using the nested fixed point algorithm. Let  $i = 1, \dots, I$  index the population of consumers. Each consumer faces the discrete choice among  $j = 0, \dots, J_t$  options in  $t = 1, \dots, T$  markets.  $j = 0$  is the outside option of not purchasing any of the  $J_t$  available differentiated products. Consumer  $i$ 's indirect utility from purchasing product  $j$  in market  $t$  is given as follows.

$$\begin{aligned} u_{ijt} &= \delta_{jt}(\xi_t, \theta_1) + \mu_{ijt}(\xi_t, \nu_{it}, \theta_2) + \epsilon_{ijt}, \quad u_{i0t} \equiv \epsilon_{i0t} \\ \delta_{jt} &= x_{1,jt}\beta - \alpha p_{jt} + \xi_{jt} \\ \mu_{ijt} &= \sum_{k=1}^K x_{2j,t}^{(k)} \nu_{it,k} \sigma_k \end{aligned} \tag{2.1}$$

The normalization of the outside option's utility is necessary, because the level of utility is not separately identified. We follow Nevo (2000) and notationally distinguish the preference parameters that enter the estimation algorithm linearly,  $\theta_1 = (\alpha, \beta)$ , and the parameters that enter nonlinearly,  $\theta_2 = \sigma$ . Note that  $x_{1,jt}$  is the row vector of all  $K_1$  observed product characteristics in market  $t$  and  $x_{2,jt}$  is a subset of  $x_{1,jt}$  and contains the  $K_2$  observed characteristics with preference heterogeneity in the population.  $\delta_{jt}$  is the mean utility of product  $j$  and it depends on price,  $p_{jt}$ , the  $K_1$  observable characteristics,  $x_{1,jt}$ , and  $\xi_{jt}$ , the product characteristic that is unobserved by the econometrician.  $\mu_{ijt}$  is the consumer-specific deviation from mean utility, which is driven by preference heterogeneity over the  $K_2$  characteristics in the population. This heterogeneity is captured by  $\nu_{it}$ . For the model's exposition and to offer a meaningful comparison with the results in Knittel and Metaxoglou (2014), we assume that the  $K_2$  dimensions of  $\nu_t$  are independently distributed. This assumption can be relaxed and preference correlations between the  $K_2$  characteristics can be modeled to achieve more flexible substitution patterns.

$\epsilon_{ijt}$  is a consumer-product specific match value that we assume follows a Type I extreme value distribution. This assumption gives closed form expressions for the consumer-specific choice probabilities for product  $j$ .

$$s_{ijt}(\delta_t, \nu_t, \theta_2) = \frac{\exp(\delta_{jt} + \mu_{ijt}(\nu_t, \theta_2))}{1 + \sum_k \exp(\delta_{kt} + \mu_{ikt}(\nu_t, \theta_2))} \tag{2.2}$$

The model-implied aggregate market share function integrates over the consumer-specific



choice probabilities, where we let  $f(\nu_t)$  denote the population density of consumer heterogeneity.

$$s_{jt}(\delta_t, \theta_2) = \int \frac{\exp(\delta_{jt} + \mu_{ijt}(\nu_t, \theta_2))}{1 + \sum_k \exp(\delta_{kt} + \mu_{ikt}(\nu_t, \theta_2))} f(\nu_t) d\nu_t \quad (2.3)$$

This integral does not have an analytical solution, but can be solved numerically. We have to deal with the fact, however, that the population distribution of consumer preferences is not directly observed by the econometrician.<sup>4</sup> We therefore have to assume a joint distribution of preferences over the  $K_2$  characteristics and simulate  $\nu_t$  accordingly. In particular, we use  $r = 1, \dots, R$  simulation draws and assume that consumers' preference for each of the  $K_2$  characteristics is distributed standard normal, so that  $\nu_{rt,k} \sim N(0, 1)$ . The model-implied aggregate market share is the average taken over the  $R$  simulated consumer-specific choice probabilities.

$$s_{jt}(\delta_t, \theta_2) \approx \frac{1}{R} \sum_{r=1}^R s_{rjt} = \frac{1}{R} \sum_{r=1}^R \frac{\exp(\delta_{jt} + \mu_{rjt}(\theta_2, \nu_t^R))}{1 + \sum_k \exp(\delta_{kt} + \mu_{rkt}(\theta_2, \nu_t^R))} \quad (2.4)$$

### 2.2.1 Identification

The unobserved characteristic or structural error term,  $\xi$ , is a vertical product attribute. Consumer utility is increasing in  $\xi$ , so that consumers always prefer more of it. Contrary to the econometrician, both firms and consumers observe  $\xi$ , which yields a positive correlation between the error term and price. We obtain consistent estimates of the preference parameters by imposing a standard GMM-IV moment restriction. Let  $z_{jt}$  denote a row vector of  $L \geq K_2$  relevant and valid instrumental variables. The moment restriction is

$$E[\mathcal{G}(\theta_2)] = E \left[ \frac{1}{N} \sum_{t=1}^T \sum_{j=1}^{J_t} z_{jt} \xi_{jt}(\theta_2) \right] = 0, \quad (2.5)$$

where  $N = \sum_t J_t$  is the total number of observations in the sample.

Cost shifters that vary at the product level would be ideal candidates for the excluded instruments. The required data, however, is usually not available. To construct suitable instruments we make the assumption that the  $\xi$ 's are mean independent of the observed product characteristics.

$$E(\xi_t | x_t) = 0 \quad (2.6)$$

If this assumption holds, any function of the observed product characteristics qualifies as a potentially valid instrument for price. Such functions also give relevant instruments, because the observed characteristics enter each product's equilibrium pricing function. BLP use this insight to derive a set of instruments that can be viewed as a first-order

<sup>4</sup>Depending on data availability, consumer heterogeneity can be partially directly observed by using (relevant) consumer demographics. This introduces an additional term that enters utility additively separably and that interacts the observable product attributes and consumer demographics.

## 2 Simulation Error Causes Weak Identification

approximation of a pricing game played between firms: for each product  $j$  sold by firm  $f$  the observable characteristics of all products sold by the same firm are summed over and the observable characteristics of all products sold by rival firms are summed over,  $z_{jt,own} = \sum_{k \neq j, k \in \mathcal{F}_{ft}} x_{kt}$ ,  $z_{jt,other} = \sum_{k, k \notin \mathcal{F}_{ft}} x_{kt}$ . This gives  $2K_1 \geq K_2 + 1$  excluded instruments to identify the price coefficient,  $\alpha$ , and the standard deviations of the random coefficients,  $\theta_2$ .

Let  $\theta_2^*$  denote the true population preference parameters. Given a suitable weighting matrix,  $W$ , we obtain consistent and as Berry, Linton and Pakes (2004) prove asymptotically normally distributed estimates of  $\theta_2^*$  by minimizing the GMM-IV objective function, which is a norm of the sample moment  $\frac{1}{N} \sum_{t=1}^T \sum_j z_{jt} \xi_{jt}$ .<sup>5</sup>

$$\hat{\theta}_2 = \arg \min_{\theta_2} \mathcal{J}(\theta_2) = \arg \min_{\theta_2} \xi(\theta_2)' z W z' \xi(\theta_2) \quad (2.7)$$

Here,  $\xi$  and  $z$  are the vertically stacked market-specific structural error terms and instrument matrices, respectively.

### 2.3 Simulation Error and its Propagation

Our large-scale study of the BLP estimation algorithm uses 20 years of nationally aggregated car sales, which is based on millions of individual purchase decisions. Sampling error is therefore likely to be negligible in our estimations. To simplify notation in this section, we therefore abstract from sampling error and assume throughout that market shares are observed without error. Without loss of generality, we also drop the time subscript here.

For each candidate vector of the nonlinearly entering preference parameters,  $\theta_2$ , the BLP model rationalizes the observed shares in the data. The model-implied and observed shares are matched.<sup>6</sup>

$$\mathcal{S}_j = s_j(\delta^{match}, \theta_2, \nu^R) \text{ for all } j, \quad (2.8)$$

where  $\mathcal{S}$  denotes the observed market shares. Berry (1994) proves the existence of a fixed point that gives the unique solution,  $\delta^{match}$ , to this system of equations for any candidate vector  $\theta_2$  and set of simulation draws  $\nu$ . We iterate over the equation

$$\delta_j^{iter+1} = \delta_j^{iter} + \log(\mathcal{S}_j) - \log(s_j(\delta^{iter}, \nu^R, \theta_2)) \quad (2.9)$$

until the distance between successive iterates falls below the chosen convergence threshold,  $|\delta_j^{iter+1} - \delta_j^{iter}| \leq \epsilon_{inner}$  for all  $j$ . If this inequality holds, the current update for the vector of mean utilities is accepted as  $\delta^{match}$ . We obtain the corresponding vector of structural error terms as the residual of a two-stage least squares regression of  $\delta^{match}$  on

<sup>5</sup>We assume that the weighting matrix is homoscedastic,  $W = (z'z)^{-1}$ .

<sup>6</sup>The magnitude of the relative approximation error,  $|(\mathcal{S}_j - s_j)/\mathcal{S}_j|$ , is bounded from above by the inner convergence threshold,  $\epsilon_{inner}$ .

the observed product characteristics. This step also delivers the estimates of the linearly entering parameters.

$$\delta_j^{match} = x_{1,j}\beta - \alpha p_j + \xi_j^{match} \quad (2.10)$$

$\xi^{match}$  is therefore a function of  $\theta_2$ , the observed market shares and product characteristics, and the set of simulation draws. To reduce notation, we do not explicitly write out the dependence on the observed shares and product characteristics. Thus, we let  $\xi^{match}(\theta_2, \nu)$  denote the vector of unobserved characteristics that matches the observed shares with the model-implied shares. It is important to note that this only gives the true vector of unobserved product characteristics,  $\xi^*$ , if it is evaluated at the population vector of preference parameters and the population distribution of preferences,  $\xi^* = \xi^{match}(\theta_2^*, \nu^*)$ .

### 2.3.1 The Propagation of Simulation Error in the Sample Moment

At each iteration of the contraction mapping, (2.9), the model-implied aggregate market shares must be computed using (2.4). Simulation error is introduced because of differences between the consumer population and the simulated sample of consumer preferences using  $R$  simulation draws.<sup>7</sup> Following Berry, Linton and Pakes (2004), we define the vector of simulation error as follows.

$$e^R \equiv s(\xi^{match}(\theta_2, \nu^*), \theta_2, \nu^*) - s(\xi^{match}(\theta_2, \nu^R), \theta_2, \nu^R) \quad (2.11)$$

By construction, simulation error vanishes for any candidate parameter vector  $\theta_2$  if we were able to impose the true population distribution of preference heterogeneity,  $\nu^*$ , when computing the model-implied aggregate market shares.

Berry, Linton and Pakes (2004), show that  $\xi^{match}(\theta_2, \nu^R)$  can be expanded around the point  $(\theta_2^*, \nu^*)$ .

$$\xi^{match}(\theta_2, \nu^R) \approx \underbrace{\xi^{match}(\theta_2, \nu^*)}_{\text{exact inversion}} - \underbrace{\left[ \frac{\partial s(\xi^*, \theta_2^*, \nu^*)}{\partial \xi'} \right]^{-1}}_{\text{effect of simulation error}} e^R \quad (2.12)$$

Here,  $[\partial s / \partial \xi']$  is the matrix of market share derivatives with respect to the unobservable product characteristics with dimension  $J \times J$ .

The first term on the right-hand side is the estimate of the structural errors that we would obtain if we could match the population distribution of preference heterogeneity exactly in the numerical integration of the aggregate shares. We only use a sample of  $R$  simulation draws, however, which causes deviations of the model-implied shares from their observed sample counterparts at  $\xi^*$ . How these deviations affect our computations

<sup>7</sup>We ignore an additional potential error here. We impose that consumer preference heterogeneity follows a normal distribution. The true preference distribution could be non-normal, which would lead to biased parameter estimates.

## 2 Simulation Error Causes Weak Identification

of  $\xi^{match}$  depends on how sensitive  $\xi^{match}$  is with respect to changes in the entries of the model-implied aggregate market share vector,  $s$ . This sensitivity is measured by the inverse of the matrix  $[\partial s / \partial \xi']$ . The smaller the derivatives, the larger is the distortion of  $\xi^{match}$  that is caused by simulation error. Thus, the inversion of aggregate market shares magnifies simulation error in the estimates of the structural error term.

By distorting the estimates of  $\xi$ , simulation error propagates in the sample moments and thereby in the GMM-IV objective function. Plugging (2.12) into the sample moment gives

$$\mathcal{G}(\theta_2, \nu^R) \approx \frac{1}{N} \sum_{t=1}^T \sum_{j=1}^{J_t} z_{jt} \left( \xi^{match}(\theta_2, \nu^*) - \left[ \frac{\partial s(\xi^*, \theta_2^*, \nu^*)}{\partial \xi'} \right]^{-1} e^R \right), \quad (2.13)$$

which stresses that the computed sample moment depends explicitly on the simulation error that is caused by the specific set of draws  $\nu^R$ . There is an analogy to the definition of weak identification in Stock, Wright and Yogo (2002) for nonlinear GMM estimation. Given that  $e^R$  is random and propagates into the GMM-IV objective function, the shape and location of the objective function, (2.7), are affected. There can be several values for  $\theta_2 \neq \theta_2^*$  for which the objective function attains a local minimum. This explains how many local minima with widely varying parameter estimates and economic implications are obtained with a crude numerical integration approach.

Berry, Linton, and Pakes (2004) also show that the extent of the magnification depends on the number of products in the market. In equilibrium as more and more products enter a market it must be the case that product-level market shares fall. This is because in the BLP model each product is substitutable with every other product to some extent. Specifically, it is assumed that all shares move inversely proportional with  $J$  (Condition S/equation (20) in Berry, Linton and Pakes (2004)). The derivatives of the shares with respect to  $\xi$  are proportional to market shares and therefore also decline with  $J$ . As simulation error is scaled by the inverse of  $[\partial s / \partial \xi']$ , the magnification of simulation error is greater in samples with many products. To bound simulation error as the number of products becomes large, the number of simulation draws must grow proportionally with the square of the number of products in the market.

## 2.4 Estimation Setup

We study how numerical integration accuracy affects the behavior and outcomes of the BLP nested fixed point estimation algorithm using the original automobile market data from BLP. This data set covers 20 years of annually aggregated car model-level sales for the United States starting in 1971.<sup>8</sup> We think this choice presents two advantages.

---

<sup>8</sup>For a detailed description of the data set, see BLP.

First, this is a real world data set where the number of products ranges from 72 to 150 and that is based on a large sample of individual consumer purchases. Sampling error, therefore, is likely to be negligible, while simulation error should play a substantial role in this setting. Second, the same data set has been used by Knittel and Metaxoglou (2014; henceforth KM) to carefully document several numerical instabilities in the BLP estimation algorithm. The study is exemplary in terms of its replicability and transparency and has motivated researchers to more carefully implement and report the outcomes of their BLP model estimations (Hellerstein and Goldberg (2013)). We therefore base our large-scale study of the BLP estimation algorithm on their replication files to demonstrate that the reported numerical instabilities are solved once the numerical integration of the model-implied aggregate shares is performed accurately. Specifically, we estimate exactly the same specification using the same set of instruments and random starting guesses for  $\theta_2$ .

### 2.4.1 Model Specification

Consumers' indirect utility is specified as follows.

$$u_{ijt} = \beta_{i0} + hpwt_{jt}\beta_{i1} + space_{jt}\beta_{i2} + aircon_{jt}\beta_{i3} + mpg_{jt}\beta_{i4} - \alpha_i price_{jt} + \xi_{jt} + \epsilon_{ijt}, \quad (2.14)$$

where *hpwt* is the horsepower-weight ratio, *space* is the length times the width of the car, *aircon* is a dummy indicating whether the car has air conditioning built in and *mpg* measures the car's miles per gallon. Except for *space* all observable characteristics, including price and the constant term, have a random coefficient. The specification therefore estimates 5 random coefficients in total. We assume that the random coefficients are distributed normally and independently. Thus,  $\alpha_i = \alpha + \sigma_p \nu_{i,p}$  and  $\beta_{i,k} = \beta_k + \sigma_k \nu_{i,k}$  with  $\nu_{i,k}, \nu_{i,p} \sim N(0, 1)$  for  $k = 1, \dots, K_2 = 5$ .

### 2.4.2 Instruments

We use the instruments from the Knittel and Metaxoglou (2014) replication files. These are the standard characteristics-based or BLP-type instruments. Using all five non-price product characteristics, including the constant, these instruments sum over the characteristics of all other cars produced by the same firm, and sum over the characteristics of all cars produced by rival firms. We therefore have 10 instruments for price and the 5 nonlinearly entering parameters. Given that the literature on approximately optimal instruments shows that these standard characteristics-based instruments can be weak and thereby yield weak identification of the random coefficients, it is important for us to show that for specification (2.14) this is not the case. We simply run the first-stage regression of price on the instruments for two cases. First, we only explain the variation in price using the excluded (BLP-type) instruments. This regression gives an F-statistic of 43.9.

## 2 Simulation Error Causes Weak Identification

Second, we use the full instrumental variable matrix that also contains the observed non-price characteristics, which given their assumed exogeneity instrument for themselves. Not surprisingly, this gives a higher F-statistic, namely roughly 248. To assess whether the observed characteristics drive out the excluded instruments, we compute the F-statistic for the null that only the coefficients of the excluded instruments are zero. This F-statistic has a value of 43.7, almost unchanged from the first-stage regression without the observed characteristics. In both cases, we comfortably pass the rule of thumb that the F-statistic should be greater than 10. The excluded instruments also comfortably pass the critical values reported in table 1 of Stock, Wright and Yogo (2002). Gandhi and Houde (2016) also suggest an ex-ante test of weak identification that estimates the standard logit model, which the BLP model collapses to for  $\theta_2 = \mathbf{0}$ , with the excluded instruments as additional regressors. If the estimated coefficients for the instruments are zero, then these instruments fail to reject the logit model and thereby suggest that there is no preference heterogeneity among consumers. We test the hypothesis that the coefficients for the excluded instruments are all zero in this setting and can comfortably reject the null at confidence levels beyond 99 percent. We conclude that our results are driven by simulation error and not by weak instruments.<sup>9</sup>

### 2.4.3 Numerical Integration

The existing literature has made extensive use of Monte Carlo, quasi-Monte Carlo and quadrature methods (Heiss and Winschel (2008)). For a given computational burden qMC and quadrature methods are capable of giving better approximations than standard Monte Carlo methods. Nevo (2001) and Sovinsky Goeree (2008), respectively, use Halton draws and antithetic sampling to increase simulation efficiency, for example.

We base the numerical integration of the model-implied aggregate shares, (2.4), on two Monte Carlo simulation methods: the standard Monte Carlo approach and modified latin hypercube sampling draws (MLHS draws). Hess, Train and Polak (2006) find that in finite samples MLHS draws perform roughly on par with Halton draws. We also find that roughly 4 to 5 times as many standard Monte Carlo draws are needed to attain the same integration accuracy as with a given number of MLHS draws. For our study MLHS draws offer the advantage that it is straightforward to obtain measures of how the number of simulation draws affects the spread of estimation outcomes. We can simply compute the variance of some estimation outcome for a given number of simulation draws. With standard Halton draws or any quadrature method this is no longer the case, because for these approaches the simulation draws or nodes are based on deterministic number sequences. By construction, therefore, for a given number of draws or nodes there is no variation across different estimations. To obtain a measure of simulation error in the

---

<sup>9</sup>This is also evident by the fact that the only model input that varies in our estimations is the specific set of simulation draws.

estimation outcomes, we would have to compute error bounds for these methods, which are model-specific and cumbersome to implement.<sup>10</sup>

We use 8 different numbers of draws for both simulation approaches that range from 50 to 10,000.<sup>11</sup> To exclude the possibility that our findings are due to any specific set of draws, we generate 50 independently sampled sets of  $\nu$  for each of the 8 different numbers of draws.

Therefore, with the 50 starting guesses for  $\theta$  from KM, each number of draws requires us to estimate specification (2.14) 2,500 times. With 8 different numbers of simulation draws and 2 simulation approaches, we estimate the BLP model 40,000 times.

#### 2.4.4 Optimization Algorithms and Inner Convergence Threshold

An important part of the Knittel-Metaxoglou critique is that the choice of optimization algorithm can have a substantial effect on the estimation outcomes. Similarly, Dubé, Fox and Su (2012) caution that a loose inner convergence threshold can produce many local minima with widely varying estimates. We investigate both of these aspects in our setting and with an accurate numerical integration approach we find the choice of optimization algorithm to be irrelevant (see the top panel of Table 2.5) and the impact of the inner convergence threshold to be of second order (see the bottom panel of Table 2.5). We therefore base all of our 40,000 estimations in the main part of our study on a trust region optimizer with an analytical gradient<sup>12</sup> and on a stringent inner convergence threshold of  $10^{-16}$ .

#### 2.4.5 Benchmark Comparison and Additional Computational Details

We deviate from the implementation of the nested fixed point algorithm in some aspects from KM. The changes that we implement make the algorithm more robust and enforce a uniform convergence threshold for the market share inversion throughout. We detail these changes in the Appendix, where we also define the criteria for a candidate parameter vector to identify a local minimum. Using the same set of 50 Monte Carlo draws as KM, we can also demonstrate that our changes do not fundamentally impact the Knittel-Metaxoglou critique at this level of numerical integration accuracy. Table 2.1 presents the results of estimating specification (2.14).

We find that 44 of the 50 random starting guesses for  $\theta_2$  yield a local minimum. Rounding the objective function values of these minima to two digits, we obtain 5 minima that range between 207.72 and 226.94. This is a more narrow range than that reported by

<sup>10</sup>We have also found that numerical integration on sparse grids works well in the majority of estimations, but when a poor guess is used for the nonlinearly entering parameters the method produces inadmissible negative market shares in about 10 percent of the estimations.

<sup>11</sup>These numbers are 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000.

<sup>12</sup>Specifically, we use Matlab's `fminunc` optimizer algorithm. This corresponds to KM's DER1-QN1 optimizer.

Table 2.1: Estimated Random Coefficients Using KM's 50 Monte Carlo Draws

	Min 1	Min 2	Min 3	Min 4	Min 5
price	0.328**	0.182	0.162	0.107**	0.134**
constant	7.480**	2.720**	5.232**	2.001	1.598**
hpwt	2.565	1.063	0.165	5.781**	1.481
aircon	8.800**	0.484	3.629	0.425	4.231**
mpg	0.098	0.687	0.134	1.767**	1.163**
$\mathcal{J}(\hat{\theta}_2)$	207.7	215.1	216.0	224.6	226.9
$\bar{\eta}_{jj}$	-10.53	-7.782	-5.787	-4.606	-5.387
$\boldsymbol{\eta}$	-1.007	-1.374	-0.946	-0.945	-1.263
Wald-statistic	23.26	87.49	72.69	112.4	93.74

Note: \* and \*\* indicate statistical significance at the 95 and 99 percent confidence levels, respectively. Only the estimated random coefficients,  $\theta_2$ , are shown. All inputs to the estimation, including the 50 simulated draws for consumer preference heterogeneity,  $\nu_{KM}$ , are identical to those used by KM. We compute HAC standard errors.  $\bar{\eta}_{jj}$  is the average own-price elasticity and  $\boldsymbol{\eta}$  is the aggregate demand elasticity averaged over all 20 markets. The null hypothesis of the Wald test is  $\theta_2 = 0$ .

KM. This indicates that at least some of the lack of robustness in their estimation results could stem from scaling issues, which we avoid (see the Appendix).<sup>13</sup> Overall, however, the Knittel-Metaxoglou critique is broadly reaffirmed. For each random coefficient the ratio of its greatest to smallest point estimate across the 50 starting guesses is at least 3 (price) and reaches up to 35 (hpwt). The model-implied average own-price elasticity and the aggregate demand elasticity vary by factors of roughly 2.3 and 1.4, respectively. Moreover, the statistical significance of individual random coefficients changes substantially across minima. In fact, for each coefficient it is possible to select a minimum where that coefficient is either statistically significant or insignificant at the 95 percent confidence interval. Finally, the Wald statistic we obtain by testing the estimated BLP model against the simple logit model also ranges widely from 23 to 112.

## 2.5 40,000 BLP Model Estimations

We present the outcomes of the 40,000 BLP model estimations in two parts. First, we demonstrate how simulation error propagates in the GMM-IV sample moments and thereby in the objective function of the estimator. This propagation explains the numerical instabilities documented by KM. Moreover, simulation error can be reduced substantially by increasing the number of simulation draws and thereby raising the accuracy of numerical integration.

<sup>13</sup>KM's high cutoff of 30 for the Euclidean norm of the gradient is likely to contribute to a wider range of outcomes, too.



Second, we document how the mean and spread of the estimation outcomes and the corresponding economic predictions change with the number of simulation draws. From 500 draws onwards the spread of the estimation outcomes is falling monotonically in the number of simulation draws for both integration approaches. The estimated parameters and economic implications fall into increasingly narrowing intervals. With regards to the mean of the estimation outcomes, our findings show that simulation error biases the estimation outcomes in the sample of U.S. automobile market data. Thus, as the number of simulation draws changes, so do the means of the estimation outcomes.

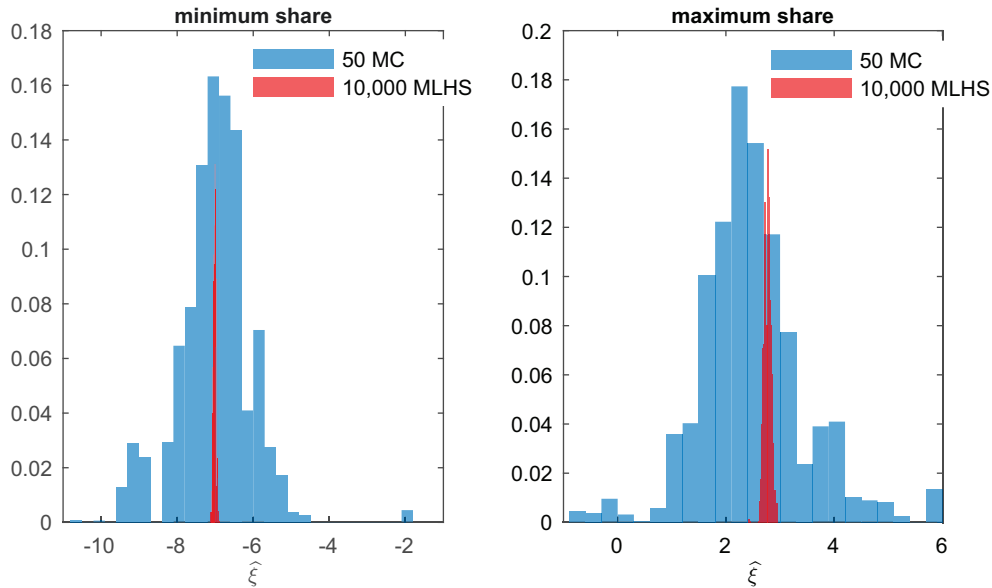
### 2.5.1 Weak Identification Caused by Simulation Error

Simulation error propagates in the estimates of the structural error term. Figure 2.1 shows how the number of simulation draws, which is inversely related to the magnitude of simulation error, affects the estimates of the structural error term. Both panels plot the empirical distributions of the estimated unobservable characteristic for the products with the smallest and largest market shares in the sample. These distributions are based on our least accurate numerical integration approach, namely 50 Monte Carlo draws (blue), and our most accurate approach, 10,000 MLHS draws (red). For each of these integration approaches the model is estimated 2,500 times and each estimation that converges to a local minimum gives us one estimate of the structural error.

The differences between the distributions are remarkable. Using only 50 Monte Carlo draws, the variances of the estimated structural errors are 0.843 and 0.952 for the products with the smallest and greatest market shares in the sample, respectively. If we use 10,000 MLHS draws, instead, we obtain corresponding variances of only 0.001 and 0.004. In terms of 99 percent confidence intervals, with 50 Monte Carlo draws, the estimate of the unobservable attribute for the products with the smallest and greatest shares are, respectively, the ranges from -9.5 to -4.8 and -0.6 to 5.8. Using our most accurate numerical integration approach gives the corresponding confidence intervals of -7.1 to -6.9 and 2.6 to 3.0. Adopting a crude integration approach, therefore, produces simulation error that easily overwhelms the estimates of the error terms. This holds across the sample. We obtain qualitatively identical figures for the products with the mean and median market shares, for example. Thus, simulation error randomly perturbs the estimates of each product's unobserved characteristic and thereby affects the shape of the GMM-IV objective function.

