BENJAMIN S. SKRAINKA

THREE ESSAYS ON PRODUCT DIFFERENTIATION

THREE ESSAYS ON PRODUCT DIFFERENTIATION

Computational Tools for Applied Research, Evaluating Model Behavior, and Geographic Demand

BENJAMIN S. SKRAINKA

A Dissertation submitted to the Department of Economics in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Ph.D.)

Faculty of Social and Historical Sciences
University College London (UCL)
January 2012

DECLARATION

I, Benjamin Stephen Skrainka, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Chicago, January 2012

Benjamin S. Skrainka

ABSTRACT

This thesis develops computational and applied tools to study differentiated products. The core of the thesis focuses on Berry, Levinsohn, and Pakes's [1995] (BLP hereafter) model of differentiated products. First, I examine how polynomial-based methods for multi-dimensional numerical integration improve the performance of the model. Unlike Monte Carlo integration, these rules produce reliable point estimates and standard errors as well as increasing the accuracy and execution speed of the estimation software. Next, I conduct a large scale simulation study to investigate both the asymptotic and finite sample behavior of the BLP model using the traditional instruments formed from characteristics of rival goods and also supply-side cost shifters, which are necessary for asymptotic identification. The final part of the thesis evaluates the 2003 merger of Morrisons and Safeway by combining a discrete/continuous choice model of demand with census data to construct a geographic distribution of demand. I use this distribution to model the interaction between the location of consumers and stores, focusing on the welfare implications of the merger.

Dedicated to my parents who continue to believe in me and helped me get this far.

CONJOINT WORK

Chapter 2 is, with few alterations, the paper Skrainka and Judd [2011] which I wrote with Kenneth L. Judd, whose primary contribution was to help in framing the research question and planning the Monte Carlo experiments. I develop some of the ideas from this paper further in Chapter 3 which is solely my own work. My work as a research assistant for Beckert et al. [2009] influenced my thinking in Chapter 4, however I analyze and apply their data in a novel manner to model the geographic nature of competition in the 2003 Morrisons-Safeway merger as well as its impact on welfare. All errors are my own.

An expert is someone who through their own sad,
bitter experience has made a sizable fraction
of all possible errors in a field.

— Niels Bohr, as told to me by John A. Wheeler

ACKNOWLEDGMENTS

FUNDING

The primary funding for my research was provided by the Economic and Social Research Council (ESRC) grant RES-589-28-0001 to the Centre for Microdata Methods and Practice (Cemmap) at the Institute for Fiscal Studies (IFS). In addition, the Hoover Institution on War, Revolution and Peace, the National Science Foundation (grant 25-3455-00-0-290) and the Computation Institute at the University of Chicago provided funding and/or hosted me during my studies.

DATA

Rachel Griffith and the IFS kindly provided access to the datasets used in Chapter 4, including the Kantar Worldpanel,[1] the key dataset used to estimate consumer demand for groceries in the evaluation of the 2003 Morrisons-Safeway merger. I supplemented this data with the Institute on Grocery Distribution (IGD) data on grocery store characteristics, UK Office of National Statistics (ONS) census data, and post code data.

ADVICE AND MORAL SUPPORT

I am deeply grateful to my supervisor Lars Nesheim and Kenneth L. Judd. Their generosity has helped me to develop the knowledge, skills, and tools to become an Economist.

Ian Crawford provided sage counsel about more than just bicycles and enriched my time at Cemmap. Antonio Guarino also provided crucial advice and support at key points in my studies.

[1]This dataset was originally branded as the Taylor Nelson Sofres Worldpanel (TNS).

Conversations with many individuals including, particularly, Andrew Chesher, JP Dubé, Che-Lin Su, Günter Hitsch, Jesse Shapiro, Chris Conlon, Yeşim Orhun, and James Heckman have sharpened my thinking about Berry, Levinsohn, and Pakes's [1995] model of differentiated products (BLP), the Morrisons-Safeway merger, and applied IO in general. Todd Munson taught me about the TAO of solvers.

My fellow students at Cemmap – Konrad Smolinski, Nicolas Roys, Michaël Béggore Bret, Maysun El-Attar, Maria Casanova, Jelmer Ypma, Le-Yu Chen, and Sami Stouli – as well as the ever delightful Dan Rogger eased the PhD process, both through insightful discussions and invaluable levity when things seemed darkest. Finally, Adam Trexler and Shane Brighton were 'good wheels' who kept me a bit saner and a bit more on course with regular rides through the country lanes ... or at least around Regents Park yet again.

Lastly, Walter Beckert and Jeremy Lise, my committee, have generously taken time from their own research to help me take the next step in my career.

Of course, all errors are my own.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

AMST Asda, Morrisons, Sainsbury's, and Tesco

BLP Berry, Levinsohn, and Pakes's [1995] model of differentiated products

CC UK Competition Commission

CSD Complex Step Differentiation

DGP Data Generating Process

EL empirical likelihood

FOCs first order conditions

GOR Government Official Region

GSCOP Groceries Supply Code of Practice

HTC High Throughput Computing

IGD Institute on Grocery Distribution

LAD Limited Assortment Discounters

MC Monte Carlo

MPEC Mathematical Programming with Equilibrium Constraints

NaN Not a Number

NFP Nested Fixed Point

OA Output Area

ONS UK Office of National Statistics

PADS Petascale Active Data Store

PBS Portable Batch System

pMC pseudo-Monte Carlo

PQRS price, quality, range, and service

qMC quasi-Monte Carlo

RMSE Root Mean Squared Error

SCOP Supermarket Code of Practice

SGI Sparse Grids Integration

SQP Sequential Quadratic Programming

TNS Taylor Nelson Sofres Worldpanel

1

INTRODUCTION: COMPUTATIONAL TOOLS AND PRODUCT DIFFERENTIATION

Two themes unite this thesis: computational tools and product differentiation. In Chapter 2 and Chapter 3, I demonstrate the benefits of polynomial-based quadrature rules for numerical approximation of multi-dimensional integrals (Chapter 2) and of large-scale simulation methods to characterize the behavior of an estimator (Chapter 3). Both of these tools are applicable to a range of problems in theoretical and applied Economics. Furthermore, these methods are particularly useful for computing accurate answers to practical questions, such as evaluating welfare consequences or the properties of an estimator for realistically-sized datasets where asymptotic results may not hold. To illustrate the power of these tools, I analyze the Berry et al. [1995] model of differentiated products, now a mainstay of the Industrial Organization literature, using these methods. The final chapter of this thesis uses better data and computational resources to construct a model of the geographic interactions between supermarkets and their customers in order to evaluate the 2003 merger between Morrisons, plc, and Safeway, plc, in the United Kingdom.

Chapter 2 studies how the interaction between different quadrature rules and a solver affects estimation results. I focus on Monte Carlo and polynomial-based rules. Monte Carlo