As a second illustration, we therefore trace out how the shape of the GMM-IV objective function depends on the number of simulation draws. As the starting point for this exercise we take the parameter estimate from our global minimum candidate, which is based on numerical integration using 10,000 MLHS draws:  $\tilde{\theta}_2 = (1.52, 5.84, 3.39, 0.41, 0.10)'$ . It is not essential that we pick this specific point. We would obtain qualitatively identical results at other candidate values of  $\theta_2$ . We hold all  $\sigma$  values constant except for  $\sigma_{price}$ .

Figure 2.1: Empirical Distribution of  $\hat{\xi}$  for Selected Products

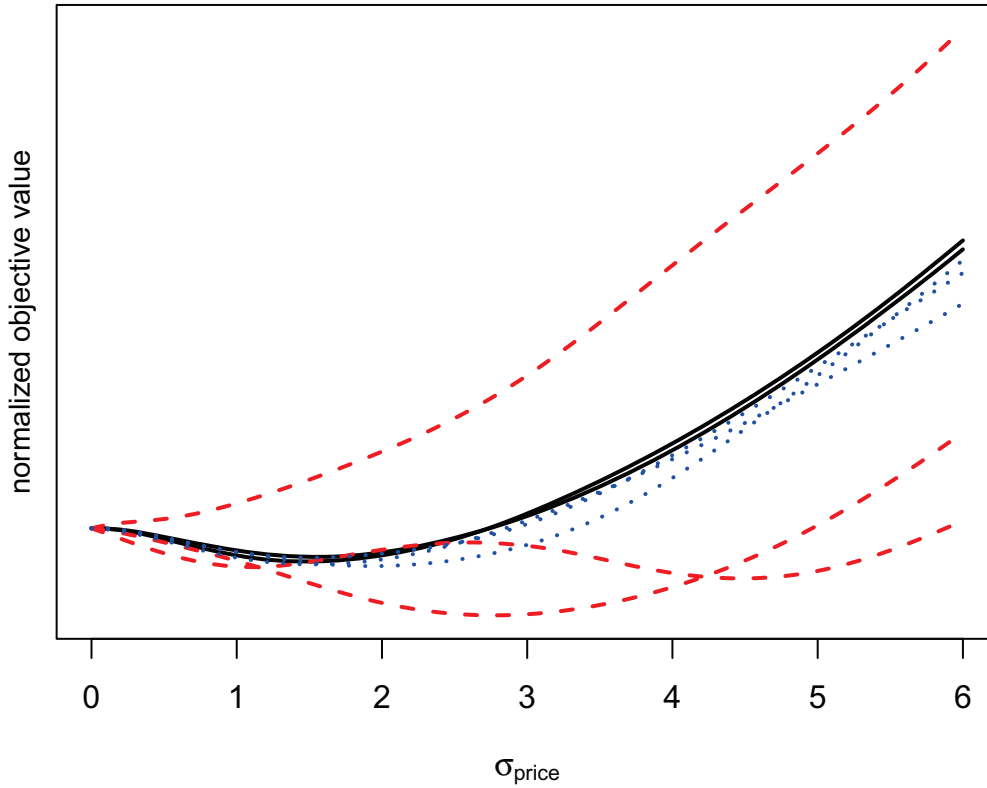


Note: Both panels show the empirical relative frequency plots for the estimated structural error term across 2,500 estimations of the BLP model for a given number of simulation draws. The 2,500 estimations are based on 50 independently sampled sets of preference heterogeneity for a given number of draws. For each of these 50 sets, we estimate the BLP model using 50 random starting guesses. With 50 MC draws only 1,562 of the 2,500 estimation runs converge to a local minimum. For the 10,000 MLHS draws all estimations converge. Estimations that fail to converge are not included in the plots.

Figure 2.2 plots the shape of the objective function along the  $\sigma_{price}$ -dimension for several sets of simulation draws. The red dashed, blue dashed and solid green lines are respectively based on three independently generated samples of 50, 500 and 5,000 MLHS draws. With 50 draws, the shape of the objective function changes markedly across the three sets. One of the three sets attains a local minimum at zero, which implies no preference heterogeneity along the price dimension. This outcome is strongly rejected by our full set of estimations. A second set produces a shape that yields two minima along the price dimension. The third set has only one local minimum, but gives a biased estimate of  $\sigma_{price}$  compared with our global minimum candidate. For 500 MLHS draws we can already see that the shape of the objective function stabilizes. There is only one local minimum for all sets, which is not located at zero. There is, however, visible variation in the location of the minima. For 5,000 MLHS draws the three sets generate objective functions that appear to be congruent. The shape of the objective is stable, the minimum at zero is ruled out and the local minima across the three sets are located very close to our global minimum candidate.

Lastly, we present a variation on this theme. Instead of tracing out the objective in one dimension for different sets of simulation draws, we fix the nonlinearly entering parameters

Figure 2.2: Shape of the Objective Function at the Global Minimum Candidate



Note: The objective function is plotted along its  $\sigma_{price}$  dimension for 9 sets of Monte Carlo simulation draws: 3 sets using 50 draws (red dashed), 3 sets using 500 draws (blue dotted), and 3 sets using 5,000 draws (black solid). To ensure that all objective functions share the same value at  $\sigma_{price} = 0$ , we subtract the objective value at that point from each of the 9 objective function plots.

at our global minimum candidate,  $\tilde{\theta}_2$ , and only vary the set of simulation draws. For each set, we therefore evaluate the objective function at exactly the same point. Without simulation error, there would be no variation across the objective function values that we obtain. To fix notation, let  $\mathcal{J}(\xi(\tilde{\theta}_2), \nu_i^m)$  denote the objective function value that we obtain at  $\tilde{\theta}_2$  using the particular set of simulation draws  $\nu_i^m$ , where we use simulation approach  $m = \{MC, MLHS\}$  and generate  $i = 1, \dots, 1000$  independent samples. We vary the number of draws between 50 and 100,000. Table 2.2 presents the results.

The spread in objective function values is striking. With only 50 Monte Carlo draws we see a range of roughly 2,500 for the objective function values. As we hold everything else constant, the different random samples of  $\nu$  are the sole driver of this effect. To assess how the variation in objective values changes with the number of simulation draws across the independently drawn samples, we report the coefficient of variation. For only 50 draws we obtain coefficients of roughly .5 and .34 for Monte Carlo and MLHS integration,

## 2 Simulation Error Causes Weak Identification

Table 2.2: Objective Function Values Obtained Using Monte Carlo and MLHS Draws

# draws	Monte Carlo draws				MLHS draws			
	$\bar{\mathcal{J}}$	$\sigma_{\mathcal{J}}$	$\sigma_{\mathcal{J}}/\bar{\mathcal{J}}$	range of $\mathcal{J}$	$\bar{\mathcal{J}}$	$\sigma_{\mathcal{J}}$	$\sigma_{\mathcal{J}}/\bar{\mathcal{J}}$	range of $\mathcal{J}$
50	726.7	360.5	.496	2,471	381.6	130.8	.343	1,002
500	286.9	70.6	.246	455.3	251.9	37.9	.151	235.7
5,000	242.6	22.3	.092	141	238.1	11.1	.047	72.3
100,000	236.9	4.8	.020	30.8	236.8	2.2	.009	14.8

Note:  $\bar{\mathcal{J}}$  and  $\sigma_{\mathcal{J}}$  denote the mean and standard deviation of the objective function values for each number of simulation draws.

respectively. Given that these figures are based on evaluating the objective function at exactly the same point, this variation is indeed substantial. As we raise the number of simulation draws, however, we can observe a large drop in the coefficients of variation. For 10,000 draws the Monte Carlo and MLHS integration approaches deliver coefficients of around .07 and .03. Raising the number of draws further to 100,000 pushes the coefficient of variation for the MLHS approach below 1 percent, while its counterpart for Monte Carlo integration is 2 percent. Finally, we can see that the mean of the objective function tends towards the same value of roughly 237 for both simulation methods.

These results show that simulation error randomly disturbs the point estimates of the structural error terms. These error terms directly enter the sample moment and thereby affect the shape of the GMM-IV objective function. Substantial simulation error then produces many local minima with widely ranging parameter estimates.

### 2.5.2 Estimation Results

We present three sets of estimation outcomes. First, we trace out how an increasing number of simulation draws affects the behavior and robustness of the BLP estimator. Second, we turn to the point estimates of the 5 random coefficients and their statistical significance. Third, we examine the model-implied economic predictions by computing the own-price elasticities at the product level and the predicted price, profit, and consumer welfare effects of a counterfactual merger between Chrysler and GM.

#### Behavior and Robustness of the Nested Fixed Point Estimator

We characterize the behavior of the estimator by examining the range and number of the identified minima. Table 2.3 shows that an increase in the number of simulation draws tightens the range and reduces the number of the identified local minima for both integration approaches. The pattern can be succinctly summarized using the coefficient of variation. Increasing the number of draws from 50 to 10,000 reduces the coefficient of

Table 2.3: Range and Spread of the Identified Minima

# draws	Monte Carlo draws				MLHS draws			
	$\bar{\mathcal{J}}$	$\sigma_{\mathcal{J}}/\bar{\mathcal{J}}$	range of $\mathcal{J}$	# Minima	$\bar{\mathcal{J}}$	$\sigma_{\mathcal{J}}/\bar{\mathcal{J}}$	range of $\mathcal{J}$	# Minima
50	198.6	.323	283	126	179.6	.361	254	149
100	195.5	.311	265	129	183.9	.302	286	136
200	199.6	.298	278	128	188.0	.240	222	132
500	209.3	.201	222	123	204.4	.154	157	111
1,000	211.5	.146	166	99	207.5	.123	148	102
2,000	214.7	.135	159	93	221.7	.078	87	65
5,000	225.1	.090	101	80	229.3	.038	53	38
10,000	230.1	.065	82	57	232.1	.029	40	32

Note: MLHS stands for modified latin hypercube sampling.  $\bar{\mathcal{J}}$  and  $\sigma_{\mathcal{J}}$  denote the mean and standard deviation of the objective function values for each number of simulation draws. To count the number of unique minima we take all identified minima from the 2,500 estimations that are run for each number of draws and round the objective function values to whole numbers.

variation for Monte Carlo integration from roughly 32 percent to 6.5 percent. For MLHS draws the decrease is more substantial from 36.1 percent to 2.9 percent. Concomitantly, the number of unique minima is reduced by a factor exceeding 2 and close to 5 for Monte Carlo and MLHS draws, respectively. The 32 unique minima that are identified using our most accurate numerical integration approach are obtained across 50 independent samples of preference draws. Thus, there is less than one minimum per set of draws. Moreover, as we compute the model aggregate shares more accurately the reduced number of minima fall into a narrowing range of values.

As a measure of the estimator's robustness we use the fraction of starting guesses that yield a local minimum. With 50 simulation draws, we see in Table 2.4 that for both simulation approaches a large fraction of estimations fails to converge to a local minimum. For Monte Carlo simulation this fraction is roughly 40 percent, while for MLHS draws almost 45 percent of the estimations fail to converge to a local minimum. With 500 simulation draws this fraction of failed estimation runs drops below 20 percent for both approaches and from 5,000 Monte Carlo draws and 2,000 MLHS draws almost every estimation run identifies a local minimum. With 10,000 draws both approaches return a local minimum for all estimation runs. Thus, with high integration accuracy the particular starting guess has no effect on whether the estimator converges to a minimum or not.

Additionally, to evaluate the computational complexity of identifying a candidate minimum we trace out how the number of simulation draws affects the number of iterations in the estimator's inner loop, the nested fixed point, and the number of objective function evaluations, the outer loop, that are required for convergence. Table 2.4 shows that

Table 2.4: Behavior of the Nested Fixed Point Estimator

# draws	Monte Carlo draws			MLHS draws		
	fraction minima	objective calls	inner iterations	fraction minima	objective calls	inner iterations
50	0.625	115.7	40.44	0.545	119.8	43.73
100	0.752	104.4	37.55	0.586	102.6	39.17
200	0.768	89.52	34.56	0.712	97.07	36.73
500	0.821	79.03	33.00	0.842	84.78	33.49
1,000	0.874	73.75	32.52	0.949	76.06	33.30
2,000	0.925	68.62	31.71	0.993	66.14	31.67
5,000	0.999	61.11	30.65	0.999	60.21	30.21
10,000	1.000	58.34	29.84	1.000	56.33	29.72

Note: MLHS stands for modified latin hypercube sampling. All statistics are computed as averages across all estimations for a given number of draws that converge to a local minimum. The number of objective calls is the number of GMM-IV objective function evaluations the optimization algorithm requires to converge to a candidate minimum.

this measure of computational complexity is roughly identical across the two simulation approaches. In terms of the number of objective function evaluations we see a substantial reduction when raising the number of draws from 50 to 10,000. The latter requires around 57 iterations, while the former needs more than 115 evaluations to arrive at a local minimum candidate. We also obtain a sizable reduction in the number of iterations in the contraction mapping from more than 40 to less than 30.

Finally, we examine how sensitive the estimation outcomes are with respect to the choice of the optimization algorithm and the choice of the inner convergence threshold for the aggregate market share inversion. The top panel of Table 2.5 shows how the choice of optimization algorithm affects the outcomes of the estimation. The results are based on running the 2,500 estimations each for 50, 5,000 and 10,000 MLHS draws with different optimization algorithms. We select one representative algorithm from three classes of optimization approaches. The Nelder-Mead algorithm falls into the category of derivative-free optimizers, the BFGS optimizer is a quasi-Newton optimizer that is derivative-based and lastly, simulated annealing belongs to the class of stochastic optimizers. For the sake of brevity, we focus on the average values of the objective function,  $\sigma_{price}$  and the estimates that are based on MLHS draws only. We obtain qualitatively identical outcomes for the remaining coefficients and the outcomes that we obtain using standard Monte Carlo draws. With 50 draws, we can see differences in the average outcomes and their empirical 95 percent confidence intervals across the optimization approaches. The estimates that we obtain with simulated annealing stand out in particular. Similar to KM, we have found that this optimization algorithm does not converge within a reasonable amount of

Table 2.5: Choice of Optimizer and Convergence Threshold

Optimizer Effect						
draws	Nelder-Mead		BFGS		Simulated Annealing	
	$\bar{\mathcal{J}}$	$\bar{\sigma}_{price}$	$\bar{\mathcal{J}}$	$\bar{\sigma}_{price}$	$\bar{\mathcal{J}}$	$\bar{\sigma}_{price}$
50	175.1 [22.53, 251.7]	2.466 [1.337, 4.228]	170.3 [24.34, 251.7]	2.662 [1.334, 5.912]	359.8 [234.9, 628.5]	0.919 [.037, 2.466]
5,000	234.4 [227.6, 244.4]	1.472 [1.327, 1.533]	233.2 [227.6, 242.0]	1.484 [1.428, 1.531]	309.7 [259.6, 405.7]	1.000 [.089, 1.866]
10,000	231.4 [225.4, 237.9]	1.460 [1.423, 1.485]	231.6 [225.4, 236.9]	1.465 [1.446, 1.484]	319.5 [253.2, 481.6]	1.03 [.243, 1.888]

Convergence Threshold Effect						
$\epsilon_{inner}$	50 MLHS draws			10,000 MLHS draws		
	$10^{-4}$	$10^{-9}$	$10^{-16}$	$10^{-4}$	$10^{-9}$	$10^{-16}$
$\bar{\mathcal{J}}$	241.4 [197.8, 311.2]	175.8 [43.88, 251.7]	170.3 [24.3, 251.7]	245.8 [224.6, 265.5]	231.6 [225.4, 236.9]	231.6 [225.4, 236.9]
$\bar{\sigma}_{price}$	1.488 [.859, 2.462]	2.453 [1.335, 4.089]	2.662 [1.334, 5.911]	1.394 [1.169, 1.922]	1.466 [1.446, 1.486]	1.465 [1.446, 1.484]

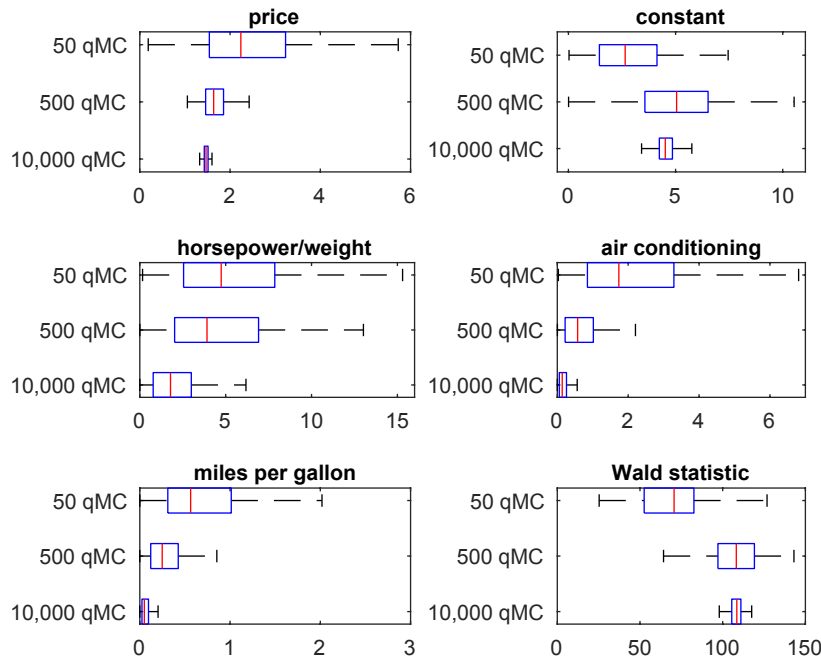
Note:  $\bar{\sigma}_{price}$  is the average of the estimated  $\sigma_{price}$  coefficients. The 2.5<sup>th</sup> and 97.5<sup>th</sup> quantiles of the outcome distributions for the objective function values and  $\sigma_{price}$  are shown in square brackets. To conserve space, we only report the outcomes from the MLHS simulation approach. The results are based on running the full 2,500 estimations each for a given number of draws. Thus, the top panel is based on a total of 22,500 estimations and the bottom panel is based on 15,000 estimations. The reported results for the simulated annealing optimizer are based on 1,000 iterations of the optimizer. As can be seen, the simulated annealing optimizer did not converge at this point.

time. For the Nelder-Mead and quasi-Newton approaches the differences in estimation outcomes turn out to be negligible for both 5,000 and 10,000 MLHS draws. Thus, with a sufficiently accurate numerical integration of the aggregate market share function, the choice of optimization algorithm becomes irrelevant in our setting.

In the bottom panel of Table 2.5 we present evidence on the role of the inner convergence threshold, which Dubé, Fox and Su (2012) demonstrate to have a major impact on the behavior of the BLP estimator. We run the 2,500 estimations each using 50 and 10,000 MLHS draws with three different inner convergence thresholds:  $10^{-4}$ , which is the loose threshold defined by Dubé, Fox and Su (2012),  $10^{-9}$  and  $10^{-16}$ . We impose the latter for all of our 40,000 estimations. With only 50 MLHS draws we indeed find that the convergence threshold of the nested fixed point has a measurable impact on the estimation outcomes. The estimates of  $\sigma_{price}$  and the identified minima of the objective differ across the three different thresholds. With 10,000 MLHS draws, however, only the very lax criterion of  $10^{-4}$  delivers results that differ markedly. The lax criterion yields a wider range for the identified minima and  $\sigma_{price}$ . The more stringent criteria of  $10^{-9}$  and  $10^{-16}$  are virtually

## 2 Simulation Error Causes Weak Identification

Figure 2.3: Range of Random Coefficient Estimates and Their Joint Statistical Significance



Note: To make changes in and around the medians of the point estimates easier, we do not plot the outliers. Moreover, to conserve space, we only show the box plots for the qMC estimations. The MC counterparts are qualitatively identical. The Wald statistic is distributed chi squared with 5 degrees of freedom. The null hypothesis is that all random coefficients are zero,  $\theta_2 = 0$ .

identical in terms of the estimation outcomes. Thus, a sufficiently high simulation accuracy also substantially diminishes the impact of the nested fixed point's convergence threshold.

### Estimated Random Coefficients and their Statistical Significance

To assess whether the estimated random coefficients are jointly statistically significant, we compute the Wald statistic for each local minimum. The null hypothesis is that the standard logit model is true, so that  $H_0 : \theta_2 = 0$ .<sup>14</sup> We do not reject the null for one out of a total of 33,479 identified minima. The evidence in favor of consumer preference heterogeneity is therefore overwhelming. This is also in line with the evidence above that our findings are not driven by weak instruments, but by the propagation of simulation error.

The boxplots in Figure 2.3 clearly show, however, that we obtain a lot of uncertainty in the random coefficient's point estimates when low numbers of draws are used to simulate

<sup>14</sup>The test statistic follows a chi squared distribution with the degrees of freedom being equal to the number of entries in  $\theta_2$ . At a 95 percent confidence level and with 5 random coefficients, the critical value for the Wald statistic is roughly 11.07.



$\nu$ . The range for the point estimates tightens drastically, however, as we move to 10,000 draws for both integration approaches. The random coefficient for price, for example, lies in a range between roughly 1.2 and 1.67 with a mean of 1.47 with 10,000 MLHS draws. The corresponding range for 50 MLHS draws is roughly 0.2 to 7.6 with a mean of 2.6. The random coefficient for the constant has a mean of 4.5 across all identified minima and also lies in a tight range. For air conditioning and miles per gallon the point estimates strongly tend toward zero. We see a similar trend for the random coefficient that is placed on the horsepower-weight ratio. Compared to the other four coefficients, however, the range of the point estimates is still quite large for this random coefficient. We would need an even higher number of draws to tighten this range further. This finding also illustrates that some random coefficients can be challenging to estimate. Nevertheless, when we examine the individual statistical significance of the random coefficients, a clear pattern emerges, which also applies to the estimated preference heterogeneity for the horsepower-weight ratio. The t-statistic for  $\sigma_{price}$  indicates that this coefficient is highly statistically significant. In fact, with 10,000 MLHS draws there are only 8 cases, where the t-statistic drops below 2.<sup>15</sup> For  $\sigma_{constant}$ , we observe a similar pattern. Out of 2,500 estimations only 207 yield a t-statistic below 1.65 and only 358 estimations produce t-statistics below 2. For each  $\sigma_{mpg}$  and  $\sigma_{air}$  there is not a single case out of 2,500 estimated minima where the t-statistic exceeds 2. For  $\sigma_{hpwt}$ , there are only 19 such instances. Thus, with sufficient integration accuracy, it turns out that only the random coefficients on price and the constant are statistically significant. The average value of the Wald statistic, however, increases with the number of simulation draws even though we are left with only two statistically significant random coefficients. The range of the Wald statistic also tightens considerably.

### Model-Implied Economic Outcomes

We assess how sensitive the model-implied economic predictions are to numerical approximation error by characterizing the distribution of own-price elasticities and the simulated effects of a merger between GM and Chrysler. Table 2.6 presents statistics on the first and second moments of the distribution of own-price elasticities. We summarize the first moment of the distribution by showing the range and mean of the average own-price elasticity. With only 50 draws, we obtain the widest range, which reaches from roughly -44 to around -2.3. Moving to 10,000 draws reduces this dispersion substantially. The mean of the average own-price elasticity increases to -8.5 and the range covers only roughly -11 to -6.8 for Monte Carlo draws and -9.6 to -7 for MLHS draws. The reduction in the standard deviation of the estimated average own-price elasticity is impressive. For the Monte Carlo and MLHS integration approaches it respectively falls from 5.3 to 0.65 and from 7.5 to only 0.32.

<sup>15</sup>We compute Eicker-White standard errors.

Table 2.6: Range of Own-Price Elasticities Using MC and qMC Integration

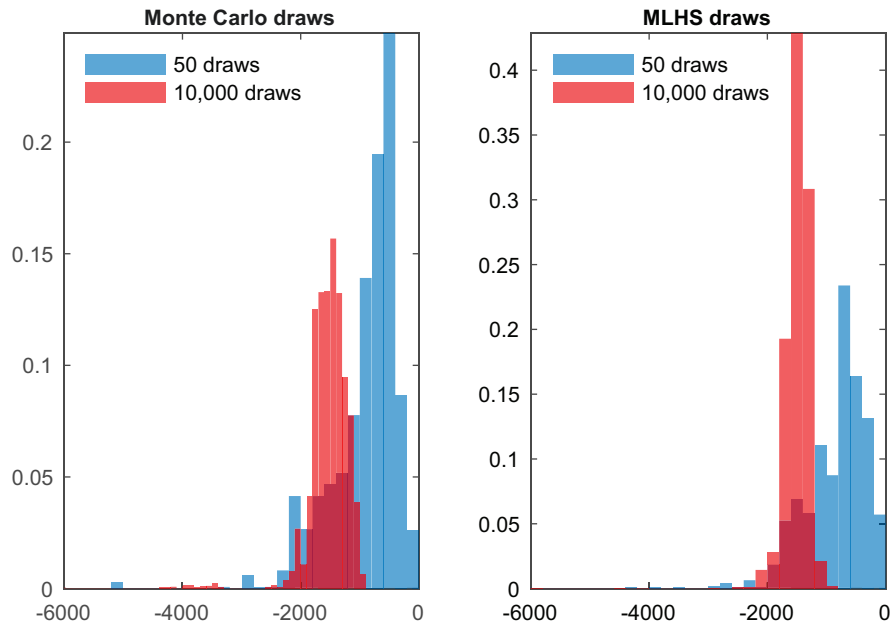
# Draws	Monte Carlo draws				
	Min.	Mean	$\bar{\eta}_{jj}$ Max.	Std. Dev.	$\eta_{jj}$ Std. Dev.
50	-43.8	-10.4	-2.31	5.32	9.97
500	-16.5	-9.13	-4.25	2.25	9.55
5,000	-11.1	-8.53	-6.14	0.84	9.12
10,000	-10.7	-8.49	-6.78	0.65	9.07
MLHS draws					
50	-43.7	-13.4	-2.36	7.47	13.0
500	-14.9	-9.47	-5.71	1.64	10.1
5,000	-9.68	-8.54	-7.33	0.41	9.15
10,000	-9.64	-8.49	-6.97	0.32	9.07

Note: MLHS stands for modified latin hypercube sampling.  $\bar{\eta}_{jj}$  denotes the average own-price elasticity. Each measure is computed across all local minima for a given number of draws. To arrive at the standard deviation of own-price elasticities for each number of draws, we average the standard deviations across all identified minima

To examine the second moment of the distribution, we compute the standard deviation of the own-price elasticities for each local minimum and average the results over all minima. Both integration approaches tend towards the same measure of the distribution's spread. The estimated standard deviation is roughly 9 when using 10,000 draws to integrate the aggregate market share function. For lower numbers of draws the spread is systematically higher.

Lastly, we perform a simulation of the equilibrium that results from a merger between GM and Chrysler. We simulate this counterfactual for each of the 20 years in the sample and average the results using units sold as weights. Figure 2.4 shows the distributions of the simulated change in consumer welfare following the merger for the two simulation approaches. The results so far are also echoed here. With only 50 draws the estimation can deliver outcomes that range from close to no detrimental effect to consumer welfare to an average annual welfare loss between 4 and 6 billion dollars. With 10,000 draws the Monte Carlo and MLHS approaches deliver a mean annual consumer welfare loss of close to 1.5 billion dollars. With 50 draws this estimate drops by roughly 40 percent to around 900 million dollars. The direction of this change is in line with how an increase in the number of simulation draws affects the estimates of own-price elasticities. We obtain consistently lower own-price elasticities for a higher number of simulation draws. Thus, demand is estimated to be overly elastic with few simulation draws. This immediately implies that the welfare losses and price changes following a merger in the market are smaller with a low integration accuracy. We surmise that this effect is driven by having

Figure 2.4: Change in Consumer Welfare



Note: The panels show the distribution of the average annual change in consumer welfare following a merger between GM and Chrysler.

sufficiently strong IV's. These effectively bound the estimates of own-price elasticities away from one. Simulation error produces a wider spread of the estimates. With a bound on own-price elasticities at one, this spread is likely to lead to an over-estimation of demand elasticities, which in turn affects the outcomes of our merger simulation. Table 2.7 shows how this biases the estimates of post-merger price and profit changes. The relatively crude approximations to aggregate market shares deliver price and profit effects that are on average too low and yield substantially wider confidence intervals. In relative terms this bias is substantial. For both simulation approaches using 10,000 draws gives an average price effect that is roughly 45 percent greater than what we obtain with only 50 draws. For the profit effect the bias is between 15 and 24 percent for the qMC and MC approach, respectively.

Simulation error produces a wider spread of economic outcomes across the different sets of minima. As with the estimates and statistical significance of the random coefficients this highlights that for low degrees of integration accuracy the specific set of simulation draws can have a substantial influence on the economic implications of BLP model estimates.

Table 2.7: Counterfactual Price and Profit Changes for the Merging Parties following a Chrysler-GM Merger

draws	Monte Carlo draws		MLHS draws	
	$\Delta p$ (percent)	$\Delta\pi$ (mln 1983 dollars)	$\Delta p$ (percent)	$\Delta\pi$ (mln 1983 dollars)
50	3.53 [1.85, 6.85]	418 [191, 640]	3.49 [1.56, 6.37]	454 [297, 683]
500	4.78 [2.61, 6.93]	503 [301, 660]	5.03 [3.21, 6.95]	526 [384, 666]
5,000	5.14 [3.81, 6.40]	521 [397, 620]	5.15 [4.22, 5.84]	524 [457, 568]
10,000	5.09 [4.17, 6.09]	519 [431, 604]	5.10 [4.56, 5.71]	521 [479, 568]

Note: The reported figures are based on simulating the GM-Chrysler merger for each of the 20 years in the sample and averaging the simulated outcomes by units sold. 95 percent confidence intervals are shown in square brackets.

## 2.6 Conclusion

The BLP model's nested fixed point estimator is indeed susceptible to numerical instabilities when the simulation error in the model's aggregate market share function is large. By substantially raising the number of simulation draws, however, the estimator's sensitivity to the specific combination of starting guess, optimization algorithm and the convergence threshold of the nested fixed point disappears. Instead, the estimator delivers an increasingly narrowing set of minima of its objective function, which also brings with it tighter sets of parameter estimates and implied economic predictions. With a larger number of simulation draws, not identifying the estimator's global minimum therefore has a more and more limited impact on the estimation outcomes. The main concern for the reliable numerical implementation of the BLP model should therefore be to select an appropriate number of simulation draws that tackles the estimator's vulnerability to simulation error.

Berry, Linton and Pakes (2004) show that in a single cross-section the BLP model's nested fixed point estimator satisfies asymptotic normality if the ratio of the number of products squared over the number of simulation draws,  $J^2/R$ , is bounded as the number of products becomes large. This asymptotic result clearly resonates with our findings and in this sense the estimator behaves as advertised. In the automobile data the number of products varies between 72 and 150 products per market with on average roughly 111 cars per year. For 10,000 draws, which bounds the ratio of  $J^2/R$  between roughly .5 and 2.3, the estimator delivers a local minimum of the objective function for every combination of starting guess and set of simulation draws.

The importance of simulation error simplifies the implementation, verification and communication of BLP model estimates relative to the guidance offered by KM and Goldberg and Hellerstein (2013). We find it unnecessary to re-estimate the model with multiple optimization algorithms once simulation error is taken seriously. This also highlights potential gains in computational efficiency. The Simplex or Nelder-Mead optimization algorithm is frequently used, because it is seen as particularly robust. We could not find any measurable difference in terms of estimation outcomes between the Nelder-Mead algorithm and a trust-region method that uses an analytical gradient. The latter approach, however, is computationally much more efficient. Similarly, the impact of the nested fixed point's convergence threshold is substantially reduced with an accurate approximation of the model's aggregate market share function. A maximum threshold of  $10^{-9}$  seems to work well for the automobile data. The loose threshold of  $10^{-4}$  should simply not be used in any setting.

We caution, however, to push our findings regarding the use of different starting guesses too far. KM have selected these 50 starting guesses after having evaluated the objective function for many more values. Thus, these guesses are likely to cover the potential parameter space well. The higher the dimensionality of the estimation problem, the more difficult it becomes to provide a good coverage of the parameter space. Therefore, all else equal, more guesses should be used for BLP models with a larger number of random coefficients. We therefore do not recommend a reduction in the number of starting guesses.<sup>16</sup> Moreover, each candidate minimum should be carefully verified. At the estimated parameter vector, the Hessian matrix must be positive definite and the norm of the gradient must be close to zero.

Dubé, Fox and Su (2012) rightly point out that the desire to speed up the estimation of BLP models confronts the researcher with the temptation to introduce approximation or simulation error. Our results highlight that giving in to this temptation is likely to backfire in the sense that it undermines the replicability and reliability of the estimation results. We realize that our findings can be viewed as raising the computational burden of BLP model estimations. Note, however, that performing relatively few estimations with many simulation draws that produce precise and reliable results are useful for answering economic questions of interest in contrast to many more estimations with substantially fewer simulation draws that give highly unstable outcomes. The researcher faces a trade-off here.

How much precision is gained by an increase in the number of simulation draws and how much does this raise the computational burden of a single model estimation? The increase in precision follows the well established Monte Carlo asymptotics that can be easily seen for  $\sigma_{\mathcal{J}}$  in Table 2.2. The standard deviation of the objective function values that is caused by simulation error changes inversely proportional with  $\sqrt{R}$ . So, raising the number of simulation draws from 50 to 10,000, a factor of 20, reduces the standard

---

<sup>16</sup>Selecting 50 starting guesses from thousands of evaluations, however, is likely unnecessary.

## 2 Simulation Error Causes Weak Identification

deviation of the objective function values roughly by a factor of 4.5. As Figure 2.5 in the Appendix indicates, the computational burden of the contraction mapping increases roughly one-for-one with the number of simulation draws. Roughly the same holds for the full estimation. We obtain an elasticity of about one for the runtime of an estimation with respect to the number of simulation draws. Thus, *ceteris paribus*, raising the number of draws by a factor of 20 raises the runtime of an estimation by the same factor. Keep in mind, however, that our change in the contraction mapping reduces the absolute runtime roughly by a factor of 2 and that a given number of MLHS draws gives approximately the same numerical precision as 4 to 5 times as many standard Monte Carlo draws. To put all of this into perspective, the average runtime of an estimation with 10,000 MLHS draws is 804 seconds, which is less than 13.5 minutes. The computational time can be further and easily reduced by realizing that each estimation can be run independently. Our 40,000 BLP estimations are in fact an “embarrassingly parallel” computational task. The speedup from parallelization almost moves one-for-one with the number of compute cores that are used. Moreover, given the extensive list of robustness checks that Knittel and Metaxoglou (2014) advise researchers to follow, comparing the computational burden of one estimation with 50 simulation draws and another with 10,000 is lopsided. Accurate numerical integration comes at the price of a higher computational burden, but rewards the researcher with reliability and therefore a substantially lessened need for extensive robustness checks. The researcher’s main concern should therefore be the reliability of the estimates and not the manageable computational burden of running the estimations.<sup>17</sup>

Finally, we re-emphasize that consistent identification requires strong and valid IV’s. Our results show, however, that simulation error can easily overwhelm the estimates of the structural error terms even when strong IV’s are used. The error propagates in the GMM-IV objective function and produces many local minima with a wide range of parameter estimates and corresponding economic implications. The accurate numerical integration of the BLP model’s aggregate market share function is therefore necessary to attain reliable identification. A high degree of numerical integration accuracy and relevant and valid IV’s are therefore complements, not substitutes.

---

<sup>17</sup>The Matlab code that we have used will be made available online and we have made available the R package `BLPestimatorR` on the CRAN repository that uses makes the same speedup of the contraction mapping.

## Appendix 2.A Computational Details

In comparison with KM, we make some modifications to the implementation of the nested fixed point estimator. These modifications affect the way the objective function is evaluated for poor parameter guesses and how the contraction mapping is implemented. KM follow the original code of Nevo (2000), which assigns very high but computable values to the objective function and analytical gradient if a specific parameter value results in numerical overflow. This issue can be easily avoided by rescaling price. We simply divide price by its standard deviation. With this rescaling we have never had to contend with overflow problems in our 40,000 BLP model estimations. Moreover, in KM’s “loose” implementation of the estimation algorithm, the convergence tolerance in the nested fixed point is dynamically adjusted. When successive iterates of (2.9) are close to each other, the convergence threshold is set to  $10^{-9}$ . If this is not the case, the threshold is set to  $10^{-6}$ . This dynamic adjustment was originally implemented by Nevo (2000) to reduce the computational burden of the estimation. Given that Dubé, Fox and Su (2012) show that a loose convergence threshold is an additional source of numerical error and given that computational power has increased dramatically over the last two decades, we enforce a uniform convergence threshold of  $\epsilon_{inner} = 10^{-16}$  throughout. Lastly, we speed up the contraction mapping by avoiding a large number of numerical divisions, which are computationally costly (see the next section for details). The computational burden is roughly reduced by half with this change. Our reformulation is equivalent to (2.9) exponentiated, as in Nevo (2000), and therefore produces the same outcomes. We do not introduce an additional source of approximation error with this modification.

### 2.A.1 Verifying Candidate Minima

We use two criteria to assess whether the output of the optimization algorithm delivers a minimum. First, it must be the case that all the eigenvalues of the Hessian at the estimated coefficient vector,  $\hat{\theta}_2$ , are strictly positive. Second, the gradient must be sufficiently close to zero. The definition of sufficiently close to zero is arbitrary to some extent. We adopt a cutoff of 0.1 for the Euclidean norm of the gradient at  $\hat{\theta}_2$ . Our qualitative results are robust to either tightening or relaxing this cutoff. This cutoff is substantially more stringent than the cutoff of 30 that is adopted by KM.

## Appendix 2.B Speeding up the Contraction Mapping

The BLP contraction mapping is given by (2.9) and we restate it here for convenience.

$$\delta_j^{iter+1} = \delta_j^{iter} + \log(S_j) - \log(s_j(\delta^{iter}, \nu, \theta_2)) \quad (2.15)$$

In contrast to the main text and without loss of generality, we switch to the product level. Nevo (2000) notes in his Appendix that taking logs is a computationally costly operation and that the computational burden of repeatedly solving the fixed point during the estimation can be reduced by exponentiating the equation.

$$w_j^{iter+1} = w_j^{iter} \frac{S_j}{s_j(w^{iter}, \nu, \theta_2)} \quad (2.16)$$

Here,  $w \equiv \exp(\delta_j)$  is the exponential of the mean utility vector. This reformulation gives a substantial speedup in computing the contraction mapping. We can improve upon this further by noting that the contraction mapping can be formulated in terms of consumer-specific choice probabilities for the outside option. Let  $v_{ij} \equiv \exp(\mu_{ij})$ .

$$w_j^{iter+1} = w_j^{iter} \frac{S_j}{\frac{1}{I} \sum_i \frac{w_j^{iter} v_{ij}}{1 + \sum_k w_k^{iter} v_{ik}}} = \frac{S_j}{\frac{1}{I} \sum_i v_{ij} s_{io}(w^{iter}, \nu, \theta_2)} \quad (2.17)$$

The reason why this reformulation presents an even lower computational burden than the formulation by Nevo (2000) is that it substantially reduces the number of divisions that have to be performed and divisions are a computationally costly operation. The main driver for these savings are that instead of computing  $J * I$  consumer-specific choice probabilities, we only compute  $I$  consumer-specific choice probabilities for the outside option. When the  $s_{ij}$ 's are computed, one big matrix division is performed. To see this, write out the matrix that gives all consumer-specific choice probabilities in the market.

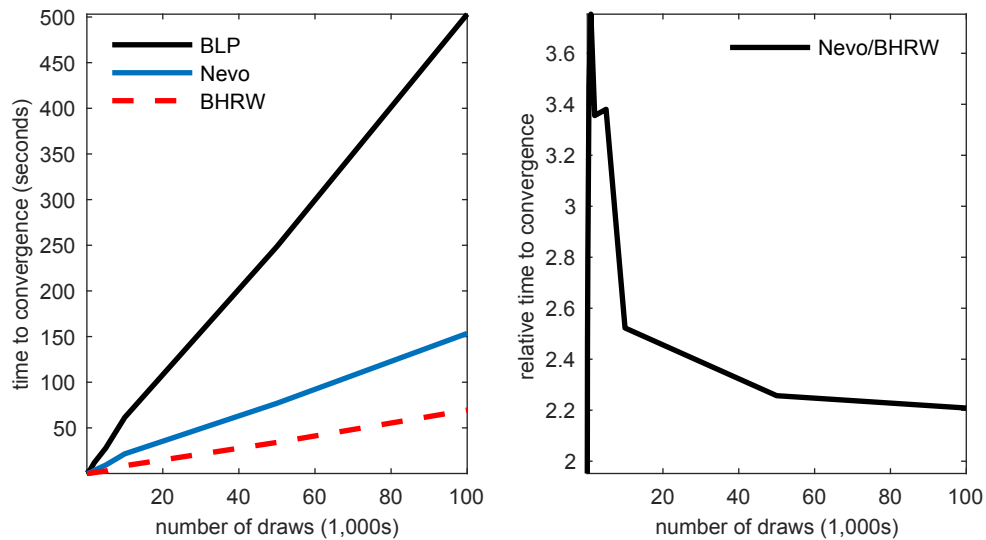
$$[s_{ij}(\delta, \nu, \theta_2)]_{i=1, \dots, I; j=1, \dots, J} = \frac{\exp(\delta * \iota'_I + [\mu_j(\nu, \theta_2)])}{\iota_J * (1 + \sum_k \exp(\delta_k * \iota'_I + [\mu_k(\nu, \theta_2)])} \quad (2.18)$$

$\iota_I$  is a vector of ones with  $I$  elements. It has the effect of stacking the vector of mean utilities horizontally  $I$  times. In the numerator  $\iota_J$  stacks the denominator vertically  $J$  times. This makes the numerator and denominator conformable and the  $s_{ij}$ 's for a whole market can be computed in one matrix operation. We are dividing a  $J \times I$  matrix by another  $J \times I$  matrix, which requires  $J * I$  divisions.

The vector of consumer-specific choice probabilities for the outside option on the other hand is given by  $[s_{io}(\delta, \nu, \theta_2)]_{i=1, \dots, I} = 1 / [(1 + \sum_k \exp(\delta_k * \iota'_I + \mu_k(\nu, \theta_2)))]$  and only requires  $I$  divisions. Unless there is only one inside product in the market, the computational burden for our reformulation in terms of the consumer-specific outside good choice probabilities



Figure 2.5: Computational Burden of Solving the Fixed Point



Note: The left panel shows the average time to reach the convergence threshold  $10^{-14}$  in the contraction mapping over KM's 50 random starting guesses for the random coefficients. The black line shows the outcomes for the original BLP contraction mapping, while the solid blue line corresponds to the fixed point formulation of Nevo (2000) and the dashed red line shows our reformulation. We evaluate 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000, 50,000 and 100,000 draws using 10 independently generated samples and average the time until convergence across these 10 sets. The right panel plots the ratio of the time to convergence for the Nevo and our fixed point formulation.

has a strictly lower computational burden. Figure 2.5 illustrates this. For the BLP car data, we solve the nested fixed point for each of KM's 50 starting guesses. We do this for numbers of draws between 50 and 100,000. The solid blue line plots the time to convergence required by the Nevo version of the fixed point, while the dashed red line corresponds to our version of the same fixed point problem. We want to emphasize that both versions need exactly the same number of iterations to reach the convergence threshold and give exactly the same  $\delta^*$  for all of the starting guesses. Formulating the contraction mapping in terms of the outside good shares yields a speedup of at least 2. This matches the speedup of the approximate BLP estimator of Lee and Seo (2015), which uses a linear approximation of the market share equation to solve the fixed point problem analytically. We attain roughly the same speedup, but solve the fixed point problem exactly and thereby retain all the properties of the original BLP estimator without introducing an additional source of approximation error that propagates in the estimator's objective function.<sup>18</sup>

<sup>18</sup>The precision with which the fixed point is solved is of course limited by the convergence threshold.

## Appendix 2.C Documentation of the R Package BLPestimatorR

### Package ‘BLPestimatorR’

May 5, 2017

**Type** Package

**Title** Performs a BLP Demand Estimation

**Version** 0.1.4

**Author** Daniel Brunner (aut), Constantin Weiser (ctr), Andre Romahn (ctr)

**Maintainer** Daniel Brunner <daniel.brunner@hhu.de>

**Description**

Provides the estimation algorithm to perform the demand estimation described in Berry, Levinsohn and Pakes (1995) <DOI:10.2307/2171802> . The routine uses analytic gradients and offers a large number of implemented integration methods and optimization routines.

**License** GPL-3

**LazyData** TRUE

**Depends** Rcpp (>= 0.12.7), mvQuad, ucmintf, numDeriv, randtoolbox

**LinkingTo** Rcpp

**RoxygenNote** 6.0.1

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2017-05-05 17:43:33 UTC

#### R topics documented:

BLPestimatorR-package . . . . .	2
estimateBLP . . . . .	3
get.BLP.dataset . . . . .	6

<b>Index</b>	<b>9</b>
--------------	----------

---

*BLPestimatorR*-package *BLP demand estimation for differentiated products*

---

**Description**

Provides the estimation algorithm to perform the demand estimation described in Berry, Levinsohn and Pakes (1995) <DOI:10.2307/2171802>. The routine uses analytic gradients and offers a large number of implemented integration methods and optimization routines.

**Details**

Package: *BLPestimatorR*  
 Type: Package  
 Version: 0.1.4  
 Date: 2017-05-05  
 License: GPL-3

**Author(s)**

Daniel Brunner (HHU of Duesseldorf / Germany)  
 Constantin Weiser (HHU of Duesseldorf / Germany)  
 Andre Romahn (HHU of Duesseldorf / Germany)  
 Maintainer: Daniel Brunner <daniel.brunner@hhu.de>

**References**

Steven Berry, James Levinsohn, Ariel Pakes (1995): Automobile Prices in Market Equilibrium <DOI:10.2307/2171802>  
 Christopher R. Knittel, Konstantinos Metaxoglou (2014): Estimation of Random-Coefficient Demand Models: Two Empiricists Perspective <DOI:10.1162/REST\_a\_00394>

**Examples**

```
# Parameters
i<-1
K<-2
Xlin_example <- c("price", "x1", "x2", "x3", "x4", "x5")
Xexo_example <- c("x1", "x2", "x3", "x4", "x5")
Xrandom_example <- paste0("x",1:K)
instruments_example <- paste0("iv",1:10)

# Data generation
BLP_data <- get.BLP.dataset(nmkt = 25, nbrn = 20,
```

*estimateBLP*

3

```

Xlin = Xlin_example,
Xexo = Xexo_example,
Xrandom = Xrandom_example,
instruments = instruments_example,
true.parameters = list(Xlin.true.except.price = rep(0.2,5),
                       Xlin.true.price = -0.2, Xrandom.true = rep(2,K),
                       instrument.effects = rep(2,10),
                       instrument.Xexo.effects = rep(1,5)),
price.endogeneity = list( mean.xi = -2,
                           mean.eita = 0,
                           cov = cbind( c(1,0.7), c(0.7,1))),
printlevel = 0, seed = 5326 )

# Estimation
BLP_est<- estimateBLP(Xlin = Xlin_example,
                     Xrandom = Xrandom_example,
                     Xexo = Xexo_example,
                     instruments = instruments_example,
                     shares = "shares",
                     cdid = "cdid",
                     productData = BLP_data,
                     starting.guesses.theta2 = rep(1,K),
                     solver.control = list(maxeval = 5000),
                     solver.method = "BFGS_matlab",

                     starting.guesses.delta = rep(1, length(BLP_data$cdid)),
                     blp.control = list(inner.tol = 1e-6,
                                         inner.maxit = 5000),
                     integration.control= list( method="MLHS",
                                                amountNodes= 100,
                                                seed= 3 ),
                     postEstimation.control= list(standardError = "robust",
                                                  extremumCheck = TRUE,
                                                  elasticities = "price"),

                     printLevel = 2)

# Show results
summary(BLP_est)

```

---

*estimateBLP*                      *Performs a BLP demand estimation.*

---

**Description**  
 Performs a BLP demand estimation.

**Usage**  
*estimateBLP*(Xlin, Xexo, Xrandom, instruments, demographics, shares, cdid,  
 productData, demographicData, starting.guesses.theta2, starting.guesses.delta,

4

*estimateBLP*

```

solver.method = "BFGS", solver.control = list(), blp.control = list(),
integration.control = list(), postEstimation.control = list(),
printLevel = 0)

```

**Arguments**

<code>Xlin</code>	character vector specifying the set of linear variables (variable names must be included in <code>productData</code> )
<code>Xexo</code>	character vector specifying the set of exogenous variables (subset of <code>Xlin</code> )
<code>Xrandom</code>	character vector specifying the set of random coefficients (variable names must be included in <code>productData</code> )
<code>instruments</code>	character vector specifying the set of instrumental variables (variable names must be included in <code>productData</code> )
<code>demographics</code>	optional: character vector specifying the set of demographic variables (must be included as list entries in <code>demographicData</code> )
<code>shares</code>	character vector specifying observed market shares (variable name must be included in <code>productData</code> )
<code>cdid</code>	character vector specifying the market identifier (variable name must be included in <code>productData</code> )
<code>productData</code>	dataframe with product characteristics
<code>demographicData</code>	optional: list with demographic data for each market (see details)
<code>starting.guesses.theta2</code>	matrix with starting values for the optimization routine (NA entries indicate the exclusion from estimation, i.e. the coefficient is assumed to be zero)
<code>starting.guesses.delta</code>	optional: numeric vector with starting guesses for the mean utility
<code>solver.method</code>	specifies the solver method in <code>optim</code> or <code>ucminf</code>
<code>solver.control</code>	list of additional arguments for the optimization routines: <code>solver.reltol</code> tolerance for the optimization routine <code>solver.maxit</code> maximum iterations for the optimization routine ... further arguments passed to <code>optim</code> or <code>ucminf</code>
<code>blp.control</code>	list of additional arguments for the BLP algorithm: <code>inner.tol</code> tolerance for the contraction mapping <code>inner.maxit</code> maximum iterations for the contraction mapping
<code>integration.control</code>	list of parameters for the BLP integration problem: <code>method</code> integration method <code>amountNodes</code> integration accuracy for Monte Carlo based integration <code>accuracyQuad</code> integration accuracy for integration by quadrature rules <code>seed</code> seed for the draws of Monte Carlo based integration <code>nodes</code> set of manually provided integration nodes <code>weights</code> set of manually provided integration weights

*estimateBLP*

5

output if TRUE, integration nodes and individual shares (sij) are available as output

postEstimation.control list of parameters specifying post estimation results:

standardError chose robust (default) or nonRobust

extremumCheck if TRUE (default), second derivatives are checked for the existence of minimum at the point estimate

elasticities character vector specifying the set of variables elasticities are calculated for

printLevel level of output information ranges from 1 (no GMM results) to 4 (every norm in the contraction mapping)

**Details**

The optimization routines are included in the packages `optim` and `ucminf`. Only gradient based methods are supported. The `ucminf` clones MATLABs' standard trust region optimization routine, which turns out to be effective in avoiding overflow problems in the BLP model. Valid arguments are `BFGS`, `BFGS_matlab`, `L-BFGS-B` or `CG`.

For solver options, the use of `solver.maxit` and `solver.rel` is recommended. For conflicts of `solver.maxit` and `solver.reltol` and arguments of the respective solvers, priority is given to the latter. Make sure that additionally provided solver control arguments are valid.

For demographics variables, list entries of `demographicData` must be named according to `demographics`. Each list entry contains a dataframe with the draws for different markets and a variable (according to the `cdid` argument) that allows to match the draws with the markets.

The logit model is used for elasticity calculation, if the variable of interest is not included in `Xrandom`. Columns of the elasticity matrix contain the variables that are changed by 1%, and rows contain effects on other products in the choice set.

**Value**

Returns an object of class `'blp'`. This object contains, among others, all estimates for preference parameters and standard errors. Necessary information for further post estimation analysis can be included as well.

**Examples**

```
# Parameters
i<-1
K<-2
Xlin_example <- c("price", "x1", "x2", "x3", "x4", "x5")
Xexo_example <- c("x1", "x2", "x3", "x4", "x5")
Xrandom_example <- paste0("x",1:K)
instruments_example <- paste0("iv",1:10)

# Data generation
BLP_data <- get.BLP.dataset(nmkt = 25, nbrn = 20,
                           Xlin = Xlin_example,
                           Xexo = Xexo_example,
```

6

*get.BLP.dataset*

```

Xrandom = Xrandom_example,
instruments = instruments_example,
true.parameters = list(Xlin.true.except.price = rep(0.2,5),
                       Xlin.true.price = -0.2, Xrandom.true = rep(2,K),
                       instrument.effects = rep(2,10),
                       instrument.Xexo.effects = rep(1,5)),
price.endogeneity = list( mean.xi = -2,
                          mean.eita = 0,
                          cov = cbind( c(1,0.7), c(0.7,1))),
printlevel = 0, seed = 5326 )

# Estimation
BLP_est<- estimateBLP(Xlin = Xlin_example,
                    Xrandom = Xrandom_example,
                    Xexo = Xexo_example,
                    instruments = instruments_example,
                    shares = "shares",
                    cdid = "cdid",
                    productData = BLP_data,
                    starting.guesses.theta2 = rep(1,K),
                    solver.control = list(maxeval = 5000),
                    solver.method = "BFGS_matlab",

                    starting.guesses.delta = rep(1, length(BLP_data$cdid)),
                    blp.control = list(inner.tol = 1e-6,
                                       inner.maxit = 5000),
                    integration.control= list( method="MLHS",
                                               amountNodes= 100,
                                               seed= 3 ),
                    postEstimation.control= list(standardError = "robust",
                                                  extremumCheck = TRUE,
                                                  elasticities = "price"),

                    printLevel = 2)

# Show results
summary(BLP_est)

```

---

*get.BLP.dataset*      *This function creates a simulated BLP dataset.*

---

**Description**

This function creates a simulated BLP dataset.

**Usage**

```

get.BLP.dataset(nmkt, nbrn, Xlin, Xexo, Xrandom, instruments,
               true.parameters = list(), price.endogeneity = list(), printlevel = 1,
               seed)

```

*get.BLP.dataset*

7

**Arguments**

<code>nmkt</code>	number of markets
<code>nbrn</code>	number of products
<code>Xlin</code>	character vector specifying the set of linear variables
<code>Xexo</code>	character vector specifying the set of exogenous variables (subset of <code>Xlin</code> )
<code>Xrandom</code>	character vector specifying the set of random coefficients (subset of <code>Xlin</code> )
<code>instruments</code>	character vector specifying the set of instrumental variables
<code>true.parameters</code>	list with parameters of the DGP
	<code>Xlin.true.except.price</code> "true" linear coefficients in utility function except price
	<code>Xlin.true.price</code> "true" linear price coefficient in utility function
	<code>Xrandom.true</code> "true" set of random coefficients
	<code>instrument.effects</code> "true" coefficients of instrumental variables to explain endogenous price
	<code>instrument.Xexo.effects</code> "true" coefficients of exogenous variables to explain endogenous price
<code>price.endogeneity</code>	list with arguments of the multivariate normal distribution
	<code>mean.xi</code> controls for the mean of the error term in the utility function
	<code>mean.eita</code> controls for the mean of the error term in the price function
	<code>cov</code> controls for the covariance of <code>xi</code> and <code>eita</code>
<code>printlevel</code>	0 (no output) 1 (summary of generated data)
<code>seed</code>	seed for the random number generator

**Details**

The dataset is balanced, so every market has the same amount of products. Only unobserved heterogeneity can be considered. Variables that enter the equation as a Random Coefficient or exogenously must be included in the set of linear variables. The `parameter.list` argument specifies the "true" effect on the individual utility for each component. Prices are generated endogenous as a function of exogenous variables and instruments, where the respective effect sizes are specified in `instrument.effects` and `instrument.Xexo.effects`. Error terms `xi` and `eita` are drawn from a multivariate normal distribution, whose parameters can be set in `price.endogeneity`. Market shares are generated by MLHS integration rule with 10000 nodes.

**Value**

Returns a simulated BLP dataset.



**Examples**

```
K<-2 #number of random coefficients
Xlin_example <- c("price", "x1", "x2", "x3", "x4", "x5")
Xexo_example <- c("x1", "x2", "x3", "x4", "x5")
Xrandom_example <- paste0("x",1:K)
instruments_example <- paste0("iv",1:10)
data <- get.BLP.dataset(nmkt = 25,
  nbrn = 20,
  Xlin = Xlin_example,
  Xexo = Xexo_example,
  Xrandom = Xrandom_example,
  instruments = instruments_example,
  true.parameters = list(Xlin.true.except.price = rep(0.2,5),
    Xlin.true.price = -0.2,
    Xrandom.true = rep(2,K),
    instrument.effects = rep(2,10),
    instrument.Xexo.effects = rep(1,5)),
  price.endogeneity = list( mean.xi = -2,
    mean.eita = 0,
    cov = cbind( c(1,0.7), c(0.7,1))),
  printlevel = 0, seed = 234234 )
```

## Index

\*Topic **math**  
BLPestimator-package, 2

\*Topic **package**  
BLPestimator-package, 2

BLPestimator (BLPestimator-package), 2

BLPestimator-package, 2

estimateBLP, 3

get.BLP.dataset, 6

## Contributions of Daniel Brunner

- Development of the R package “BLPestimatorR” that conveniently allows to apply the results of this paper (appendix 2C)
- Running pretests on the performance of different numerical integration and optimization methods
- Apply the formal results of Berry, Linton and Pakes (2004) to the setting of this paper
- Code preparation for parallel computations at the Hilbert cluster
- Robustness checks with simulated data
- Editorial contributions

---

Florian Heiss

André Romahn

Constantin Weiser



# 3 Implications of Adaptive Integration Rules for the Performance of Random Coefficient Models of Demand

## 3.1 Introduction

The model of Berry, Levinsohn and Pakes has attracted much attention in demand estimation due to the consideration of endogenous prices and consumer heterogeneity in the presence of aggregated data. Models that explicitly allow for consumer heterogeneity often produce more reliable conclusions (Nevo, 2000) and in fact, many important economic questions have been answered by their model (for an overview, see table 1 in Knittel and Metaxoglou (2014)). This, however, comes at the cost of approximating non-analytic integrals to compute implied market shares, which introduces an additional error that propagates to parameter accuracy and computational cost. The paper contributes by discussing adaptive integration rules as a remedy for adverse effects of inaccurate market share approximations.

The approximation of market shares gives an important numerical error that propagates through the estimation algorithm. It is a fundamental step of the estimation algorithm and takes place hundreds or thousands of times for different sets of parameters. The estimation of a typical simulated dataset in section 3.4, for example, includes 400,000 market share approximations. Simulation and real data results show a huge variation in parameter estimates between different inaccurate integration rules.

Given this leverage, the attempts to reduce the approximation error are well-founded. An easy and obvious way of dealing with this problem is to use a very large number of function evaluations<sup>1</sup> for the integration problem. Although the approximation error converges to zero for an infinite amount of draws, in practice, computational and memory constraints limit the amount of draws to a given level. Therefore, another approach builds on increasing the efficiency of integration rules, i.e. to get more precise market share approximations with a “better” integration rule while using the same amount of draws<sup>2</sup>.

---

<sup>1</sup>The arguments of this function are called draws or nodes. The terms draws and nodes are used interchangeably in the remainder of this paper.

<sup>2</sup>As explained later, it is not necessary that a more efficient integration rule can produce more precise results with the same computational cost.

### 3 Implications of Adaptive Integration Rules

The comparison between widely used Monte Carlo and more efficient quasi Monte Carlo techniques in the simulation study, for example, shows much lower variation in point estimates for the latter. Efficient quadrature rules, like the sparse grids approach, are another option that is also considered in this paper. Monte Carlo, quasi Monte Carlo and sparse grids are labeled as standard approaches in the following.

Adaptive integration rules adjust the set of draws of a standard integration rule to the integrand to attain a *maximum* numerical efficiency. This paper uses an importance sampling technique for both, quasi Monte Carlo and sparse grids, to perform the adjustment. In cases where the relevant part of the integrand is limited to a small subset of the integration variable, adaptive rules produce much more precise approximations than standard approaches. In fact, unfortunate parameter combinations can produce integration problems where standard integration rules, even with the maximum amount of feasible draws, are incapable of reducing the approximation error to an acceptable level. In nearly all simulation and real data settings impressive accuracy results for a given market share integral are attained.

The efficiency gain of adaptive integration rules comes at the price of additional computational cost<sup>3</sup>. The process of adjusting a standard integration rule gives the first component of additional cost, which turns out to be insignificant compared to the overall computation time of the estimator. The second component stems from using different sets of draws for different market share integrals. The standard approach in the BLP estimator is to reuse the same draws in a given market, which saves a lot of function evaluations. Unfortunately, this is not in line with a tailored integration rule and increases the computational cost for adaptive integration. More precisely, computation time and memory usage increase quadratically in the amount of considered products for adaptive rules compared to a linear increase for the standard approaches. Especially in situations with many products per market, this is a clear drawback of adaptive integration rules.

To address problems with many markets, adaptive rules are not only applied for each market share integral (product level), but at an aggregated level as well (market level). The latter uses the same parsimonious standard BLP approach with an adapted integration rule for all products in one market. However, much of the possible efficiency gain can be lost, if integrands are heterogeneous. Simulation results confirm a better performance than standard integration approaches.

This paper is structured as follows: Section 3.2 introduces the BLP model and section 3.3 gives an overview of the numerical integration approaches that are used to approximate the market share integrals. Section 3.4 evaluates performance differences of the integration methods in terms of estimates' precision and computational cost in a simulation study with

---

<sup>3</sup>Computational cost comprises the amount of necessary runtime and RAM space in the following.

multiple generated BLP datasets and different dimensions of the integration problem<sup>4</sup>. Proposed integration rules are illustrated in section 3.5 with the well-known cereal example of Nevo (2000). Finally, results are summarized.

## 3.2 The Model

The ability of predicting realistic substitution patterns and the handling of endogenous variables by a standard instrumental variable approach in a nonlinear choice model helped the BLP model (Berry, Levinsohn, and Pakes (1995) and Berry (1994)) to become the workhorse in estimating models of product differentiation. This becomes possible by combining the fortunate properties of a mixed multinomial logit discrete choice model with an efficient way of calculating a large amount of parameters in a nested fixed point algorithm.

The BLP model has been the subject of intensive research in different aspects. Econometric problems, for example, range from extensions to micro data (Berry, Levinsohn, and Pakes, 2004) to the optimal choice of instruments (e.g. Reynaert and Verboven, 2014). This area also comprises asymptotic analyses with a growing amount of products (Berry, Linton, and Pakes, 2004) or markets (Freyberger, 2015). Both authors highlight the effect of large stochastic integration errors of the market share integrals on the asymptotic distribution of the estimated parameters. The integration error is also of particular importance when investigating the numerical properties of the BLP model. Skrainka and Judd (2011) show that inaccurate integration rules lead to a numerical instability of inner and outer loop behavior and wrong standard errors. Other important contributions have been made by Knittel and Metaxoglou (2014) showing that different optimization routines give different answers to the same optimization problem, or Dubé, Fox, and Su (2012) examining the effect of error tolerances in the model.

### 3.2.1 Demand Side

Identifying the drivers of market shares starts with a utility maximizing agent. Assuming that agent  $i$  in market  $t$  consuming product  $j$  experiences a utility of

$$u_{ijt} = U(\mathbf{x}_{jt}, \xi_{jt}, p_{jt}, \tau_i, \boldsymbol{\theta}) ,$$

the consumer will choose the product with the highest utility out of  $J_t$  products<sup>5</sup>. This depends on a vector of  $K$  product characteristics  $\mathbf{x}_{jt}$ , which are observed by the econometrician and the consumer, product characteristics  $\xi_{jt}$ , which are exclusively observed by

---

<sup>4</sup>In the BLP model, the dimensionality of the integration problem is determined by the amount of random coefficients.

<sup>5</sup>The definition of a market  $t$  depends on the context of the study and can be very broad ranging from different points in time to geographical differences.

### 3 Implications of Adaptive Integration Rules

the consumer and the price  $p_{jt}$ . Moreover,  $\tau_i$  denotes all consumer specific variables on utility. The set  $\boldsymbol{\theta}$  collects all variables' effects.

Without loss of generality, this paper imposes a linear functional form without consumer specific variables and focuses on unobserved heterogeneity:

$$u_{ijt} = \alpha p_{jt} + \mathbf{x}_{jt} \boldsymbol{\beta}_i + \xi_{jt} + \epsilon_{ijt} \quad . \quad (3.1)$$

$\epsilon_{ijt}$  is assumed to be i.i.d. and follows a Type I Extreme Value distribution. Moreover, agent  $i$  can choose an outside option, where the consumer obtains a utility of  $u_{i0t} = \epsilon_{i0t}$  (i.e. all parameters are normalized to zero). The random coefficient  $\boldsymbol{\beta}_i$  is a vector of dimension  $K$  with individual specific coefficients and follows a multivariate normal distribution. Without consumer specific effects, substitution patterns from one to another product resulting from a change in the environment would only depend on their market shares. This unfortunate pattern is referred to as the irrelevance of independent alternatives. For notational simplicity, this paper uses individual specific effects for every product characteristic except price, so the dimension of the integration problem equals  $K$ .

The effect  $\beta_{i,k}$  can be split in a constant and individual part, i.e.  $\beta_{i,k} = \bar{\beta}_k + \sigma_k \nu_{i,k}$  for characteristic  $k$  with

$$\beta_{i,k} \sim N(\bar{\beta}_k, \sigma_k^2) \quad \text{and} \quad \nu_{i,k} \sim N(0, 1) \quad \text{for} \quad k = 1, \dots, K \quad .$$

For the sake of simplicity, the covariances between  $K$  random variables are assumed to be zero, so there is no need to use matrix algebra for the distribution of  $\boldsymbol{\beta}_i$ .

By separating a constant and individual part, equation (3.1) can be reformulated as

$$u_{ijt} = \underbrace{\alpha p_{jt} + \mathbf{x}_{jt} \bar{\boldsymbol{\beta}} + \xi_{jt}}_{\delta_{jt}(\xi_{jt}, \boldsymbol{\theta}_1)} + \sum_{k=1}^K x_{jt,k} \sigma_k \nu_{i,k} + \epsilon_{ijt} \quad . \quad (3.2)$$

Typically,  $\boldsymbol{\theta}_1$  includes linear effects  $\alpha$  and  $\bar{\boldsymbol{\beta}}$ , while  $\boldsymbol{\theta}_2$  denotes all standard deviations  $\sigma_k$ . The constant part of utility across consumers for given a market  $t$  and an alternative  $j$  is labeled  $\delta_{jt}$ . It is the sum term in (3.2) that allows to correlate the maximum utility with a utility of a product with similar characteristics. For  $\sigma_k \neq 0$ , consumers are more likely to switch to a product  $j$  with a similar  $x_{j,k}$ , which often results in more realistic substitution patterns than predicted by a logit model.

According to the logit formula, the probability for consumer  $i$  choosing  $j$  given  $\delta_{jt}$  and  $\boldsymbol{\theta}_2$  is:

$$s_{ijt}(\delta_{jt}, \boldsymbol{\theta}_2) = \frac{\exp\left(\delta_{jt} + \sum_{k=1}^K x_{jt,k} \sigma_k \nu_{i,k}\right)}{1 + \sum_{l=1}^{J_t} \exp\left(\delta_{lt} + \sum_{k=1}^K x_{lt,k} \sigma_k \nu_{i,k}\right)} \quad . \quad (3.3)$$

In a market where all consumers are like consumer  $i$ ,  $s_{ijt}$  would be the market share of



product  $j$ . Since consumers are allowed to be different, the aggregate of individual market shares is calculated as

$$s_{jt}(\delta_{jt}, \boldsymbol{\theta}_2) = \int_{\mathbb{R}^K} \frac{\exp\left(\delta_{jt} + \sum_{k=1}^K x_{jt,k} \sigma_k \nu_{i,k}\right)}{1 + \sum_{l=1}^{J_t} \exp\left(\delta_{lt} + \sum_{k=1}^K x_{lt,k} \sigma_k \nu_{i,k}\right)} \phi(\boldsymbol{\nu}) d\boldsymbol{\nu} . \quad (3.4)$$

This integral has no analytic solution and needs to be approximated by methods discussed in section 3.3.

### 3.2.2 Estimation Algorithm

The identifying assumption is that the unobserved term  $\xi_{jt}$  is mean independent of exogenous product characteristics<sup>6</sup>,  $\mathbf{x}_{jt}$ , and cost shifters,  $\mathbf{c}_{jt}$ :

$$E(\xi_{jt} | \mathbf{z}_{jt}) = 0 ,$$

with  $\mathbf{z}_{jt} = \{\mathbf{x}_{jt}, \mathbf{c}_{jt}\}$  and  $\mathbf{z}_{jt} \in \mathbf{Z}$ , where  $\mathbf{Z}$  defines the set of all instrumental variables.

The set of parameters  $\hat{\boldsymbol{\theta}}_2$  is obtained by equalizing sample moments to population moments in an GMM approach as suggested by Hansen (1982). This results in the following optimization problem:

$$\hat{\boldsymbol{\theta}}_2 = \operatorname{argmin} \xi(\boldsymbol{\theta}_2)' \mathbf{Z} \mathbf{W} \mathbf{Z}' \xi(\boldsymbol{\theta}_2) , \quad (3.5)$$

with the weighting matrix  $\mathbf{W}$ .

For a given  $\boldsymbol{\theta}_2$ , the unobservable term  $\xi_{jt}$  is defined as the following difference:

$$\xi_{jt}(\boldsymbol{\theta}_2) = \delta_{jt} - \alpha p_{jt} - \mathbf{x}_{jt} \bar{\boldsymbol{\beta}} .$$

The calculation of  $\xi_{jt}$  becomes possible by combining a contraction mapping with an IV approach. As pointed out by Berry (1994), under very general conditions there exists a unique  $\delta_{jt}$  that equals predicted and observed market shares. BLP propose a contraction mapping to exclude the estimation of  $\delta_{jt}$  from the minimization problem by equating  $s_{jt}(\delta_{jt}, \boldsymbol{\theta}_2)$  and  $s_{jt}^{\text{obs}}$ :

$$\delta_{jt}^{n+1} = \delta_{jt}^n + \ln \left( \frac{s_{jt}^{\text{obs}}}{s_{jt}(\delta_{jt}^n, \boldsymbol{\theta}_2)} \right) . \quad (3.6)$$

This is done until the log term falls below a predefined tolerance. A common starting value is  $\delta_{jt}^{n=0} = 0$ . Note that  $\boldsymbol{\theta}_2$  needs to be given in order to calculate  $s_{jt}(\delta_{jt}, \boldsymbol{\theta}_2)$ .

The final value of  $\delta_{jt}$  is regressed on  $p_{jt}$  and  $\mathbf{x}_{jt}$  with  $\mathbf{Z}$  as instruments, which gives consistent estimates for  $\alpha$  and  $\bar{\boldsymbol{\beta}}$ .

<sup>6</sup>Note that this notation requires all  $\mathbf{x}_{jt}$  to be exogenous. If this is not the case for any  $x_{jt} \in \mathbf{x}_{jt}$ ,  $x_{jt}$  would be treated as the endogenous price variable in the following.

### 3 Implications of Adaptive Integration Rules

Parameters  $\theta_2$  are varied by an optimization routine, which is called the outer loop. In each step of this optimization,  $\delta_{jt}$  is obtained by the contraction mapping in an inner loop. Eventually, the resulting set  $\hat{\theta}_2$  minimizes equation (3.5).

## 3.3 Market Share Integration

The market share integral in equation (3.4) gives the following general integration problem:

$$A = \int_{\mathbb{R}^K} f(\boldsymbol{\nu})\phi(\boldsymbol{\nu})d\boldsymbol{\nu} . \quad (3.7)$$

The  $K$ -dimensional vector of random variables  $\boldsymbol{\nu} \in \mathbb{R}^K$  is evaluated by  $f(\boldsymbol{\nu})$ .  $f(\boldsymbol{\nu})$  represents the consumer specific probability of buying a specific product, so it is bounded between zero and one.

### 3.3.1 Standard Approaches

#### Simulation

Monte Carlo simulation (MC) uses a sample of  $R$   $K$ -dimensional vectors  $\boldsymbol{\nu}$  (according to the PDF) to evaluate  $f(\boldsymbol{\nu})$ . The average of these evaluations gives the approximation:

$$S_R = \frac{1}{R} \sum_{r=1}^R f(\boldsymbol{\nu}_r) . \quad (3.8)$$

Taking draws can be illustrated by surveying individual preferences for product characteristics of  $R$  different individuals from a population (Train, 2009). According to the law of large numbers, it holds that  $S_R$  converges in probability to  $A$  by the rate of  $\sqrt{R}$  and it is asymptotically normal distributed according to the central limit theorem.

Quasi MC methods (qMC) provide a better coverage of the integral by spreading the draws more evenly (Bhat, 2001). This results in a higher approximation accuracy for a given amount of draws and often a better convergence rate than MC. Halton sequences are based on this mechanism and use a deterministic sequence of numbers. To maintain the statistical properties of random draws, this paper uses randomized Halton sequences (Bhat, 2003).

In the BLP model, the variance of MC and qMC based market share approximations propagates to the variance of the estimated parameters. Both, Freyberger (2015) and Berry, Linton, and Pakes (2004), show this effect in their asymptotic analysis for a growing amount of markets and products and conclude that a sufficient amount of draws has to be used in order to minimize the impact of the integration error.

### Quadrature

Gaussian quadrature rules constitute another approach in numerical integration, where the integral is approximated by a weighted sum of function evaluations. The evaluation of  $f(\boldsymbol{\nu})$  takes place at different points, also called nodes, in the best possible deterministic way given a desired accuracy level  $l$ . A detailed explanation of quadrature rules is given in Davis and Rabinowitz (2007).

In the one-dimensional case, the integral  $A$  is approximated by

$$G_l = \sum_{r=1}^{R_l} f(\nu_r) w_r \quad \text{with } \nu_r \in \mathbb{X}_l \text{ and } w_r \in \mathbb{W}_l . \quad (3.9)$$

Each accuracy level comes with a different set of  $\mathbb{X}_l$  and  $\mathbb{W}_l$ , where both sets consist of  $R_l$  elements. Nodes and weights depend on the PDF, which is specified in the integration problem.

The accuracy of the integral approximation increases in the number of nodes. This is based on the Weierstrass theorem stating that any function can be approximated by a polynomial arbitrarily closely, as long as the function is smooth. With Gaussian quadrature rules, the approximation error is zero, if  $f(\nu)$  is a polynomial of order  $p = 2R_l - 1$ .

Straightforward extensions to multiple dimensions often come with an exponential growth of nodes in the number of dimensions. A prominent example is the product rule, which is a straightforward and well-known extension of a univariate quadrature rule: for a  $K$ -dimensional integral,  $f(\boldsymbol{\nu})$  needs to be evaluated at all combinations of univariate sets of nodes. The evaluations are weighted by the product of the underlying univariate weights:

$$(G_{l_1} \otimes G_{l_2} \otimes \cdots \otimes G_{l_K})[f] = \sum_{r_1=1}^{R_{l_1}} \cdots \sum_{r_K=1}^{R_{l_K}} f(\nu_{r_1}, \dots, \nu_{r_K}) \prod_{k=1}^K w_{r_k} . \quad (3.10)$$

Assuming that in every dimension  $R_l$  nodes are used, the number of function evaluations equals  $(R_l)^K$ . In this case, the product rule integrates  $f(\boldsymbol{\nu})$  exactly, if the function is a multivariate polynomial<sup>7</sup> with a maximum exponent of  $2R_l - 1$ . The number of 10,000 function evaluations is often regarded as a computationally feasible upper limit in the BLP context, which is exceeded quickly with the product rule. This limits the application of integration rules with an exponential growth of nodes and motivates the use of a more parsimonious approach.

The sparse grids approach (SG) is also based on univariate quadrature rules, but saves a substantial amount of function evaluations by combining nodes of different accuracy levels in different dimensions such that while being very fine in one dimension, they are coarse in the others. Therefore, the amount of function evaluations increases only polynomially as  $K$  is rising. This, however, comes at the cost of relying on a more restricted

<sup>7</sup>A multivariate polynomial is defined as  $g(\mathbf{x}) = \sum_{t=1}^T a_t \prod_{k=1}^K x_k^{j_{t,k}}$ .

### 3 Implications of Adaptive Integration Rules

class of polynomials to approximate  $f(\boldsymbol{\nu})$ . A detailed discussion of the approach from an econometric point of view is provided in Heiss and Winschel (2008). Bungartz and Griebel (2004) discuss the approach from a mathematical point of view.

The SG rule is based on multiple product rules, each with a different set of accuracy levels:

$$S_u = \sum_{q=u-K}^{u-1} (-1)^{u-1-q} \binom{K-1}{u-1-q} \sum_{\mathbf{l} \in \mathbb{N}_q^K} (G_{l_1} \otimes G_{l_2} \otimes \dots \otimes G_{l_K})[f] .$$

Accuracy levels  $\mathbf{l}$  for a particular product rule, i.e. for a given  $q$ , are obtained by the following combination technique:

$$\mathbb{N}_q^K = \left\{ \mathbf{l} \in \mathbb{N}^K : \sum_{k=1}^K l_k = K + q \right\} .$$

The SG rule is exact, if  $f(\boldsymbol{\nu})$  is a polynomial with a total order<sup>8</sup> of  $2u - 1$ . A rising  $u$  goes along with a higher approximation accuracy of  $S_u$ . For sufficiently smooth functions, convergence rates and error levels can be much better than MC and qMC.

For practical implementations, nodes and weights can be obtained from software packages or readily available grids.<sup>9</sup> The integral is then simply approximated as:

$$S_u = \sum_{r=1}^{R_u} f(\boldsymbol{\nu}_r) w_r .$$

#### 3.3.2 Adaptive Integration

Integration rules that adjust any standard approach based on the function  $f(\boldsymbol{\nu})$  to increase the integration efficiency are called “adaptive” in the remainder of this paper. One common way for MC rules is importance sampling, where replacing the standard normal density by a well-chosen density assures more accurate results. This section introduces adaptive integration rules by adjusting the set of draws via a Gaussian importance sampler.

Based on equation (3.7), the following change of variables can be performed, if the new density has the same support as  $\phi(\boldsymbol{\nu})$ :

$$\int_{\mathbb{R}^K} f(\boldsymbol{\nu}) \phi(\boldsymbol{\nu}) d\boldsymbol{\nu} = \int_{\mathbb{R}^K} \frac{f(\boldsymbol{\nu}) \phi(\boldsymbol{\nu})}{\phi(\mathbf{L}^{-1}(\boldsymbol{\nu} - \mathbf{a}))} \phi(\mathbf{L}^{-1}(\boldsymbol{\nu} - \mathbf{a})) d\boldsymbol{\nu} . \quad (3.11)$$

$\Delta$  collects a vector of means  $\mathbf{a}$  and a matrix  $\mathbf{L}$ . For a given set of these parameters, the change of variables gives a function

<sup>8</sup>Given the representation of a multivariate polynomial in the former footnote, the total order is defined as  $\max_{t=1, \dots, T} \sum_{k=1}^K j_{t,k}$ .

<sup>9</sup>See [www.sparse-grids.de](http://www.sparse-grids.de) (Heiss and Winschel, 2008) or the R package mvQuad (Weiser, 2016).

$$\int_{\mathbb{R}^K} h(\mathbf{z}, \Delta) \phi(\mathbf{z}) d\mathbf{z} \quad \text{with } \mathbf{z} = \mathbf{L}^{-1}(\boldsymbol{\nu} - \mathbf{a}) \quad , \quad (3.12)$$

which can be approximated by standard approaches.

For an optimal choice of  $\Delta$ , the transformed function  $h(\mathbf{z}, \Delta)$  becomes a smoother function than  $f(\boldsymbol{\nu})$  making it more suitable for polynomial approximation or stochastic integration. Intuitively stating, the same amount of more cleverly chosen draws induces a lower approximation error than the regular integration rule.

This paper uses an efficient importance sampling algorithm to obtain  $\Delta$ . This method performs well for MC/qMC (Richard and Zhang, 2007) and SG (Heiss, 2010) and is based on minimizing the estimated variance of the weighted distance between  $f(\boldsymbol{\nu})\phi(\boldsymbol{\nu})$  and  $\phi(\mathbf{L}^{-1}(\boldsymbol{\nu} - \mathbf{a}))$ :

$$\hat{\Delta} = \underset{\Delta}{\operatorname{argmin}} \sum_{r=1}^R \left[ \ln \left( \frac{f(\boldsymbol{\nu}_r) \phi(\boldsymbol{\nu}_r)}{\phi(\mathbf{L}^{-1}(\boldsymbol{\nu}_r - \mathbf{a}))} \right) + c \right]^2 f(\boldsymbol{\nu}_r) w_r \quad . \quad (3.13)$$

In a quadrature setting,  $\boldsymbol{\nu}_r$  and  $w_r$  denote nodes and weights of a regular SG rule, for qMC or MC these are the random draws and weights  $w_r = \frac{1}{R}$ . The minimization gives a weighted least squares problem that can be solved analytically for  $\mathbf{a}$  and  $\mathbf{L}$ . Note that the use of too few initial draws leads to an imprecise estimate of  $\mathbf{a}$  and  $\mathbf{L}$ .

### 3.3.3 Integration in the BLP Model

The properties of the integration problem in the BLP model assure that the approximation error can be bounded arbitrarily close to zero. In practice, however, integrals are evaluated at a finite set of draws restricting the error to a given level. For a given amount of draws, this error is significantly reduced by adaptive integration. This is important, because the approximation error translates to incorrect point estimates<sup>10</sup>.

Unfortunate combinations of  $\delta_{jt}$ ,  $\sigma_k$  and  $x_{jt,k}$  can induce small intervals of  $\nu_k$  with a large contribution to the choice probability. Moreover, these small intervals can be located in the tails of the standard normal distribution. At least one outlier in  $\delta_{jt}$ ,  $\sigma_k$  or  $x_{jt,k}$  is a necessary condition for this phenomenon<sup>11</sup>. Figures 3.1 and 3.2 demonstrate the effects of the following parameter variations on an exemplary market share integral in a market with 5 products, while all other model parameters are kept constant:

- Setting A:  $x_1, x_2 \sim U(0, 1)$ ,  $\sigma_1 = \sigma_2 = 1$ ,  $\delta_j = 0$

<sup>10</sup>Intuitively, this can be explained by examining the contraction mapping in equation (3.6) at  $s_{jt}^{\text{obs, true}}$  (i.e. observed shares without any sampling error) and  $s_{jt}(\delta_{jt}^{\text{true}}, \boldsymbol{\theta}_2^{\text{true}})$ : in this case, there is no need for further iteration. However, in the presence of approximation error the contraction mapping will lead away from the true to a defected  $\delta_{jt}$ .

<sup>11</sup>The effect can be derived analytically by examining the impact of  $\delta_{jt}$ ,  $\sigma_k$  and  $x_{jt,k}$  on  $f(\boldsymbol{\nu})$  and  $\frac{\partial f(\boldsymbol{\nu})}{\partial \nu_k}$ . A formal analysis is not considered in the following due to two reasons: (i) numerous assumptions about the relation of the parameters in  $f(\boldsymbol{\nu})$  would be necessary and (ii) the results have no immediate use for the considered adaptive integration approaches.

### 3 Implications of Adaptive Integration Rules

- Setting B:  $x_1, x_2$  from A with outliers,  $\sigma_1 = \sigma_2 = 5, \delta_j = -15$

Sharp declining integrands with a very limited relevant region of  $\nu$  as in setting B are problematic for a specific numerical integration rule. That is, for MC and qMC methods it gets more unlikely to capture those regions with a low amount of draws. Quadrature rules are also problematic, since polynomials give a poor approximation of these functions. Note that even if true parameters of the demand model give well-behaving integrands, the estimation algorithm varies parameters extensively. Comparing the error level of standard integration rules in the convergence plots of figures 3.1 and 3.2 visualizes the problem.

The accumulated error resulting from every inaccurate market share approximation in every step of the estimation procedure impacts the shape of the GMM function severely. Figure 3.3 is based on the data from setting A and shows the GMM function value for different values of  $\sigma_1$ . Graphs are based on a standard randomized Halton rule with a total of 200, 10,000 and 100,000 draws, respectively. For each amount of draws, five different sets of generated draws are considered. The upper panel with sets of just 200 draws shows a huge variation in resulting GMM functions and results in very different point estimates. Moreover, multiple local minima can be introduced (see, for example, the cyan colored line) making the problem dependent on the starting guess. These problems diminish with more precise integration rules. The lowest panel, for example, shows almost no variation in the GMM function between different sets of draws. As explained in the following, switching from standard to adaptive integration rules attains the same effect on the GMM as increasing the amount of draws with a standard integration rule.

Performing efficient importance sampling adjusts the draws to the *relevant* region of  $\nu$  and assures integration with a maximal numerical efficiency. In the right panel of figure 3.2 (a) and (b) for example, it is easily observable that adaptive integration, here demonstrated by randomized Halton draws and SG nodes, evaluate the integrand at points with a large contribution to the integral.

Higher integration efficiency reduces the approximation error for a given amount of draws. The convergence plots in figures 3.1 and 3.2 give the integration accuracy (measured by the RMSE<sup>12</sup>) of the exemplary integrand depending on the used precision<sup>13</sup>. Adaptive versions of standard integration rules decrease the error level significantly. Low accuracy levels with the adaptive Halton rule in the convergence plot of setting B show a situation where too few initial draws result in an imprecise estimation of the adaptivity parameters.

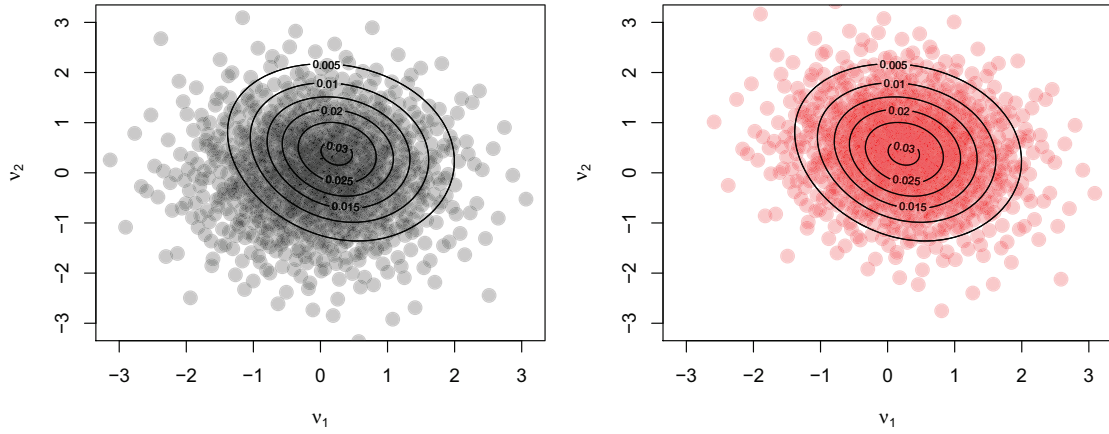
For integration rules evaluating all integrals in a given market at the same set of draws, parts of the market share approximation can be used for different integrals. This is, because the term  $\delta_{jt} + \sum_{k=1}^K x_{jt,k} \sigma_k \nu_{i,k}$  is part of every integral in a given market (either in the nominator or in the denominator) and each market share approximation uses the

<sup>12</sup>To capture the uncertainty of MC and randomized Halton sequences, respective errors are based on 50 different sets of draws for each accuracy level.

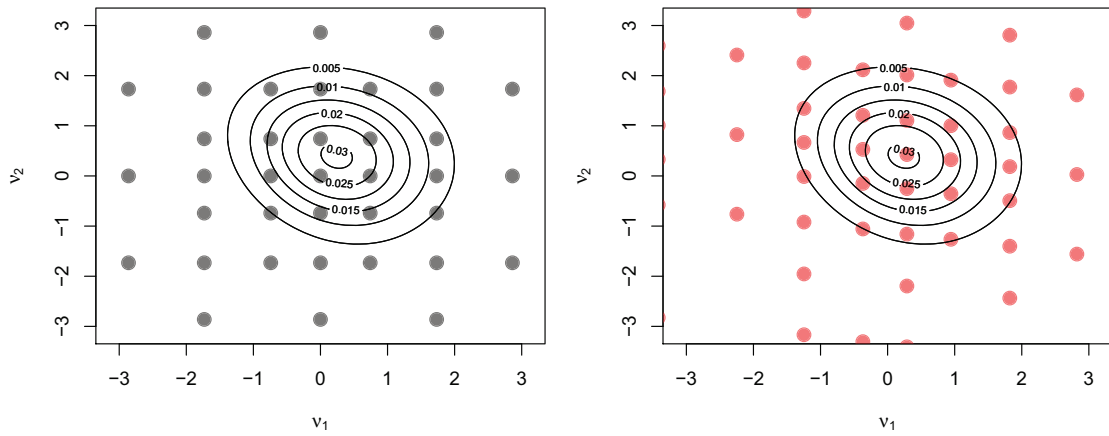
<sup>13</sup>The maximum number of available nodes for SG in two dimensions is 921.

Figure 3.1: Exemplary Integrand in Setting A

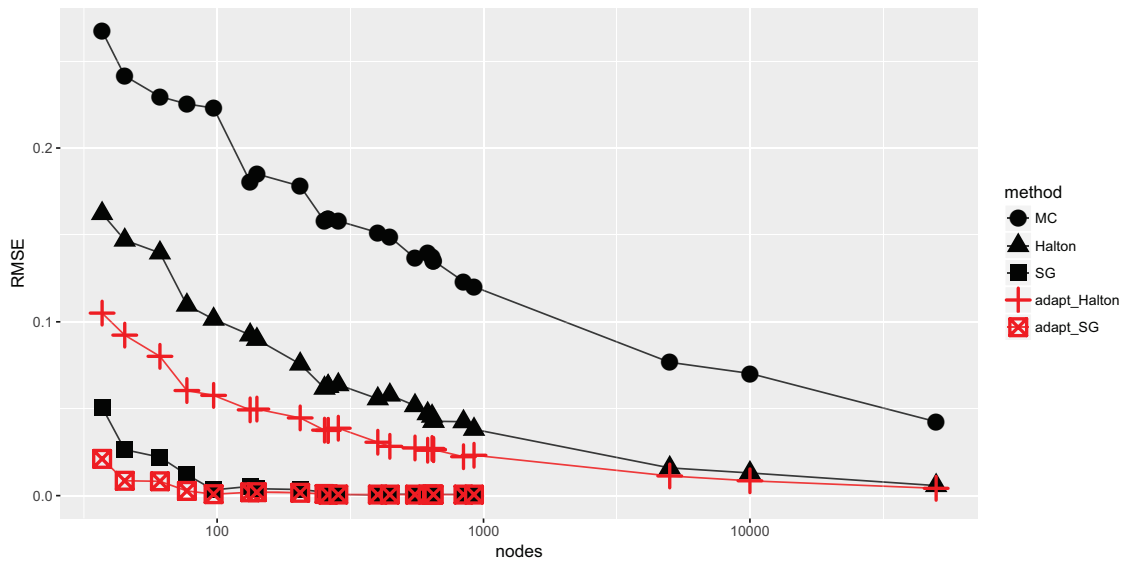
(a) Regular and Adaptive Randomized Halton Rule with 1000 Draws



(b) Regular and Adaptive SG Rule (Nested Gauss-Hermite with Accuracy Level 7)



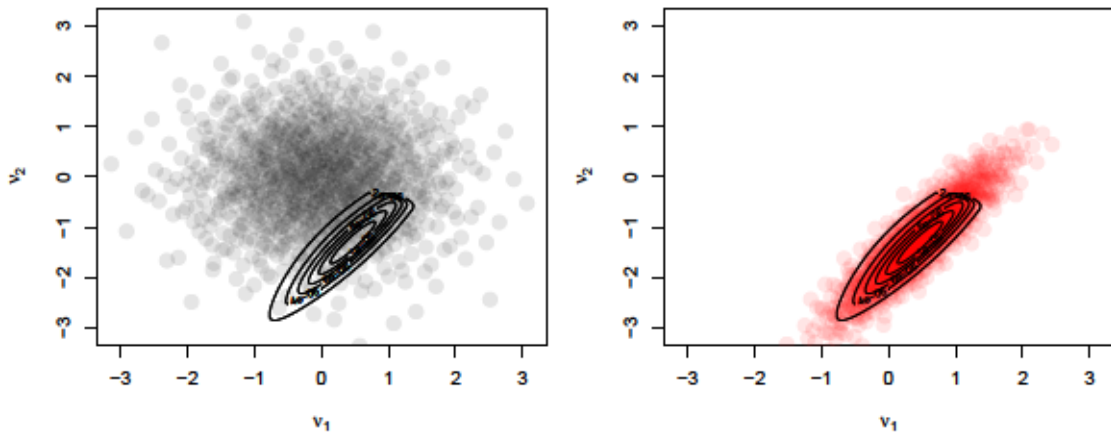
(c) RMSE for Different Amounts of Draws and Nodes



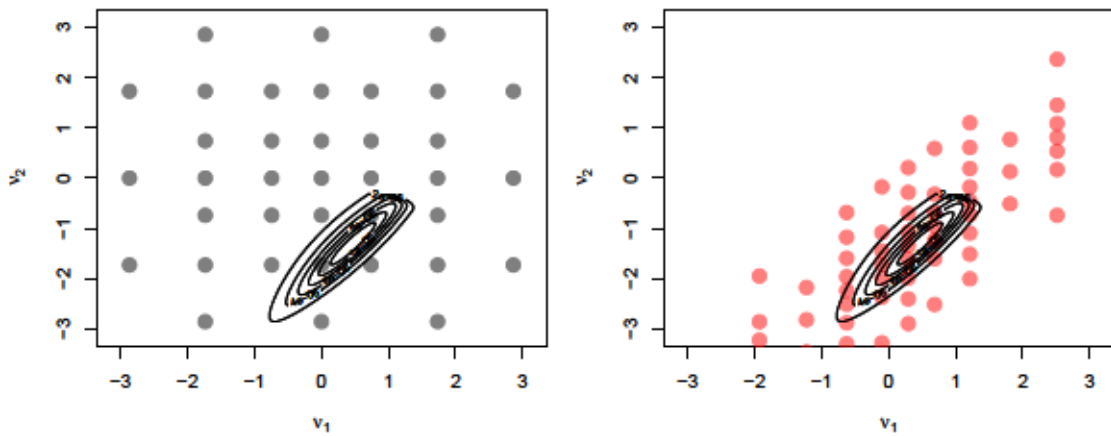
### 3 Implications of Adaptive Integration Rules

Figure 3.2: Exemplary Integrand in Setting B

(a) Regular and Adaptive Randomized Halton Rule with 1000 Draws



(b) Regular and Adaptive SG Rule (Nested Gauss-Hermite with Accuracy Level 7)



(c) RMSE for Different Amounts of Draws and Nodes

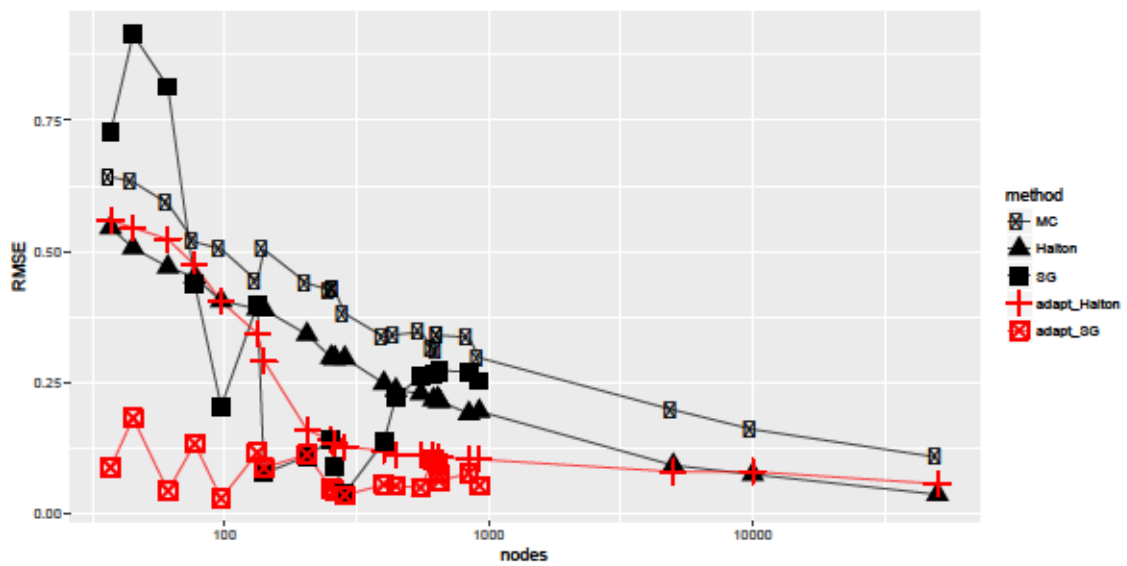




Figure 3.3: Impact of Inaccurate Integration Rules on the GMM Function

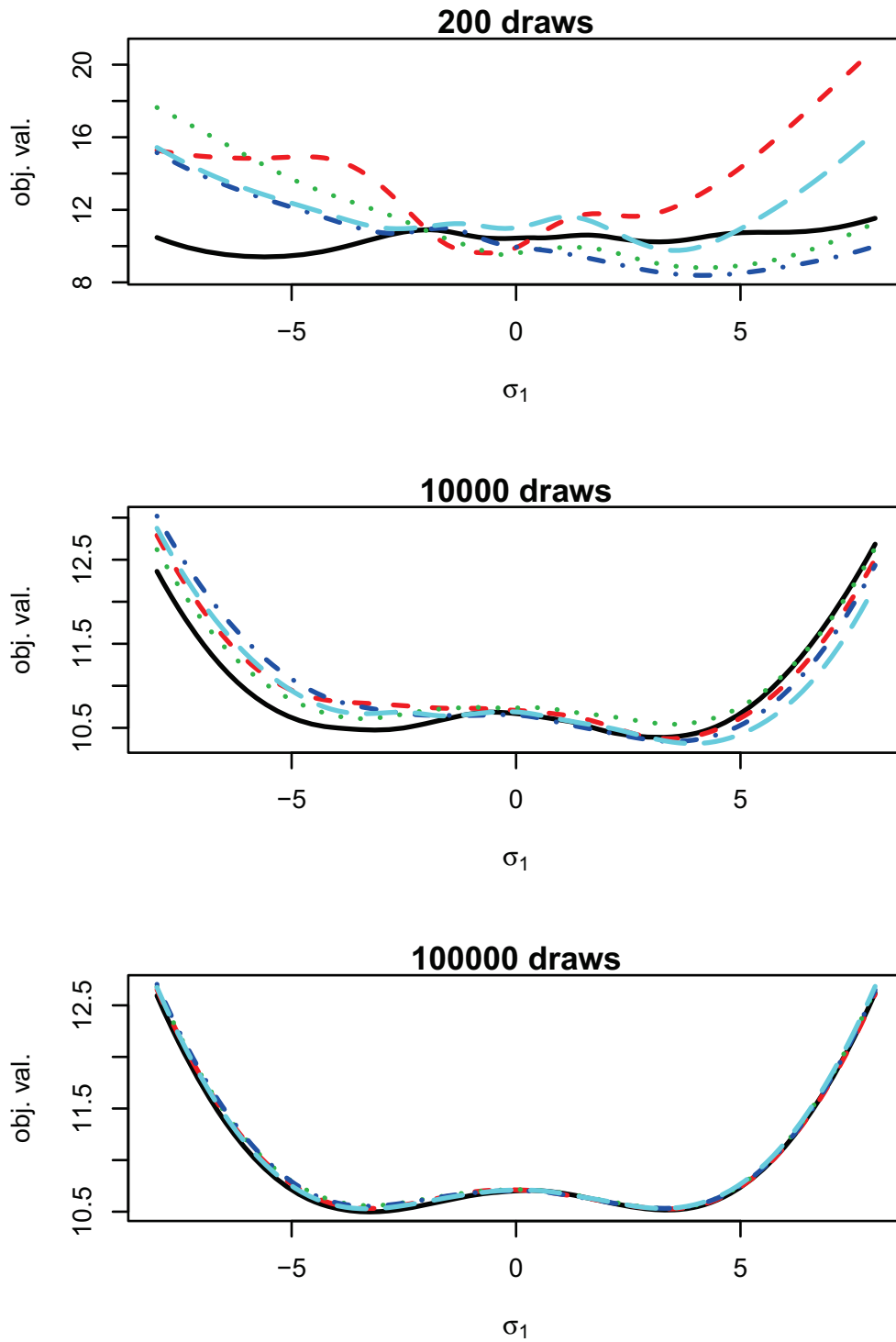
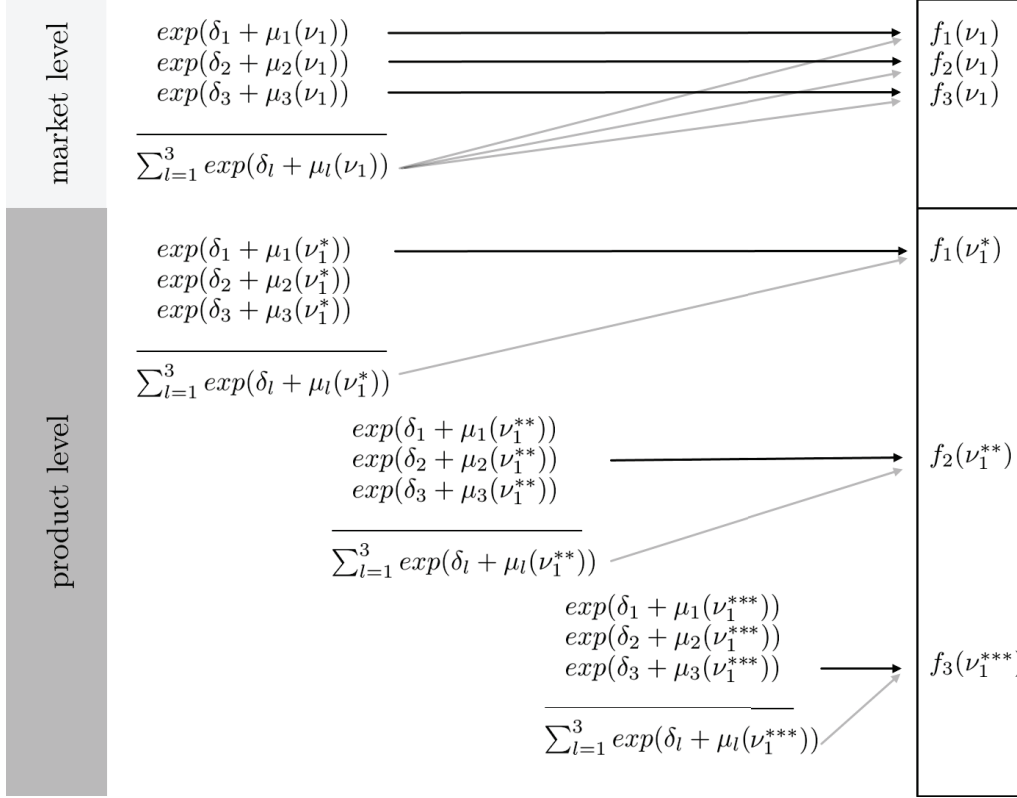


Figure 3.4: Computational Cost with Product- and Market-Wise Integration Rules



same set of  $\nu_{i,k}$ . For the calculation of  $J_t$  market shares in market  $t$ , this term only needs to be computed once for every product. With the adaptive approaches, the set of  $\nu_{i,k}$  is different for each integral, so for every integral the term has to be computed individually for all products<sup>14</sup>. The amount of function evaluations and RAM space to store the results is proportional to  $\sum_{t=1}^T J_t^2$  (compared to  $\sum_{t=1}^T J_t$  for the standard approaches). Figure 3.4 demonstrates this by an example with three products in one market. The upper panel uses the evaluation point  $\nu$  for every integral resulting in a total of 3 evaluations. The lower panel, however, uses an individual point for every market share integral, which gives a total of 9 function evaluations.

While the quadratic increase of computational cost is an issue in a setting with large  $J_t$ , adaptive integration can be the only way for misbehaving integrands to obtain an acceptable approximation error with a feasible amount of draws. The change of the convergence behavior between setting A and B indicates that this is the case for problems where very small regions of  $\nu$  contribute a lot to the integral.

To prevent the problem of computational cost for large  $J_t$  to some extent while benefiting of a higher integration efficiency, the performance of an additional approach is examined: draws are adapted market-wise based on an additive overlapping of all functions  $f(\nu)$  in

<sup>14</sup>The same problem would arise, if different draws for each integral are used with standard MC or qMC.

a given market. The efficiency gain of the numerical integration approach is preserved, if the integrands are homogeneous within one market. Details of this approach can be found in appendix 3.D.

## 3.4 Results

### 3.4.1 Data Generating Process

The performance of each discussed integration approach is assessed by evaluating 100 simulated datasets on different levels of integration accuracy. An emphasis is put on the precision of BLP point estimates and on computational cost.

Each dataset includes 300 observations ( $T = 20$  markets and  $J = 15$  products) and is estimated for different accuracy/ integration method/ dimension combinations. For the purpose of precision evaluation, estimates from each combination are compared to the set of “true” random coefficients that result from an estimation with (almost) no integration error. All  $\sigma_k$  in the DGP range between 0 and 4 making sure to include a mix of integration problems with moderate difficulty and more challenging ones. All settings across datasets result in a total of 72,500 BLP estimations. Other parameter choices and the design of the DGP are based on Reynaert and Verboven (2014) and described in appendix 3.A.

All statistics in this section are based on the starting guess of the optimization routine that produces the lowest GMM objective at the respective point estimates out of five starting guesses. The BFGS solver converged for every estimation. Additional computational details of the estimation routine are given in appendix 3.B.

### 3.4.2 Precision

Results are related to random coefficients since they define the optimization problem<sup>15</sup>. Precision of each accuracy/ integration method/ dimension combination is measured by the root mean squared error (RMSE) across datasets :

$$\text{RMSE}_k = \sqrt{\sum_{i=1}^{100} \frac{(\hat{\sigma}_{k,i} - \sigma_{k,i}^{true})^2}{100}}, \quad \text{for } k = 1, \dots, K . \quad (3.14)$$

The random coefficient  $\sigma_{k,i}^{true}$  is estimated with a very accurate randomized Halton rule with 100,000 draws. Reported RMSEs in all tables are averaged across dimensions due to reasons of clarity and readability. So, for a given dimension  $n$ , the reported RMSE equals  $\frac{1}{n} \sum_{k=1}^n \text{RMSE}_k$ .

Tables 3.1 and 3.2 show a strong dependency between the number of draws and the dispersion of the point estimates. Moreover, a higher accuracy level always goes along

<sup>15</sup>Due to the functional relationship, precision of linear parameters is affected as well.

Table 3.1: RMSE of  $\hat{\sigma}$  for Non-Adaptive Measures (Simulated Data)

<b>Dimension 1</b>	MC	Halton	SG
17 draws	0.92	0.89	0.26
19 draws	0.94	0.80	0.17
31 draws	0.97	0.76	0.09
33 draws	0.98	0.70	0.06
<b>Dimension 2</b>	MC	Halton	SG
37 draws	1.03	1.00	0.41
45 draws	1.01	0.98	0.41
97 draws	0.90	0.62	0.34
401 draws	0.66	0.38	0.08
<b>Dimension 3</b>	MC	Halton	SG
93 draws	1.09	0.78	0.63
165 draws	0.84	0.66	0.42
237 draws	0.71	0.52	0.39
919 draws	0.50	0.24	0.13
<b>Dimension 4</b>	MC	Halton	SG
33 draws	1.12	1.21	1.33
201 draws	0.85	0.83	0.56
761 draws	0.69	0.49	0.23
4489 draws	0.34	0.26	0.16
<b>Dimension 5</b>	MC	Halton	SG
51 draws	1.34	1.58	1.14
401 draws	0.88	0.68	0.55
2033 draws	0.57	0.30	0.40
3793 draws	0.36	0.24	0.33

Table 3.2: RMSE of  $\hat{\sigma}$  for Adaptive Measures (Simulated Data)

<b>Dimension 1</b>	adaptHalton (market)	adaptSG (market)	adaptHalton (product)	adaptSG (product)
17 draws	0.59	0.18	0.50	0.05
19 draws	0.49	0.04	0.38	0.01
31 draws	0.39	0.06	0.45	0.02
33 draws	0.33	0.03	0.35	0.01
<b>Dimension 2</b>	adaptHalton (market)	adaptSG (market)	adaptHalton (product)	adaptSG (product)
37 draws	0.60	0.33	0.30	0.17
45 draws	0.47	0.42	0.34	0.10
97 draws	0.33	0.21	0.16	0.05
401 draws	0.21	0.08	0.10	0.08
<b>Dimension 3</b>	adaptHalton (market)	adaptSG (market)	adaptHalton (product)	adaptSG (product)
93 draws	0.31	0.26	0.21	0.17
165 draws	0.30	0.27	0.14	0.10
237 draws	0.24	0.17	0.11	0.08
919 draws	0.16	0.04	0.11	0.03
<b>Dimension 4</b>	adaptHalton (market)	adaptSG (market)	adaptHalton (product)	adaptSG (product)
33 draws	0.70	0.64	0.44	0.37
201 draws	0.25	0.32	0.33	0.16
761 draws	0.19	0.18	0.22	0.12
4489 draws	0.05	0.15	0.07	0.03
<b>Dimension 5</b>	adaptHalton (market)	adaptSG (market)	adaptHalton (product)	adaptSG (product)
51 draws	0.60	0.59	0.45	0.35
401 draws	0.28	0.35	0.22	0.16
2033 draws	0.25	0.23	0.18	0.11
3793 draws	0.08	0.22	0.06	0.15

### 3 Implications of Adaptive Integration Rules

with a decreasing RMSE. This is not surprising, because a higher amount of draws gives a lower approximation error in the market share integrals. In general, randomized Halton sequences perform much better than MC.

In small dimensions SG outperforms MC/ qMC rules. In higher dimensions, however, the disadvantage of MC/ qMC diminishes, because with these methods errors across different dimensions cancel each other out. Moreover, they have a better coverage in higher dimensions than regular SG.

The relation between errors from regular integration rules (table 3.1) and adaptive rules (table 3.2) is in line with section 3.3: if the draws for integrand evaluation are shifted to the relevant parameter space, integration accuracy and therefore parameter precision increases. Moreover, with adaptive rules, the weaker performance of non-adaptive quadrature in higher dimensions is compensated to some extent.

Adaption on a product level, i.e. for every single market share integral, gives the most accurate results in all settings. Errors of market-wise adaption ranges between product level adaption and no adaption<sup>16</sup>.

#### 3.4.3 Computational Cost

Computational time in seconds of each accuracy/ integration method/ dimension combination across datasets is visualized by a categorized box plot in figure 3.5. Note that time differences do not only capture the direct effect of a change in the method or accuracy level but also indirect effects like convergence to a different local minimum. Differences in memory usage are approximately proportional to the computation time differences.

Adaptive methods on the product level show a significant increase of computational cost reflecting the direct relation between time and integral evaluations. For details about the source of additional evaluations see section 3.3.3. Adaption on the market level shows no clear time increase that is needed for the adaption of integration rules.

In combination with tables 3.1 and 3.2, it can be stated that the selection of a method should not be guided by time advantages, even if time is a critical factor. Instead, only after a particular level of precision is achieved time consuming integration methods can be excluded.

## 3.5 An Application to the Cereal Market

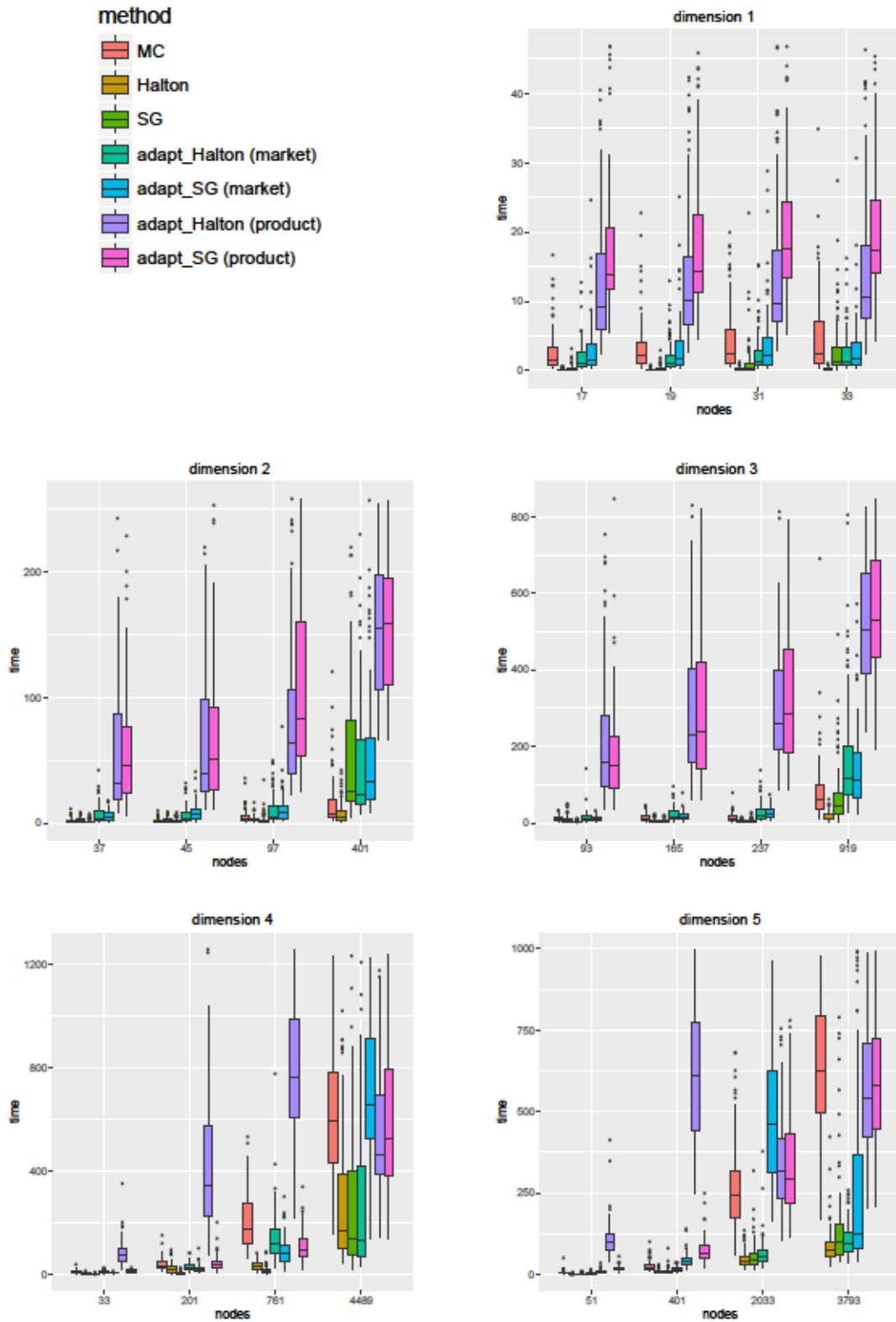
Effects on numerical properties of the BLP estimation procedure are often illustrated by the well-known training data for the cereal market from Nevo (2000). Following this

---

<sup>16</sup>Additional simulations (not included in this paper) show that more difficult integration problems decrease the probability that relevant parameter spaces of the integration problem are overlapping. In some cases, this can lead to a poor performance of market-wise adaption making the method somewhat problem dependent.

### 3.5 An Application to the Cereal Market

Figure 3.5: Total Computation Time in Seconds Across Datasets



### 3 Implications of Adaptive Integration Rules

Table 3.3: RMSE of  $\hat{\sigma}$  for Non-Adaptive Measures (Cereal Data)

	MC	Halton	SG
33 draws	0.54	0.20	0.02
201 draws	0.22	0.05	0.00
761 draws	0.11	0.02	0.00
4489 draws	0.04	0.01	0.00

Table 3.4: RMSE of  $\hat{\sigma}$  for Adaptive Measures (Cereal Data)

	adaptHalton (market)	adaptSG (market)	adaptHalton (product)	adaptSG (product)
33 draws	0.04	0.02	0.03	0.01
201 draws	0.02	0.00	0.01	0.00
761 draws	0.02	0.00	0.02	0.00
4489 draws	0.02	0.00	0.02	0.00

approach, Nevo’s study is reestimated with different sets of draws and accuracy levels for the discussed integration rules.

The original setting of Nevo’s exemplary analysis remains unchanged with one important exception: effects of demographics are assumed to be zero. This has two convenient consequences. First, there is no need to make use of the very limited data on demographics (just 20 data points per market and demographic) and second, this simplifies the problem to an integral over a normal density.

The cereal dataset is estimated with 50 different sets of draws for MC/qMC rules and only with one set of (deterministic) nodes for SG. Estimates are compared to the “true” set of parameters that are obtained from a randomized Halton integration rule with 100,000 draws. Tables 3.3 and 3.4 show the average RMSE of four random coefficients.

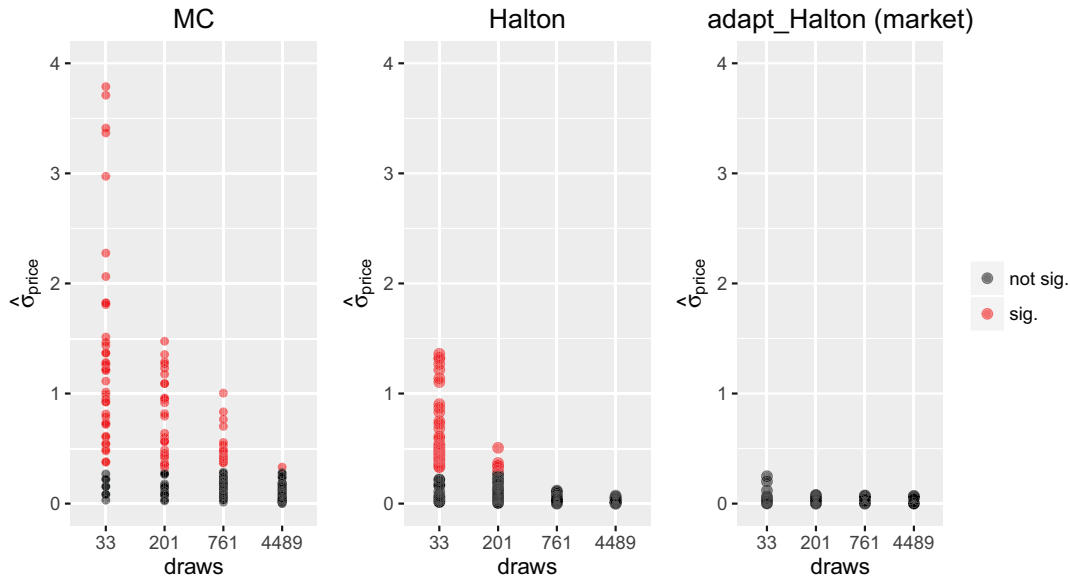
The baseline estimation with 100,000 draws gives a minimum of the GMM objective at 189.9432. Inaccurate integration introduces false local minima, which results in point estimates being different from the baseline estimation.

RMSEs of the different integration approaches tell the same story as with simulated data: adaption on the product level performs slightly better than on the market level and qMC performs better than MC. SG gives impressive results in all settings<sup>17</sup>.

In the cereal example, computational cost is up to 20 times higher with adaptive approaches. For the case of 4,489 draws and adaptive Halton sequences, this comprises a

<sup>17</sup>This might be surprising given the results for higher dimensions in section 3.4. Note, however, that due to  $\theta_2$  close to zero, the integrand is near the logit model. Polynomials easily approximate these nearly constant functions.



Figure 3.6: Convergence of  $\hat{\sigma}_{price}$  (Cereal Data)

total memory usage of 1 gigabyte RAM<sup>18</sup> and an average computation time of 3.7 hours<sup>19</sup> (compared to 40 megabytes and 20 minutes with a standard Halton rule). While still being feasible, consider another hypothetical example with the most unfortunate product/market combination: all observations in just one market. This would give an increase of computational cost up to a factor of 2,000 compared to the non-adaptive case resulting in about 85 gigabytes of required RAM. This demonstrates that in some cases, precise standard integration or market-wise adaptive rules might be the only options.

One random coefficient that is of particular importance from an economic point of view is the random coefficient of price. Figure 3.6 shows a selection of three methods and again demonstrates the importance of accurate integration. Though the coefficient is insignificant, the majority of inaccurate standard integration methods estimate a large range of significant coefficients. This means that without accurate numerical integration, researchers can find any (significant) result just by making use of the integration noise.

### 3.6 Conclusion

Inaccurate market share approximations in the BLP model are a major source of error. The complexity of the integration problem largely depends on parameters of the underlying choice model. Difficult cases have a very small region on the integrand's support with a

<sup>18</sup>This results from considering a matrix filled with 4,489 draws (for 94 markets with 24 products) represented by floating point numbers with single precision:  $94 \cdot 24^2 \cdot 4489 \cdot 4 \text{ byte} \approx 1 \text{ gigabyte}$ .

<sup>19</sup>Although the direct effect of additional function evaluations increases computation time by a factor of 20, the faster convergence induced by more accurate integration reduce the time disadvantage to a factor of 10 in the present case.

### *3 Implications of Adaptive Integration Rules*

high contribution to the market share. Since parameters are varied extensively by the estimation algorithm, these difficult integrands can be encountered quite often during estimation. In the worst case, this gives unreliable results even with 10,000 or more function evaluations. This paper applies and evaluates alternative integration approaches that fit stochastic and deterministic integration rules to the specific integrand.

Adaptive integration rules are more efficient for a given number of function evaluations, because only the relevant part of the integrand's support is considered. Their application is most useful for ill-behaved integrands where adaptive integration gives accurate approximations with a feasible amount of integration draws. Real data and simulation examples confirm the superior performance of adaptive approaches.

The standard approach in the BLP model is to reuse integrand evaluations, i.e. using the same set of draws for every product in a given market. This parsimonious approach is not in line with individually fitted integration rules, which can be infeasible with very large datasets. Adaption at a market level is therefore considered as well. With this approach, the optimal set of draws is based on all integrands in a given market and this set is shared across integral approximations. Market level adaption gives more accurate results than standard approaches in nearly all settings.

Aiming for precise market share approximation, either with adaptive approaches or many integration draws, drives computational cost. Attempts to reduce this cost in the BLP model are well-founded. In practice, datasets can become so large that the estimation easily becomes infeasible with highly accurate integration rules. As demonstrated by the results of this paper, one should not be tempted to use low-cost integration rules. Instead, it is highly recommended to run at least multiple estimations with different sets of draws to check the result's robustness. The standard MC approach always gives the most inaccurate estimates, so qMC or SG rules should be preferred. Whenever possible, highly accurate adaptive integration rules based on the techniques in this paper are the best option to produce reliable estimates.

The strong dependence of estimates and integration accuracy translates into fuzzy economic interpretations that build on the estimated BLP model. Implications of the model, for example cross price elasticities or welfare analyses, become prone to error and manipulation. Accurate integration prevents the researcher from being exposed to this weak points.

## Appendix 3.A Data Generating Process

The use of different simulated datasets requires the formulation of a DGP. This section gives details about the characteristics and chosen parameters of this process, which is based on the simulation setting in Reynaert and Verboven (2014).

Data is created as a balanced set, where each market has the same amount of products  $J$ . The total number of observations is therefore the number of markets,  $T$ , times  $J$ . For every generated dataset in this paper, there are six variables that enter the model linearly: price labeled  $p$  and five other product characteristics  $x_1$  to  $x_5$ .  $x_1$  to  $x_5$  are generated by a uniform distribution between zero and two. An additional set of 10 cost shifters (labeled  $c_{jt,1}$  to  $c_{jt,10}$ ) is drawn from the same distribution. All  $x$  and  $c$  define the set of exogenous variables  $\mathbf{Z}$ , because they are independent of the structural error term  $\xi$ .

For a dimension  $K$  in the integration problem, the first  $x_1, \dots, x_K$  product characteristics are modeled as random coefficients. Price is left out as a random coefficient due to notational simplicity.

Prices correlate with exogenous variables and the structural error term to introduce artificial endogeneity. Prices are computed as a function of  $\mathbf{Z}$  and a random variable  $\omega$ .  $\xi$  and  $\omega$  are drawn from a multivariate normal distribution with given means ( $\mu_\xi$  and  $\mu_\omega$ ) and variance/ covariances. This gives the following prices of product  $j$  in market  $t$ :

$$p_{jt} = \sum_{i=1}^{10} \gamma_{1,i} c_{jt,i} + \sum_{i=1}^5 \gamma_{2,i} x_{jt,i} + \omega_{jt} .$$

Without loss of generality, this specification implies perfect competition since the error term does not include any demand dependent component, so prices equal marginal costs.

In this paper, prices and exogenous parameters define the set of variables that enter the problem linearly. The true mean utility  $\delta$  can be calculated as:

$$\delta_{jt}^{true} = \alpha p_{jt} + \sum_{k=1}^5 \bar{\beta}_k x_{jt,k} + \xi_{jt} .$$

The resulting market shares are given by the integral with  $K$  random coefficients:

$$s_{jt}(\delta_{jt}^{true}, \boldsymbol{\theta}) = \int_{\mathbb{R}^K} \frac{\exp\left(\delta_{jt}^{true} + \sum_{k=1}^K x_{jt,k} \sigma_k \nu_{i,k}\right)}{1 + \sum_{l=1}^J \exp\left(\delta_{lt}^{true} + \sum_{k=1}^K x_{lt,k} \sigma_k \nu_{i,k}\right)} \phi(\boldsymbol{\nu}) d\boldsymbol{\nu} . \quad (3.15)$$

These integrals are approximated by an accurate integration rule (randomized Halton draws with 100,000 draws). Therefore, the resulting generated observed market shares are assumed to be without simulation error.

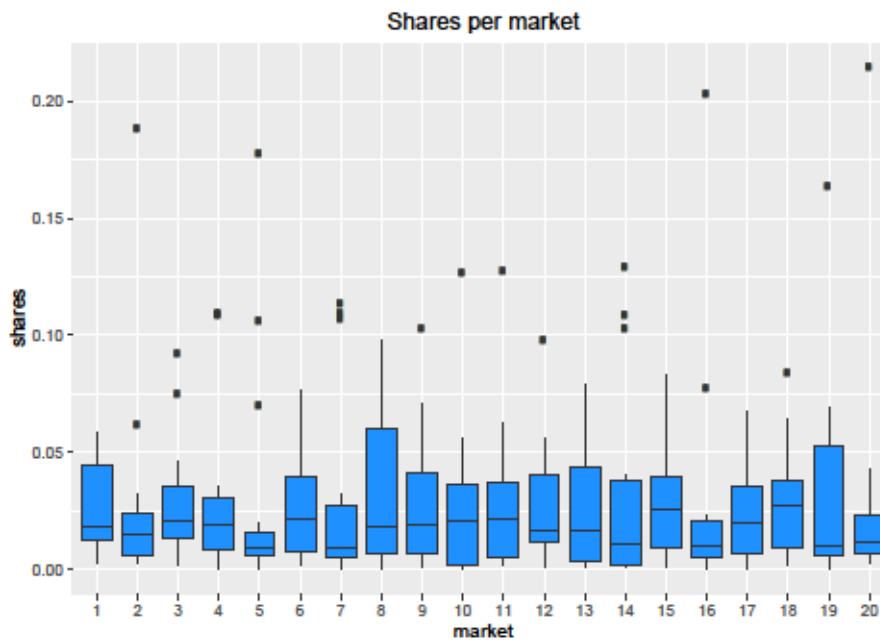
Table 3.5 shows the chosen values for the parameters. Figure 3.7 gives information about the resulting market shares in a typical BLP dataset based on the DGP.

### 3 Implications of Adaptive Integration Rules

Table 3.5: DGP Parameters

parameter	value
J	15
T	20
$\mu_\xi$	-2
$\mu_\omega$	0
$\text{Cov}(\xi, \omega)$	$\begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$
$\gamma_1$	2 2 2 2 2 2 2 2 2 2
$\gamma_2$	1 1 1 1 1
$\alpha$	-0.2
$\bar{\beta}$	0.2 0.2 0.2 0.2 0.2
$\sigma_k$	0.5

Figure 3.7: Exemplary Distribution of Market Shares



## Appendix 3.B Computational Details

This paper uses the statistical software `R` to perform the simulation study. Parts of the code are based on the well-known and widely used Nevo (2000) code.

Computations are performed on the HILBERT cluster at the University of Düsseldorf. Frequently used functions are coded in `C++` and are implemented in `R` with the `Rcpp` package. Nodes and weights of the SG approach are provided by the `mvQuad` `R` package (Weiser, 2016). The GMM objective is minimized with a boxed `R` solver applying the quasi Newton method BFGS (Broyden, 1970). This method uses analytic gradients and an approximated Hessian. Moreover, 5 starting values, an inner loop tolerance of  $10^{-9}$  and an outer loop tolerance of  $10^{-6}$  is used. Several checks indicate that results are robust with regard to tighter tolerance levels and starting values.

For standard MC and qMC, the same set of draws in a given market is used to approximate all related market share integrals. However, sets differ across markets.

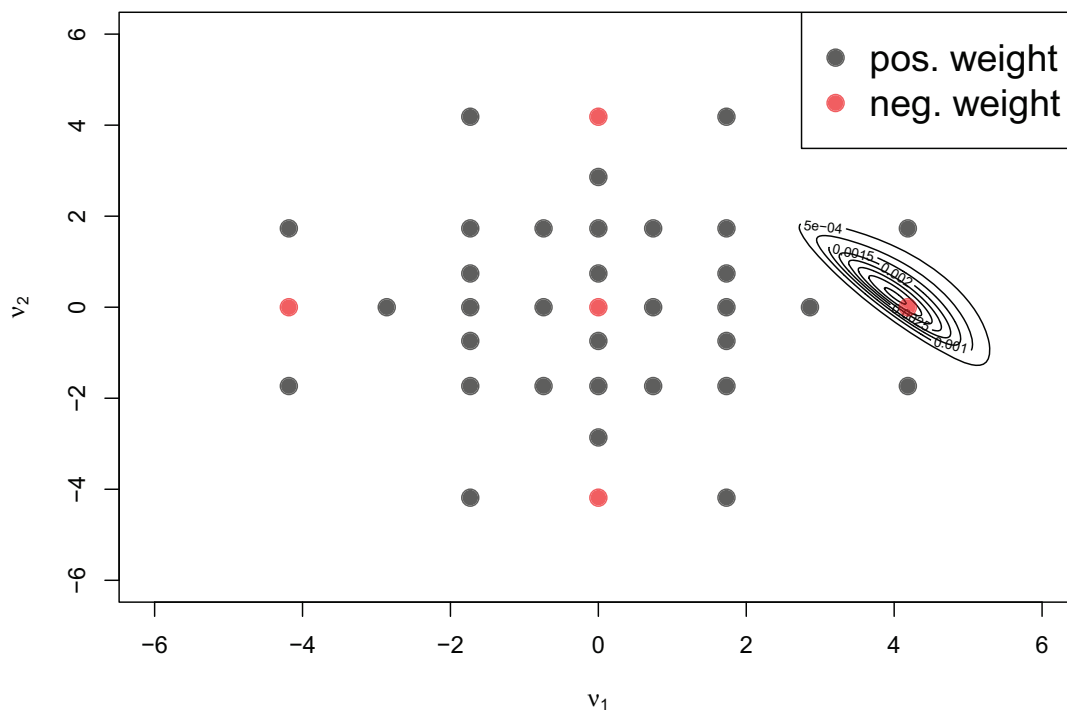
For the adaptive approaches, the researcher is free to decide when to update the set of draws. Two extreme cases are either to update them one single time at the beginning or every time a new integral needs to be calculated in the inner loop. The first possibility might not be the best choice due to the fact that the integral based on the starting values of  $\delta_{jt}$  and  $\theta$  changes a lot in the iterative procedure. So, it is unlikely that the adjusted set of draws will perform better in computing the final integral. The second possibility causes a lot of function calls while there might be little change between two iteration steps. This paper strikes a balance between the two extreme cases: The grid adjustment takes place for the first five iterations in every contraction mapping to account for the effect of a changing  $\delta$  on the integrand. This assures that new draws fit to the function they are approximating and limits the required function calls to a feasible level. Moreover, the algorithm performs the adaption only if the change between two outer loop iterations is significant ( $|\theta^n - \theta^{n+1}| > 0.01$ ) or there has been an error in the previous step (including undefined function evaluations or non-convergence of the contraction mapping).

### Appendix 3.C Problems with SG Integration

Due to their construction, SG rules work with negative weights in the integral approximation. There are cases, when these weights cause negative market share approximations. Figure 3.8 demonstrates this with a two dimensional integrand and a nested Gauss Hermite integration rule with accuracy level 9. Red colored nodes indicate function evaluations that enter the approximation with a negative weight. In this example, the unfortunate location of the integrand leads to a negative integral approximation, since other function evaluations with a positive weight are close to zero.

Negative approximations are not only problematic for an economic interpretation, but impedes the contraction mapping from converging. It turns out that resetting the value of  $\delta$  to the starting value of zero every time a negative share is computed works very well for the simulation study and the Nevo data. There are only few exceptions where the contraction mapping iterates up to the predefined maximum number.

Figure 3.8: Negative Integral Approximation with SG



## Appendix 3.D Adaptive Integral Approximation on the Market Level

To preserve the parsimonious way of market share approximations in the BLP model with standard integration rules while benefiting from highly efficient adaptive approaches, integration rules on the market level are proposed.

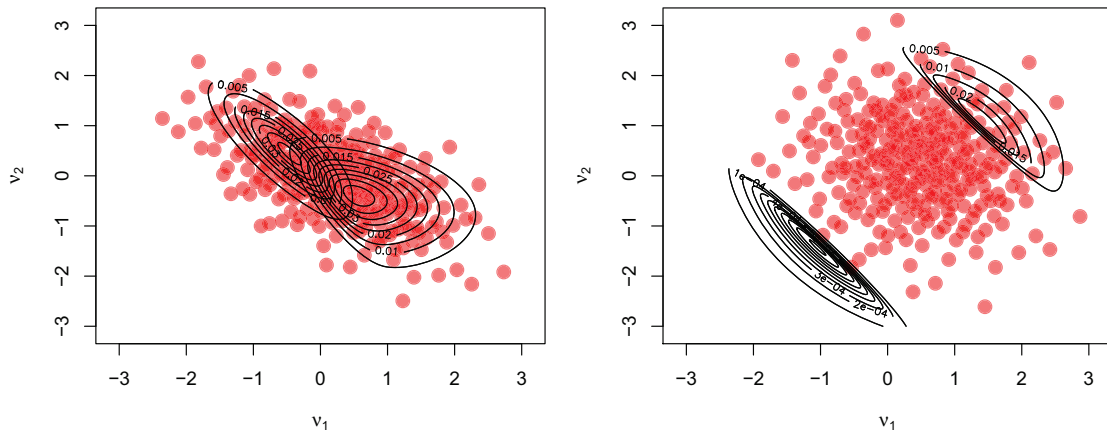
Using this integration method, the first step consists of additively overlaying all integrands in a given market  $t$ , which gives the following integration problem:

$$\int_{\mathbb{R}^K} \underbrace{(f_1(\boldsymbol{\nu}) + \dots + f_{J_t}(\boldsymbol{\nu}))}_{f_{\text{total}}(\boldsymbol{\nu})} \phi(\boldsymbol{\nu}) d\boldsymbol{\nu} .$$

While the approximation of this integral is not of interest in the BLP model, the function  $f_{\text{total}}(\boldsymbol{\nu})$  can be used to compute optimal evaluation points that consider all integrands in one market. This market level approach allows to reduce computational cost of the product level integration rule by sharing the set of optimal draws across  $J_t$  integrands.

The market level approach works well, if integrands are located close by. The left panel of figure 3.9 demonstrates this situation with an example of two market share integrands in one market. The right panel visualizes the opposite: if the location of integrands differs a lot, the market level approach will lose much of its efficiency. Both graphs use an adaptive randomized Halton rule with 300 draws on the market level. Simulation results in section 3.4 indicate that integrands in all markets tend to behave as in the left panel.

Figure 3.9: Market Level Integration with a Low and High Efficiency Loss







# 4 Copulas for Aggregated Demand Models with Partly Observed Heterogeneity

## 4.1 Introduction

The model of Berry, Levinsohn, and Pakes (1995) has become the workhorse for structural demand estimation due to a number of favorable attributes. This includes the explicit consideration of unobserved and observed consumer heterogeneity. Observed heterogeneity requires the handling of demographic variables and their joint distribution. For situations, in which only limited information about the joint distribution is available, this paper proposes copula functions. Copula modeling, being a very popular technique in statistics itself, allows to combine marginal distributions and a dependence parameter to a joint distribution in a very flexible fashion. Aggregated demand models give an interesting and important application for copulas and to the best of my knowledge no attempts have been made to combine these two topics. The main contribution of this paper lies in demonstrating the copula modeling for an exemplary demand estimation from the banking literature.

An important requirement for the approximation of BLPs' market share equation in the presence of demographics is the knowledge of their joint distribution, or at least the availability of extensive micro data across markets. With significant demographic effects, consumers from markets with a similar demographic structure tend to substitute products by other products with similar characteristics, which helps to overcome the restrictive substitution patterns of the logit model. Popular examples for applications of demand estimation with demographics are given in Nevo (2001) or Berry, Levinsohn, and Pakes (2004).

In empirical applications, the knowledge of the demographics' joint distribution or an reliable empirical counterpart is a standard that is hard to reach, which can impede a direct demand estimation. The problem often stems from legal restrictions that oblige providers of publicly available demographic information to protect individuals' privacy<sup>1</sup>. This results in protective measures like averaging data over years, providing only marginal or discretized demographic information. Obviously, this is of particular relevance for data on a fine geographic level. However, using a fine geographical market definition to describe

---

<sup>1</sup>See, for example, Title 13 U.S. Code (for the US) or §16 BStatG (for Germany).

a homogeneous group of consumers can be required in the BLP model to assure that the assumption of an identical unobserved mean utility for every consumer in a given market holds. The widely used American Community Survey (ACS) is an example for a source of demographic information that is heavily anonymized on finer geographic levels. Especially the exclusive information about marginal distributions requires the researcher to model the dependence structure separately before approximating market share integrals.

Copula functions allow to model an unknown joint distribution, if either marginal distributions are known, but it is not clear how the joint distribution of random variables can be specified, or a simple concept like the correlation coefficient does not sufficiently describe the dependence (see, for example, Trivedi and Zimmer (2007)). An important characteristic of copulas is the possibility of modeling marginal distributions and the dependence structure in different steps. This paper uses micro data from a broad geographic level<sup>2</sup> to estimate the dependence parameter. In a second step, the dependence parameter is combined with marginal distributions from nested areas on a finer geographic level<sup>3</sup> to construct the joint distribution. This procedure builds on the assumption that the dependence structure is the same for the two different geographic layers. Although the data situation in this paper seems to be tailored for the two-step procedure, information about the dependence structure can be extracted from far less information<sup>4</sup>.

The different steps of copula modeling are demonstrated by estimating the demand for deposits in commercial banks. As stated by Dick (2008), “Understanding the form of demand and consumer behavior in banking has several immediate uses. [...] The estimates of consumer preferences across bank characteristics can be used to analyze the effects of potential mergers or various other changes in the environment on consumer welfare.” The model of Dick (2008) is extended in two dimensions: (i) instead of a nested logit model, the model of BLP is used for demand estimation and (ii) the demographic variables age and income are incorporated in the model. The analysis gives an illustration of the proposed copula approach and is not meant to identify a causal relationship between variables. Moreover, the selection of demographics in (ii) just serves as an example for the copula modeling and is unlikely to comprise all economically relevant demographics<sup>5</sup>.

Due to the copula modeling, estimation results become available for a market definition with limited demographic information and indicate that bank products with similar deposit rates serve as substitutes. To demonstrate the implications of different data qualities the estimation is performed on two geographical layers, i.e. different market definitions: the

---

<sup>2</sup>This step comprises the use of Public Use Microdata Sample (PUMS) data containing micro data for every US state. For more details, see section 4.4.

<sup>3</sup>Data on this stage is taken from ACS on the level of statistical areas, called core based statistical area (CBSA).

<sup>4</sup>For all copula classes in this paper, for example, there exists a functional relation between Kendall's Tau and the dependence parameter. In this case, the researcher needs only a single reported parameter to proceed with step two.

<sup>5</sup>Even though, comparable studies like Ho and Ishii (2011) also rely on the use of age and income.

state level (broad market definition) and the core based statistical area (CBSA) level (fine market definition). For the broad market definition, PUMS micro data are available, so the demand estimation is straightforward since draws from the joint distribution can be generated directly. On the second and more realistic market level, marginal distributions are fitted by copula functions to generate a joint distribution. For all states, the Clayton copula provides the best fit predicting a very strong dependence of young ages and low income (strong left tail dependence). After this step, the BLP estimation routine can be applied.

This paper is structured as follows: The BLP model is briefly introduced in section 4.2. Section 4.3 reviews the properties of different copula classes and goodness of fit measures. In section 4.4, the application of copula functions is demonstrated with an example from the banking literature. Eventually, results are summarized.

## 4.2 The Model

For demand estimation, the model of BLP (Berry, Levinsohn, and Pakes (1995) and Berry (1994)) is used.

Assuming that a utility maximizing agent  $i$  in market  $t$  consuming product  $j$  experiences a utility of

$$u_{ijt} = \mathbf{x}_{jt}\boldsymbol{\beta}_i + \xi_{jt} + \epsilon_{ijt} \quad , \quad (4.1)$$

the consumer will choose the product with the highest utility out of  $J_t$  choices. Product characteristics observable by the econometrician and the consumer are given by a vector of  $K + 1$  product characteristics  $\mathbf{x}_{jt}$  (price and other product characteristics). Other product characteristics exclusively observed by the consumer are denoted by  $\xi_{jt}$  and  $\epsilon_{ijt}$ .  $\xi_{jt}$  is assumed to be constant across consumers in a market  $t$  and  $\epsilon_{ijt}$  follows a Type I Extreme Value distribution. Consumers can also choose an outside option, where a utility of  $u_{i0t} = \epsilon_{i0t}$  is realized, i.e. all parameters are normalized to zero.

Specifying the distribution of  $\boldsymbol{\beta}_i$  allows to generate realistic substitution patterns and avoids the “irrelevance of independent alternatives” effect. Usually, this results in a combination of observed and unobserved heterogeneity for the  $k$ 'th random coefficient:

$$\beta_{i,k} = \bar{\beta}_k + \sum_{r=1}^R \gamma_{k,r} d_{i,r} + \sigma_k \nu_{i,k} \quad .$$

$\nu_k$  (for  $k = 1, \dots, K$ ) is assumed to be standard normal<sup>6</sup>, denoted by the PDF  $\phi(\cdot)$ , and  $R$  demographic variables  $d_r$  (for  $r = 1, \dots, R$ ) follow a multivariate PDF  $f(\cdot)$ . This gives the following utility specification:

---

<sup>6</sup>This representation implicitly assumes zero correlation between  $\beta_{i,k}$  for different  $k$ . This assumption can be relaxed without loss of generality, but increases the complexity of the optimization problem.

$$u_{ijt} = \delta_{jt} + \sum_{k=1}^K x_{jt,k} \cdot \left[ \sum_{r=1}^R \gamma_{k,r} d_{i,r} + \sigma_k \nu_{i,k} \right] + \epsilon_{ijt} ,$$

with  $\delta_{jt} = \sum_{k=1}^K x_{jt,k} \bar{\beta}_k$  as the constant part of utility across consumers in market  $t$ . The estimation of  $\delta_{jt}$ ,  $\sigma_k$  and  $\gamma_{k,r}$  is subject of the estimation problem. With dimensions  $K \times 1$  and  $K \times R$ , respectively,  $\boldsymbol{\sigma}$  and  $\boldsymbol{\gamma}$  collect all subscripted effects and define the set of parameters  $\boldsymbol{\theta}$ .

Parts of the estimation algorithm are based on equating observed and implied market shares. For a given set of parameters, the latter is defined as:

$$s_{jt}(\delta_{jt}, \boldsymbol{\theta}) = \int_{\mathbb{R}^{K+R}} \frac{\exp\left(\delta_{jt} + \sum_{k=1}^K x_{jt,k} \left[ \sum_{r=1}^R \gamma_{k,r} d_{i,r} + \sigma_k \nu_{i,k} \right]\right)}{1 + \sum_{l=1}^{J_t} \exp\left(\delta_{lt} + \sum_{k=1}^K x_{lt,k} \left[ \sum_{r=1}^R \gamma_{k,r} d_{i,r} + \sigma_k \nu_{i,k} \right]\right)} \phi(\boldsymbol{\nu}) f(\mathbf{d}) d(\boldsymbol{\nu}) d(\mathbf{d}) . \quad (4.2)$$

This integral has no analytic solution and needs to be approximated by numerical integration rules (see, for example, Train (2009) or Heiss and Winschel (2008)).

#### 4.2.1 Estimation Algorithm

The main identifying assumption is that the unobserved term  $\xi_{jt}$  is mean independent of exogenous product characteristics  $\mathbf{x}_{jt}^*$  and a set of instruments  $\mathbf{IV}_{jt}$ :

$$E(\xi_{jt} | \mathbf{x}_{jt}^*, \mathbf{IV}_{jt}) = 0 . \quad (4.3)$$

Equalizing sample moments to population moments in a GMM approach gives the following optimization problem:

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin} \xi(\boldsymbol{\theta})' \mathbf{Z} \mathbf{W} \mathbf{Z}' \xi(\boldsymbol{\theta}) . \quad (4.4)$$

Exogenous variables are denoted as  $\mathbf{Z}$ , with  $\mathbf{z}_{jt} \in \mathbf{Z}$  and  $\mathbf{z}_{jt} = \{\mathbf{x}_{jt}^*, \mathbf{IV}_{jt}\}$ , and  $\mathbf{W}$  is a weighting matrix. As pointed out by Berry (1994), under very general conditions there exists a unique set of  $\boldsymbol{\delta}$  that equals predicted and observed market shares. BLP propose a contraction mapping to invert equation (4.2) in an inner loop:

$$\delta_{jt}^{n+1} = \delta_{jt}^n + \ln \left( \frac{S_{jt}}{s_{jt}(\delta_{jt}^n, \boldsymbol{\theta})} \right) . \quad (4.5)$$

After reaching a predefined tolerance level at  $n^*$ ,  $\delta_{jt}^{n^*}$  is regressed on  $\mathbf{x}_{jt}$  with  $\mathbf{Z}$  as instruments to obtain a consistent estimate of  $\bar{\beta}_k$  for  $k = 1, \dots, K$ . With these estimates, the value of  $\xi(\boldsymbol{\theta})$  can be calculated. An optimization routine varies  $\boldsymbol{\theta}$  in an outer loop and converges to  $\hat{\boldsymbol{\theta}}$ . Note that the contraction mapping is performed at each outer loop iteration step. The minimization of the GMM objective gives asymptotically normally distributed estimates (Berry, Linton, and Pakes, 2004). A convenient implementation of

the estimation strategy is provided with the R package `BLPestimatorR` (Brunner, Weiser, and Romahn, 2017).

### 4.3 Modeling Demographics' Joint Distribution

The market share integral approximation in equation (4.2) requires a precise knowledge of the density  $f(\mathbf{d})$ . This section introduces copula functions as a tool that allows a flexible modeling of this joint distribution, if only limited information about the dependence structure is available.

Trivedi and Zimmer (2007) motivate the use of copulas for situations where either marginal distributions are known, but it is not clear how the joint distribution of random variables can be specified, or a simple concept like the correlation coefficient does not sufficiently describe the dependence. From an econometric point of view, the copula approach has many useful applications with an emphasis on financial econometrics. Asset pricing and credit risk management make extensive use of copulas to model the joint risk of different financial assets (see, for example, Cherubini, Luciano, and Vecchiato (2007)). Actuarial sciences give another area, where copulas are used to model the probability of joint events, like multiple deaths (see, for example, the work of Clayton (1978)). A comprehensive literature review is provided in Frees and Valdez (1998).

#### General Properties of Copulas

The term copula and the fundamental idea goes back to the work of Sklar (1959), Hoeffding (1940) and Hoeffding (1941) and is based on the relation between the joint distribution of variables  $d_1$  and  $d_2$  and their marginal distributions<sup>7</sup>:

$$F(d_1, d_2) = C(F_1(d_1), F_2(d_2)) = C(u_1, u_2) \quad . \quad (4.6)$$

In this representation,  $F$  denotes the joint CDF according to PDF  $f$  in equation (4.2).  $F_i$  is the marginal distribution of  $d_i$ , so  $F_1(d_1)$ , for example, is defined by  $\lim_{d_2 \rightarrow \infty} F(d_1, d_2)$ .  $C$  is unique for continuous variables<sup>8</sup>. Because  $d_i$  follows the CDF of  $F_i$ , it holds that  $F_i(d_i) = u_i \sim \text{unif}(0, 1)$ .

The modeling with copulas imposes a parametric function for  $C$ , labeled as  $C_\omega$ . Information about the dependence structure is collected in  $\omega$ , which is a scalar value for copula classes considered in this paper. This gives a parametric version of equation (4.6):

<sup>7</sup>Only due to notational simplicity, the following properties of copula functions are limited to the bivariate case.

<sup>8</sup>For discrete variables, the copula is only "uniquely determined", meaning that the copula is only defined for  $u_1 \in \text{Range}(F_1)$  and  $u_2 \in \text{Range}(F_2)$ . Otherwise, the inverse of  $F_1$  and  $F_2$  is not defined and so is the copula function.

$$F(d_1, d_2) = C_\omega(u_1, u_2) .$$

Since a copula function represents  $F$  in terms of uniform random variables, it shares the same properties as a multivariate CDF. For continuous variables,  $C$  is defined on the unit square  $[0, 1]^2$  and maps variables to  $[0, 1]$ . According to Sklar's theorem, the function is grounded, i.e.  $C(u_1, 0) = C(0, u_2) = 0$  and it holds that  $C(u_1, 1) = u_1$  and  $C(1, u_2) = u_2$ . Copulas are two increasing meaning that  $\frac{\partial C^2}{\partial u_1 \partial u_2} \geq 0$ , if second derivatives are defined. Moreover, they are bounded by the Fréchet-Hoeffding lower and upper bound functions:

$$\max(u_1 + u_2 - 1, 0) \leq C(u_1, u_2) \leq \min(u_1, u_2) .$$

For more detailed information about the theory of copulas, see Joe (1997) or Nelsen (2007).

### Copula Classes

The choice of a functional class for  $C_\omega$  has important consequences for the dependence structure that can be modeled. A favorable attribute of a copula class is to include as much as possible functional relations between Fréchet-Hoeffding lower and upper bounds to attain a wide range of possible dependence structures. The properties of four popular and commonly used copula classes are introduced in the following. The summary is based on the structure in Trivedi and Zimmer (2007).

The Gaussian copula belongs to the category of elliptical copulas and imposes the following function:

$$C_\omega(u_1, u_2) = \Phi_\omega(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) .$$

$\Phi$  is the CDF of a standard normal distribution and  $\Phi_\omega$  the normal CDF with correlation parameter  $\omega$ .  $\omega$  is bounded between -1 and 1, where each bound represents the Fréchet-Hoeffding lower and upper bound, respectively. Gaussian copulas can model a symmetric left and right tail dependence. An example of the copula's PDF with  $\omega = 0.6$  and standard normal margins is given in the upper left of figure 4.1.

Clayton, Frank and Gumbel copulas represent three prominent classes in the category of Archimedean copulas. These classes are based on a generator function  $\varphi$ , with  $\varphi : [0, 1] \rightarrow [0, \infty]$  having continuous derivatives on  $(0, 1)$ ,  $\varphi(1) = 0$ ,  $\varphi(0) = \infty$ ,  $\varphi'(t) < 0$  and  $\varphi''(t) > 0$  for  $t \in (0, 1)$ . Functions with these properties ensure the existence of an inverse  $\varphi^{-1}(t)$ , which is used for the generation of a Archimedean copula:

$$C_\omega(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2)) .$$

As shown below, the functional form of  $\varphi$  includes the parameter  $\omega$ . Copula classes in the Archimedean category choose different functions for  $\varphi$  allowing to capture very

different dependence structures.

The Clayton copula, for example, is given by:

$$C_{\omega}(u_1, u_2) = (u_1^{-\omega} + u_2^{-\omega} - 1)^{-\frac{1}{\omega}} .$$

$\omega$  ranges between 0 (independence) and  $\infty$  (upper Frechet-Hoeffding bound), i.e. only positive dependence can be modeled. Clayton copulas can only capture a strong left tail dependence. The upper right panel of figure 4.1 demonstrates this property for the copula's PDF with  $\omega = 2$  and standard normal margins.

The Frank copula takes the form:

$$C_{\omega}(u_1, u_2) = -\omega^{-1} \log\left(1 + \frac{(e^{-\omega u_1} - 1)(e^{-\omega u_2} - 1)}{e^{-\omega} - 1}\right) .$$

The Frechet-Hoeffding upper bound is attained for  $\lim_{\omega \rightarrow \infty}(C_{\omega}(u_1, u_2))$ , the lower bound for  $\lim_{\omega \rightarrow -\infty}(C_{\omega}(u_1, u_2))$  and  $\omega = 0$  represents independence. Dependence is strong in the middle of the distribution and weak in the tails. This property is visualized in the middle left panel of figure 4.1 for the copula's PDF with  $\omega = 15$  and standard normal margins.

The Gumbel copula imposes the following function:

$$C_{\omega}(u_1, u_2) = \exp(-(-\log(u_1) - \log(u_2))^{\frac{1}{\omega}}) .$$

The parameter  $\omega$  is defined between 1 (independence) and  $\infty$  (Frechet-Hoeffding upper bound), i.e. only positive dependence can be considered. Strong dependence is modeled in the right tail of the distribution as shown in the middle right panel of figure 4.1 (copula's PDF with  $\omega = 3$  and standard normal margins).

As explained above, the case of independence is a subset of all four classes. With independence, all functional relations - either asymptotically or analytically - break down to:

$$C_{\text{indep}}(u_1, u_2) = u_1 \cdot u_2 .$$

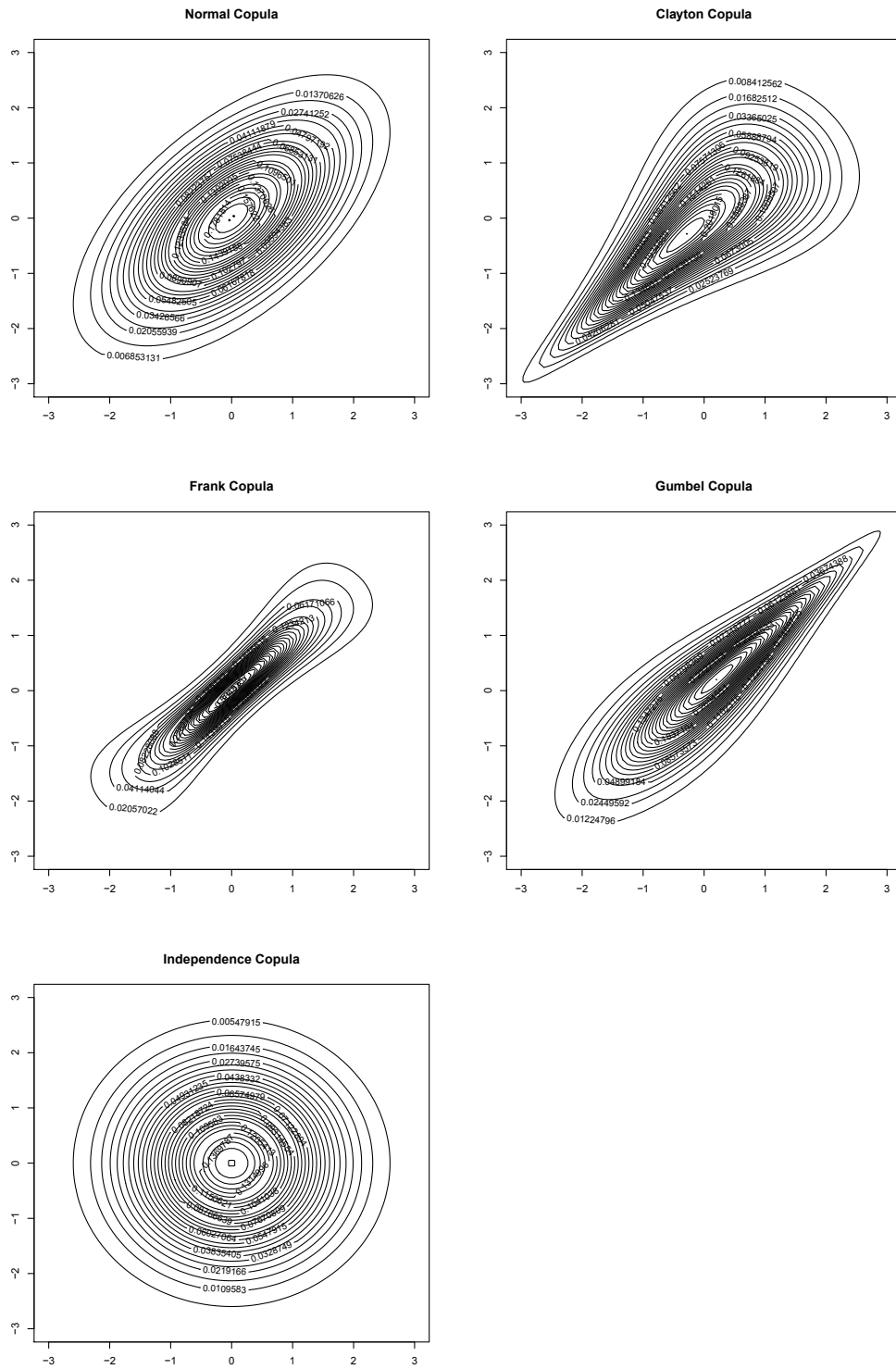
The copula's PDF is shown in the lower left panel of figure 4.1.

The following paragraphs describe the estimation of  $\omega$  and the selection between different fitted copula classes.

### Estimation of the dependence parameter

A Maximum Likelihood (ML) approach is used to estimate parameters and ensures that standard inferential results can be applied. The procedure depends mainly on how many assumptions about the marginal distributions can be made. A fully parametric approach imposes, besides the functional form for the copula, parametric marginal distributions. In

Figure 4.1: Dependence Structure of Different Copula Classes





this case, the estimation consists out of one step, because likelihood contributions can be derived directly. If marginals are not known, they can be approximated by an empirical counterpart and ML is applied in a second step. This approach is often labeled as canonical ML (CML) and discussed in Genest, Ghouli, and Rivest (1995).

In both situations, a likelihood contribution is defined as:

$$\begin{aligned} c(F_1(d_1), F_2(d_2)) &= \frac{d}{dd_1 dd_2} C_\omega(F_1(d_1), F_2(d_2)) \\ &= \frac{\partial C_\omega(F_1(d_1), F_2(d_2))}{\partial F_1(d_1) \partial F_2(d_2)} \cdot f_1(d_1) \cdot f_2(d_2) \ , \end{aligned} \quad (4.7)$$

with

$$f_i(d_i) = \frac{\partial F_i(d_i)}{\partial d_i} \ .$$

This gives the definition of a general log-likelihood function:

$$\begin{aligned} \mathcal{L}_N(\omega) &= \sum_{i=1}^N \ln(f_1(d_{1,i})) + \ln(f_2(d_{2,i})) \\ &\quad + \ln \left( \frac{\partial C_\omega(F_1(d_1), F_2(d_2))}{\partial F_1(d_1) \partial F_2(d_2)} \Big|_{d_1=d_{1,i}, d_2=d_{2,i}} \right) \ . \end{aligned} \quad (4.8)$$

Since  $f_i(d_i)$  does not depend on any parameters to be estimated by ML, it drops out of the maximization problem. For non-parametric marginals,  $F_i(d_i)$  is not known and needs to be approximated with an estimate  $\hat{F}_i(d_i)$ . In the CML approach, this can be done by the empirical CDF or a kernel density estimate, which is useful, if one is not willing to rely on parametric assumptions about the marginals. Eventually, an estimate for  $\omega$  is obtained by:

$$\hat{\omega}_{\text{CML}} = \underset{\omega}{\operatorname{argmax}} \sum_{i=1}^N \ln \left( \frac{\partial C_\omega(\hat{F}_1(d_1), \hat{F}_2(d_2))}{\partial \hat{F}_1(d_1) \partial \hat{F}_2(d_2)} \Big|_{d_1=d_{1,i}, d_2=d_{2,i}} \right) \ . \quad (4.9)$$

This procedure ensures consistent and normally distributed estimates for  $\omega$ , although the variance formula is more involved than in the standard likelihood approach<sup>9</sup>.

Empirical copulas are not capable of estimating a dependence parameter, but are often used as a baseline to compare the fit of copula classes as described in the next paragraph. The empirical copula is completely non-parametric and estimated as (Deheuvels, 1979):

$$C_N \left( \frac{i_1}{N}, \frac{i_2}{N} \right) = \frac{1}{N} \sum_{k=1}^N \mathbb{1} \left\{ \frac{\operatorname{rank}(d_{1,k})}{N+1} \leq \frac{i_1}{N}; \frac{\operatorname{rank}(d_{2,k})}{N+1} \leq \frac{i_2}{N} \right\} \ .$$

<sup>9</sup>In practice, software like the R package *copula* (Yan, 2007) easily calculates all necessary statistics.

### Goodness of fit

Evaluating goodness of fit measures provides guidance in selecting a copula class and detecting a possible misspecification. Marginal distributions are fitted non-parametrically in the present paper, so the only misspecification possible is the dependence structure imposed by the chosen copula class.

There are multiple informal and formal methods to evaluate the goodness of fit. Starting with the informal criteria, a graphical analysis to detect possible dependence structures that fit to a particular copula class and the comparison of Bayesian information criteria based on the likelihoods are a good guidance for choosing from multiple copula classes (see, for example, Trivedi and Zimmer (2007)).

Another informal method, proposed by Ane and Kharoubi (2003), evaluates the “distance” between the estimated copula of a given class and the empirical copula by

$$\text{distance} = \sum_{i=1}^N (C_N(\mathbf{u}) - C_{\hat{\omega}}(\mathbf{u}))^2 \quad ,$$

with  $C(\mathbf{u}) = C(\hat{F}_1(d_{1,i}), \hat{F}_2(d_{2,i}))$ , such that the class with the smallest distance gives optimal choice.

Another approach is to formally test, if the chosen copula is appropriate. An unfavorable property of this approach is that observed data have to meet demanding requirements in order to support the hypothesis of a correctly specified model. If the data do not fit in all considered copula classes, all specified models are rejected making it hard to select at least the best approximation to reality. Moreover, due to the unknown acceptance error not rejecting the null does only provide weak evidence for a correctly specified model. Given these drawbacks, formal tests are less common in the applied econometric literature and disregarded in the paper’s empirical application. For the sake of completeness, however, the most common test is introduced in the following.

Genest, Remillard, and Beaudoin (2009) review different goodness of fit tests based on the null that a copula function belongs to the chosen class with an estimate for  $\omega$  :

$$H_0 : C \in C_{\hat{\omega}} \quad .$$

One common approach is the calculation of a Cramer-von Mises test statistic given by

$$S_N = N \int_{[0,1]^2} (C_N(\mathbf{u}) - C_{\hat{\omega}}(\mathbf{u}))^2 dC_N(\mathbf{u}) \quad .$$

If the empirical copula is much different from the theoretical distribution under  $H_0$ , large test statistics induce a rejection of the null. Genest and Remillard (2008) prove the test to be consistent, i.e. the null is rejected with probability one, if  $C \notin C_{\hat{\omega}}$ . The limiting distribution of  $S_N$  depends on the copula class and the true parameter  $\omega$ , so

bootstrapping is used to obtain the respective p-value. The procedure is described in appendix A of Genest, Remillard, and Beaudoin (2009).

## 4.4 Empirical Application

### 4.4.1 Modeling the Demand of Deposits

Structural models of demand in commercial banking provide a framework for the analysis of policies. A popular example and subject of research<sup>10</sup> is the Riegle-Neal Interstate Banking and Branching Efficiency Act in 1994, which, among other things, prohibited interstate branching. As a consequence, a massive number of bank mergers and acquisitions reduced the total amount of commercial banks from 11,000 to 8,000 banks in this time period. At the same time, the amount of bank branches increased by about 10,000 up to 65,000 branches across the US<sup>11</sup>. Implications of the banking industry's reactions for consumers can be twofold: On the one hand, positive effects like the increased availability of banking services or a higher efficiency increases consumers' welfare. On the other hand, banks can use the weaker degree of competition to strategically increase prices. Modeling and estimating the demand for bank products allows to validate different theories empirically.

The depository consumer can choose out of  $j = 0, \dots, J_t$  banks, so every bank represents a differentiated product. To be available in the choice set, a bank needs at least one branch in market  $t$ . The definition of a market is varied in the empirical example and explained in section 4.4.2. The choice of a particular bank affects its observed amount of deposits, which includes checking, savings and time deposits accounts<sup>12</sup>.

Consumers' utility is specified as

$$u_{ijt} = p_{jt}\alpha_i + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt} + \epsilon_{ijt} . \quad (4.10)$$

$p_{jt}$  denotes the deposit rate of bank  $j$  in market  $t$ . Other variables and their effects are collected in  $\mathbf{x}_{jt}$  and  $\boldsymbol{\beta}$ , which are listed in table 4.4 (appendix 4.A).

To include the demographic variables age and income, the deposit rate effect is modeled as a random coefficient:

$$\alpha_i = \bar{\alpha} + \gamma_{\text{income}} d_{i,\text{income}} + \gamma_{\text{age}} d_{i,\text{age}} .$$

<sup>10</sup>See, for example, Dick (2008), Ho and Ishii (2011) or Beck, Levine, and Levkov (2010).

<sup>11</sup>For more information, see [www5.fdic.gov/hsob/HSOBRpt.asp](http://www5.fdic.gov/hsob/HSOBRpt.asp) .

<sup>12</sup>Deposits are only provided as an aggregate across depository services and consumer groups (households and businesses). While this is a limitation of the data, there is empirical evidence that the majority of consumers chooses one bank for all depository services (Amel and Starr-McCluer, 2001) and different consumer groups behave similarly (based on data from the Survey of Consumer Finances and Survey of Small Business Finances).

The model is completed by the definition of an outside good, which is computed from the deposits of thrifts and credit unions. While being straightforward to compute, this approach cannot account for consumers choosing no depository institution at all. A realistic but technically more demanding approach would be the consideration of micro moments in the estimation procedure as in Ho and Ishii (2011), which is left for future versions of this paper.

Treating just the deposit rate as a random coefficient and excluding unobserved heterogeneity is a strong simplification of a more realistic model and serves as an example for considering multiple demographics with the copula approach.

Service fees and deposit rates are treated as endogenous due to their relation to unobserved service components. Other product characteristics and the set of instruments define  $\mathbf{Z}$  in equation (4.3). Following Dick (2008), instruments comprise cost shifters including labor, operating costs, funding and environmental variables. The complete list of instruments is given in table 4.5 (appendix 4.A).

#### 4.4.2 Geographic Market Definitions

The definition of a market has important economic implications for the demand model. The assumption that the term for vertical product differentiation  $\xi_{jt}$  in equation (4.10) is identical for all consumers in market  $t$  directly depends on an appropriate market definition. As a consequence, markets are defined such that consumers in it are as homogeneous as possible.

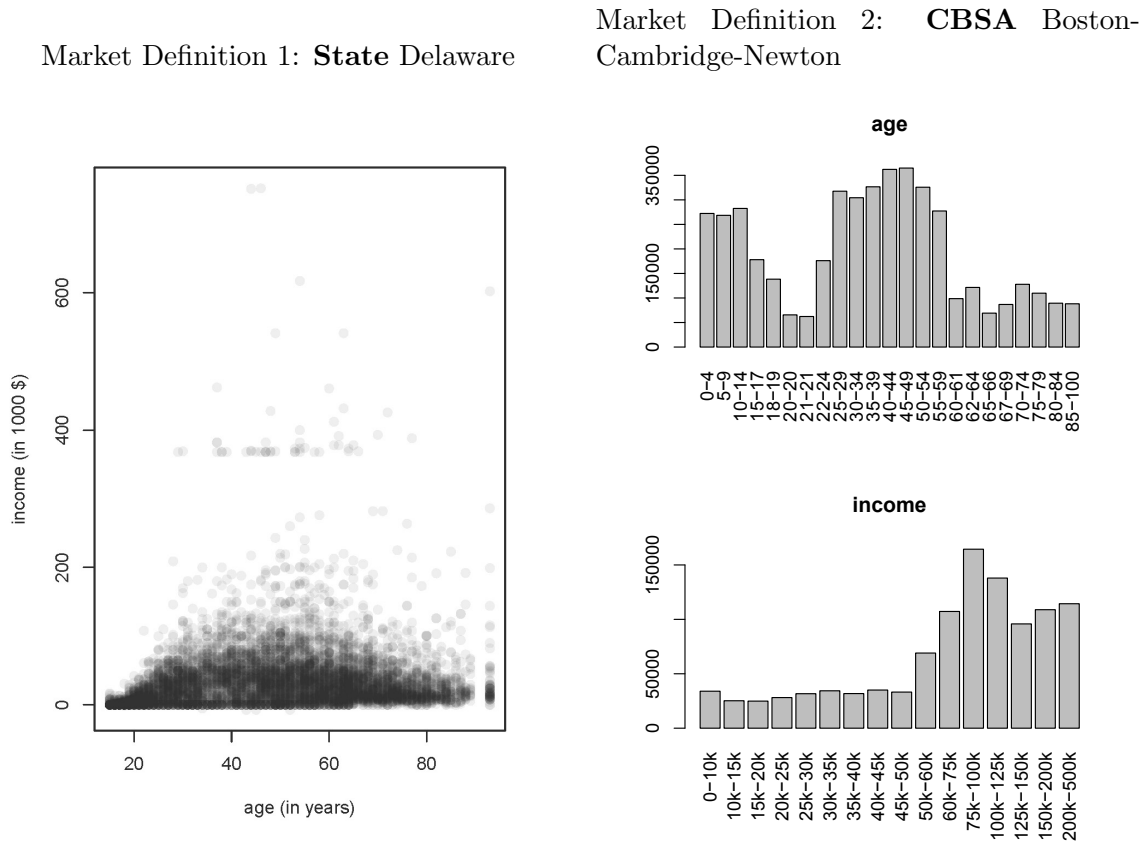
Two market definitions are used to demonstrate the consequences of different data qualities on the consideration of the demographics' dependence structure. The broad market definition uses US states to define a market (state level). One advantage of the data available on this level is its information about age and income's joint distribution. Therefore, the draws for the market share approximation can be generated directly from the empirical distribution. The left panel of figure 4.2 demonstrates the data quality for Delaware.

The second market definition uses a CBSA<sup>13</sup> as a market (CBSA level). This includes all Micro- and Metropolitan statistical areas, i.e. regions with a high degree of social and economic integration and with 10,000 and more individuals. Results in Dick (2008) are based on a similar market definition. To protect privacy, available data on this level comprises only the variables' marginal distributions. This prevents from computing market shares directly, since  $f(\mathbf{d})$  in equation (4.2) is unknown and requires further modeling as described in section 4.4.4. The right panel of figure 4.2 demonstrates the data quality for the CBSA Boston-Cambridge-Newton.

---

<sup>13</sup>A CBSA is defined geographically and conceptually by the Office of Management and Budget.

Figure 4.2: Data Quality in Different Market Definitions



### 4.4.3 Data

Multiple sources of data have to be merged to generate variables for the demand estimation. The summary of deposits (SOD) files provide information about deposits in each bank branch, credit union and thrift, which is the basic component of market shares. They also contain longitudes and latitudes allowing the mapping of branches to different markets. The variable IDSSR indicates to which bank a branch belongs. The latter is used to match bank characteristics included in the credit report (CR) with their branches. Basically all banks are legally obliged to file in these CRs every quarter and SOD half-yearly<sup>14</sup>. A CR contains balance sheet information like interest rates and fees, but also lots of other variables that might affect consumers' decisions. Bank age is taken from the National Information Center (NIC) and is also matched by IDSSR. Local labor cost on the state level is provided by the Bureau of Labor Statistics<sup>15</sup>. The housing price index on the state level is available at the Federal Housing Finance Agency<sup>16</sup>.

<sup>14</sup>Since SOD has to be reported on 6/30, CRs are used only from this date.

<sup>15</sup>For more information, see <https://www.bls.gov/oes/tables.htm> .

<sup>16</sup>For more information, see <https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx#mpo> .

To determine the CBSA of a bank branch, Topologically Integrated Geographic Encoding and Referencing (TIGER) files are used. This database contains, among other geographic layers, polygons that define a CBSA and allows to check whether a bank branch lies in the CBSA or not.

The source of demographic data depends on the market definition. The broad market definition allows to use PUMS micro data, which includes, besides age and income, much more demographic information. PUMS data is available for different geographic layers: regions, divisions, states and PUMA. The latter defines the finest geographic subdivision of the micro data as regions with at least 100,000 individuals. On the state level, PUMS provides data on 50 states and Washington D.C.. Demographic data for the 948 (in 2009) CBSAs in the second market definition is provided by the ACS summary files and is grouped into sequences that represent a category of variables<sup>17</sup>. This tabulated data gives information about how many individuals fall into predefined categories. One advantage of ACS summary files is their availability on a very fine geographic level (the finest level are block groups with only 600 to 3,000 individuals). However, even on the CBSA level, data is only provided as a 5-year average to protect the privacy of individuals and there are only information about the marginal distributions.

In the case of missing or incomplete values in any of these data sources, bank branches are deleted from the choice set. This includes branches with zero deposits or missing longitudes and latitudes, not matched bank branches with the CR data or any missing CR data. For the year 2009, this adds up to 15% of all SOD data.

Descriptives and further information about resulting variables that are based on the mentioned data sources are given in tables 4.1, 4.2, 4.4 and 4.5.

#### 4.4.4 Results

##### Dependence Modeling

Modeling the dependence structure is required for an analysis based on the second market definition, because marginal distributions of age and income are the only information available in a CBSA. Their joint distribution is obtained in a two-step procedure: First, micro data of the state, which comprises the CBSA, is used to estimate the dependence parameter with the copula approach. Second, this parameter and the corresponding copula class are used to tie the marginal distributions on the CBSA level together, which results in a semi-parametric joint distribution. Obviously, this procedure is based on the assumption that the dependence parameter is identical for each CBSA in a given state.

The estimation of the dependence parameter is performed for every state and for the four copula classes discussed in section 4.3. Estimates are based on the PUMS data set. Age information of individuals younger than 15 years is excluded from the estimation, since

---

<sup>17</sup>The relevant first three digits of the geographic identifier to extract data on the CBSA level are 311.

Table 4.1: Descriptive Statistics of Demographic Data

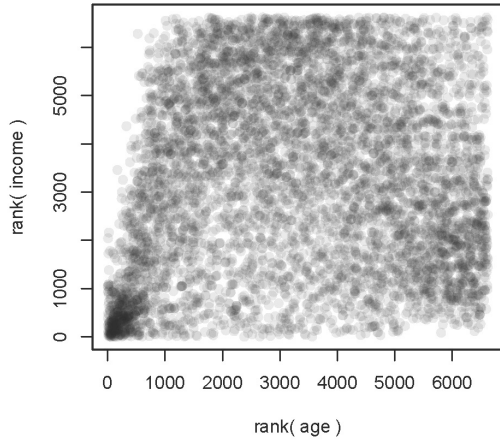
variable	state level				CBSA level		
	Mean	Min	Max	Std. dev.	Median	Min	Max
age	47.37	15	95	19.55	30 - 34	0 - 4	85+
income	32,769	-14,400	1,471,000	48,091.90	50k - 60k	< 10k	200k - 500k
source	PUMS				ACS		
geogr. units	51				948		
observations	2,560,888				286,175,684		
data quality	micro data				discretized marginal distributions		

Table 4.2: Descriptive Statistics of Product Level Data

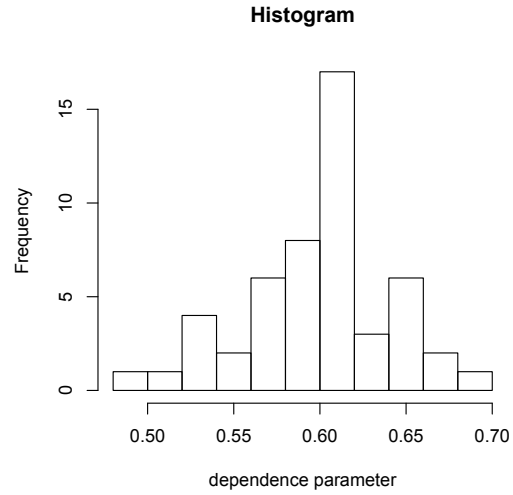
variable	state level				CBSA level			
	Mean	Min	Max	Std. dev.	Median	Min	Max	Std. dev.
share	0.007	0.000	0.654	0.029	0.046	0.000	0.786	0.077
outside share	0.217	0.017	0.562	0.111	0.077	0.000	0.904	0.147
service fees	0.002	0.000	1.355	0.018	0.004	0.000	1.355	0.020
deposit rate	0.000	0.000	0.008	0.001	0.000	0.000	0.008	0.000
employees per branch	43.839	0.000	26,216	671.162	47.418	0.000	26,216	706.246
branch density	0.001	0.000	0.508	0.013	0.002	0.000	0.197	0.005
bank age	4,008	45.429	12,166	2,381	4,347	125.429	12,166	2,527
number of states	2.119	1.000	34.000	4.163	5.280	1.000	34.000	8.143
medium size	0.020	0.000	1.000	0.139	0.085	0.000	1.000	0.279
large size	0.019	0.000	1.000	0.137	0.087	0.000	1.000	0.282
local labor cost	11.596	9.690	14.080	0.784	11.704	9.690	14.080	0.778
housing price index	44,292	29,120	95,354	12,360	45,959	29,120	95,354	13,511
non interest expenses	0.002	-0.005	0.041	0.001	0.002	-0.005	0.037	0.002
other funding cost	0.001	0.000	0.018	0.001	0.001	0.000	0.018	0.001
non performing loans	0.026	0.000	0.458	0.034	0.029	0.000	0.458	0.032
unused credit lines	0.013	0.000	6.338	0.149	0.035	0.000	6.338	0.152
holding company	0.200	0.000	1.000	0.400	0.357	0.000	1.000	0.479
equity over assets	0.111	-0.118	0.985	0.067	0.109	-0.087	0.985	0.063
rural	0.419	0.000	1.000	0.493	0.487	0.000	1.000	0.500
observations	6,045				5,413			

Figure 4.3: Visualization of Empirical and Parametric Dependence Structures

Joint Distribution of Age and Income in Ranks (Delaware)



Dependence Parameters Across States



Note that plotting the relation between two variables in ranks visualizes the dependence structure while disregarding marginal distributions.

PUMS data on income is only available for individuals older than 15 years. The marginals on the CBSA level are based on the ACS with the following adjustments: (i) age categories from 0 to 14 years are deleted, because the dependence parameter from PUMS includes no information about this part of the distribution. (ii) A uniform distribution within the ACS categories, age and income, is assumed, which seems to be realistic due to the fine category steps. For the 71 cases where a CBSA belongs to multiple states, the dependence parameter from the state with the CBSA's largest share of individuals is taken.

The left panel of figure 4.3 shows a strong dependence in the left tail of the distribution (for the example of Delaware). This typical characteristic is easily modeled by the Clayton copula, which turns out to be the best fitting copula class in all states and for both informal selection criteria. The weak negative dependence in the right half of the scatterplot, however, cannot be modeled by this copula class<sup>18</sup>. This pattern is observable in other states as well. The right panel of figure 4.3 shows the heterogeneity of dependence parameter estimates across states. All estimates are significant on a 1% level.

Eventually, market share integrals in the demand estimation on the CBSA level are based on draws from the copula function including  $\hat{\omega}$  and the CBSA's marginal distributions.

<sup>18</sup>This is probably the reason, why formal tests reject the null easily. A more realistic model, going beyond the scope of this paper, would be a mixture model that mixes two or more copula classes. This would comprise the Clayton class and another one that can capture a weak negative dependence in the right tail of the distribution.



Table 4.3: Estimation Results

variable	state level		CBSA level		
	LOGIT(i)	RC(ii)	LOGIT(iii)	LOGIT(iv)	RC(v)
<b>linear coefficients</b>					
service fees	-134.2793 (14.8830)***	-266.0855 (234.1008)	-124.8181 (11.2583)***	-82.0777 (23.6587)***	-246.8365 (233.5111)
deposit rate	647.8067 (371.5792)*	-4435.4268 (25924.6837)	3066.3751 (466.0068)***	105434.8800 (14972.4516)***	-2321.2971 (19207.3379)
deposit rate $\times \overline{\text{age}}_j$				-1941.6366 (294.3509)***	
deposit rate $\times \overline{\text{income}}_j$				-0.2616 (0.0273)***	
employees per branch	0.0002 (0.0001)***	-0.0001 (0.0001)	0.0001 (0.0001)*	0.0001 (0.0001)	-0.0001 (0.0001)
branch density	10.7254 (2.8083)***	13.7574*** (2.3777)***	18.7258 (9.3937)**	32.7742 (9.4786)**	-0.9932 (7.0646)
bank age	0.0000 (0.0000)*	0.0001 (0.0001)	0.0002 (0.0000)***	0.0001 (0.0000)***	0.0003 (0.0002)**
number of states	0.1944 (0.0215)***	-0.0545 (0.0977)	0.0802 (0.0147)***	0.0747 (0.0144)***	-0.0950 (0.0412)**
size medium	0.6646 (0.3747)*	0.9347 (1.1605)	1.0470 (0.2584)***	0.5583 (0.2472)**	0.9865 (0.7602)
size large	-0.4101 (0.5465)	2.6355 (1.9506)	0.8043 (0.3840)**	0.1961 (0.3797)	2.7642 (1.4896)*
<b>nonlinear coefficients</b>					
$\gamma_{age}$		0.2772 (2.5703)			0.8817 (1.8025)
$\gamma_{income}$		-0.0060 (0.0010)***			-0.0023 (0.0008)***
<b>other information</b>					
Wald statistic		51.443***			7.88**
observations	6045	6045	5413	5413	5413
markets	51	51	367	367	367
GMM( $\theta^*$ )	13726.75	11572.4	11132.79	15495.46	4327.88

Heteroskedastic and autocorrelation robust standard errors are in parentheses. Significance levels are denoted by 1%(\*\*\*), 5%(\*\*) and 10%(\*).  $\gamma_{income}$  and  $\gamma_{age}$  are related to the random coefficient of deposit rate. The Wald statistic tests  $H_0 : \gamma_{income} = \gamma_{age} = 0$ . To avoid overflow problems, all variables are scaled in the estimation algorithm. All estimates are rescaled and can be interpreted with respect to their original unit.

### Demand Estimation

Table 4.3 lists the results of all demand estimation settings. Columns (i) and (ii) use PUMS state level data and (iii) to (v) are based on the ACS data. For both market definitions, the first setup is a standard logit model without demographics. Setting (ii) serves as an example for the incorporation of demographic micro data on the state level in a random coefficient model, but is of limited economic use due to the broad market definition. The logit model on the CBSA level, (iii), confirms the findings of Dick (2008)<sup>19</sup>: service fees are valued negatively by the depositor and interest rates have a positive effect on indirect utility.

Setting (iv) makes use of the variation in average demographics across markets, which is interacted with the deposit rate. In markets with wealthier and older consumers (on average), deposit rates become less important.

The random coefficient model in (v) depends on the subsequent dependence modeling described in the former section. Results show that individuals have a lower marginal valuation of the deposit rate as their income increases and age decreases. Surprisingly, the deposit rate has no significant effect on the average consumer. Demographics' joint significance predicts bank products with similar deposit rates to be better substitutes than in the logit model.

Note that many other effects in the random coefficient models are insignificant. Insignificance of  $\gamma_{age}$  is probably caused by a low variation of the age distribution across geographical units and may be fixed by including additional years. The explanatory power of many linear variables in the logit model is absorbed by the consumer heterogeneity, which could be a result of the simplistic nature of the model.

Estimates are obtained by the R package BLPestimatorR (Brunner, Weiser, and Romahn, 2017). The demand estimation is performed with an inner and outer loop stopping criterion of  $1 \times 10^{-10}$ . Zero vectors serve as an initial starting value for  $\delta_{jt}$  and  $\theta$ . Market shares are approximated with 10,000 draws from the available demographic joint distribution. Second derivatives were checked to verify that each point is a minimum. Heteroskedastic and autocorrelation robust standard errors are used to compute significance levels<sup>20</sup>. Settings in (iii) to (v) are only based on 367 out of 948 CBSAs, because others do not contain any thrift or credit union branches resulting in zero outside shares.

<sup>19</sup>See setting (iv) (table 2) in her paper.

<sup>20</sup>Note that for settings based on the copula approach, standard errors do not take into account the variance of the dependence parameters. Solving the problem by a bootstrap procedure, however, is problematic due to computational cost and assumed to be negligible due to the availability of extensive micro data.

## 4.5 Conclusion

The present paper proposes copula functions to model the joint distribution of demographic variables for estimating BLPs' random coefficient model of differentiated products. This modeling becomes necessary, if, for example, providers of such data protect individuals' privacy by publishing only demographics' marginal distributions. The approximation of market share integrals within the model of BLP, however, requires the knowledge of the demographics' joint distribution.

Copula functions impose a parametric dependence structure on the joint distribution and are flexible in considering marginals. Four popular copula classes are introduced to capture different dependence structures. Once the copulas' parameters are estimated and a class is chosen, the demand estimation is straightforward. The only difficulty in the example above is to get demographic data that allows to extract the dependence structure and is related to the marginal distributions of interest. This paper takes advantage of the nested structure of different geographic layers: on a broad layer, micro data is available to estimate a dependence structure, which then is used on a fine level to combine the provided marginals. The procedure is illustrated by an application from the empirical banking literature.

In a real data example, two market definitions are used to estimate the demand for deposits in commercial banks and to demonstrate the implications of different data qualities. Micro data on the state level can be directly used to generate draws and approximate market share integrals. On the CBSA level, however, marginal distributions are the only information that is provided, which requires the researcher to model the dependence structure separately. Copulas are used to extract the latter from the respective state and tie marginal distributions on the geographic level of a CBSA together. For all states, the Clayton copula provides the best fit to the data predicting a strong left tail dependence between age and income. Using this result in the demand estimation predicts bank products with similar deposit rates to be good substitutes.

Modeling dependence as proposed in this paper allows to use realistic market definitions and contributes to reliable economic conclusions.

**Appendix 4.A Variable Description**

Table 4.4: Demand Estimation Variables

Variable Name	Definition	Data source
market share (state)	deposits of a commercial bank in a market/ total deposits in a market	DEPSUMBR, STALPBR (SOD)
market share (CBSA)	deposits of a commercial bank in a market/ total deposits in a market	DEPSUMBR (SOD), TIGER
service fees	service charges on deposit accounts/ total deposits	RIAD 4080, RCON 2200 (CR)
deposit rate	interest expenses on deposits/ total deposits	RIAD 4508, RCON 2200 (CR)
employees per branch	total number of employees per bank/ total amount of branches per bank	RSSDID (SOD), RIAD 4150 (CR)
branch density	number of branches per bank/ state square miles	RSSDID, STALPBR (SOD), Geography data (Census Bureau)
bank age	weeks between today (=2017) or closing date and opening date	DT_OPEN, DT_EXIST_TERM (NIC)
number of states	number of states in which the bank has presence	RSSDID, STALPBR (SOD)
medium size	total assets between 100 and 300 million dollar	RCON 2170 (CR)
large size	total assets larger than 300 million dollar	RCON 2170 (CR)
demogr. age (state)	distribution of age per state	ST, AGEP ( PUMS)
demogr. age (CBSA)	marginal distribution of age per CBSA	Sequence 10 (ACS)
demogr. income (state)	distribution of demographic income per state	ST, AGEP ( PUMS)
demogr. income (CBSA)	marginal distribution of demographic income per CBSA	Sequence 58 (ACS)

Table 4.5: Instrumental Variables

Variable Name	Definition	Data source
local labor cost	mean hourly wage of bank tellers per state	H_MEAN (OES Survey), STALPBR (SOD)
housing price index	index value per state	HPI (FHFA), STALPBR (SOD)
non interest expenses	noninterest expenses related to the use of premises, equipment, furniture, and fixtures / total assets	RIAD 4217 , RCON 2170 (CR), RSSDID (SOD)
other funding cost	other funding costs / total assets	RIAD 4180 , 4185, 4200, RCON 2170 (CR), RSSDID (SOD)
non performing loans	non-performing loans / total assets	all RC-N items, RCON 2170 (CR), RSSDID (SOD)
unused credit lines	unused credit lines / total assets	RCFD 3814, 3815, F164, F165 , 3817, 6550 , RCON 2170 (CR), RSSDID (SOD)
holding company equity over assets	“1” if bank is owned by a holding equity / total assets	HCTMULT (SOD) RCON 3210, 2170 (CR), RSSDID (SOD)
rural	“1” if bank operates in rural area	METROBR, MICROBR (SOD)



## Bibliography

- Amel, D. F. and M. Starr-McCluer (2001). *Market definition in banking: recent evidence*. Finance and Economics Discussion Series 2001-16. Board of Governors of the Federal Reserve System (U.S.).
- Ane, T. and C. Kharoubi (2003). “Dependence Structure and Risk Measure”. In: *The Journal of Business* 76.3, pp. 411–438.
- Beck, T., R. Levine, and A. Levkov (2010). “Big Bad Banks? The Winners and Losers from Bank Deregulation in the United States”. In: *The Journal of Finance* 65.5, pp. 1637–1667.
- Berry, S. (1994). “Estimating Discrete-Choice Models of Product Differentiation”. In: *The RAND Journal of Economics* 25.2, pp. 242–262.
- Berry, S. and P. A. Haile (2014). “Identification in Differentiated Products Markets Using Market Level Data”. In: *Econometrica* 82.5, pp. 1749–1797.
- Berry, S., J. Levinsohn, and A. Pakes (1995). “Automobile Prices in Market Equilibrium”. In: *Econometrica* 63.4, pp. 841–890.
- (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, S., O. B. Linton, and A. Pakes (2004). “Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems”. In: *Review of Economic Studies* 71, pp. 613–654.
- Bhat, C. R. (2001). “Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model”. In: *Transportation Research Part B: Methodological* 35.7, pp. 677–693.
- (2003). “Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences”. In: *Transportation Research Part B: Methodological* 37.9, pp. 837–855.
- Broyden, C. G. (1970). “The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations”. In: *IMA Journal of Applied Mathematics* 6.1, pp. 76–90.
- Brunner, D., C. Weiser, and A. Romahn (2017). *BLPestimatorR: Performs a BLP Demand Estimation*. R package version 0.1.5.
- Bungartz, H.-J. and M. Griebel (2004). “Sparse grids”. In: *Acta Numerica* 13, pp. 147–269.
- Cherubini, U., E. Luciano, and W. Vecchiato (2007). *Copula Methods in Finance (The Wiley Finance Series)*. Wiley.
- Clayton, D. G. (1978). “A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence”. In: *Biometrika* 65.1, pp. 141–151.
- Davis, P. J. and P. Rabinowitz (2007). *Methods of Numerical Integration: Second Edition (Dover Books on Mathematics)*. Dover Publications.

## Bibliography

- Deheuvels, P. (1979). “La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d’indépendance.” In: *Académie Royale de Belgique: Bulletin de la Classe des Sciences* 65, pp. 274–292.
- Dick, A. A. (2008). “Demand estimation and consumer welfare in the banking industry”. In: *Journal of Banking & Finance* 32.8, pp. 1661–1676.
- Dubé, J.-P., J. T. Fox, and C.-L. Su (2012). “Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation”. In: *Econometrica* 80.5, pp. 2231–2267.
- Frees, E. and E. Valdez (1998). “Understanding relationships using copulas”. In: *North American Actuarial Journal* 2, pp. 1–26.
- Freyberger, J. (2015). “Asymptotic theory for differentiated products demand models with many markets”. In: *Journal of Econometrics* 185.1, pp. 162–181.
- Genest, C., K. Ghoudi, and L.-P. Rivest (1995). “A semiparametric estimation procedure of dependence parameters in multivariate families of distributions”. In: *Biometrika* 82.3, pp. 543–552.
- Genest, C. and B. Remillard (2008). “Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models”. In: *Annales de l’Institut Henri Poincaré - Probabilités et Statistiques* 44.6, pp. 1096–1127.
- Genest, C., B. Remillard, and D. Beaudoin (2009). “Goodness-of-fit tests for copulas: A review and a power study”. In: *Insurance: Mathematics and Economics* 44.2, pp. 199–213.
- Goldberg, P. and R. Hellerstein (2013). “A Structural Approach to Identifying the Sources of Local Currency Price Stability”. In: *Review of Economic Studies* 80.1, pp. 175–210.
- Hansen, L. (1982). “Large Sample Properties of Generalized Method of Moments Estimators”. In: *Econometrica* 50.4, pp. 1029–54.
- Heiss, F. (2010). “The panel probit model: Adaptive integration on sparse grids”. In: *Advances in Econometrics*, pp. 41–64.
- Heiss, F. and V. Wünsch (2008). “Likelihood approximation by numerical integration on sparse grids”. In: *Journal of Econometrics* 144.1, pp. 62–80.
- Hess, S., K. E. Train, and J. 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.
- Ho, K. and J. Ishii (2011). “Location and competition in retail banking”. In: *International Journal of Industrial Organization* 29.5, pp. 537–546.
- Hoeffding, W. (1940). “Maßstabinvariante Korrelationstheorie”. In: *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin* 5, pp. 181–233.
- (1941). “Maßstabinvariante Korrelationsmasse für diskontinuierliche Verteilungen”. In: *Archiv für die mathematische Wirtschafts- und Sozialforschung* 7, pp. 49–70.
- Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Taylor & Francis Ltd.
- Knittel, C. R. and K. Metaxoglou (2014). “Estimation of Random-Coefficient Demand Models: Two Empiricists’ Perspective”. In: *Review of Economics and Statistics* 96.1, pp. 34–59.
- Lee, J. and K. Seo (2015). “A computationally fast estimator for random coefficients logit demand models using aggregate data”. In: *The RAND Journal of Economics* 46.1, pp. 86–102.



- Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer New York.
- Nevo, A. (2000). “A Practitioner’s Guide to Estimation of Random-Coefficients Logit Models of Demand”. In: *Journal of Economics & Management Strategy* 9.4, pp. 513–548.
- (2001). “Measuring Market Power in the Ready-to-Eat Cereal Industry”. In: *Econometrica* 69.2, pp. 307–342.
- Reynaert, M. and F. Verboven (2014). “Improving the performance of random coefficients demand models: The role of optimal instruments”. In: *Journal of Econometrics* 179.1, pp. 83–98.
- Richard, J.-F. and W. Zhang (2007). “Efficient high-dimensional importance sampling”. In: *Journal of Econometrics* 141.2, pp. 1385–1411.
- Sklar, A. (1959). “Fonctions de Répartition à  $n$  dimensions et leurs Marges”. In: *Publications de l’Institut de Statistique de l’Université de Paris* 8, pp. 229–231.
- Skrainka, B. S. and K. L. Judd (2011). “High Performance Quadrature Rules: How Numerical Integration Affects a Popular Model of Product Differentiation”.
- Sovinsky Goeree, M. (2008). “Limited Information and Advertising in the U.S. Personal Computer Industry”. In: *Econometrica* 76.5, pp. 1017–1074.
- Stock, J. H., J. H. Wright, and M. Yogo (2002). “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments”. In: *Journal of Business & Economic Statistics* 20.4, pp. 518–529.
- Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Trivedi, P. K. and D. M. Zimmer (2007). “Copula Modeling: An Introduction for Practitioners”. In: *Foundations and Trends in Econometrics* 1.1, pp. 1–111.
- Weiser, C. (2016). *mvQuad: Methods for Multivariate Quadrature*. R package version 1.0-6.
- Yan, J. (2007). “Enjoy the Joy of Copulas: With a Package copula”. In: *Journal of Statistical Software* 21.1, pp. 1–21.



# Eidesstattliche Versicherung

Ich, Daniel Brunner, versichere an Eides statt, dass die vorliegende Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Düsseldorf, 09.06.2017

\_\_\_\_\_

Daniel Brunner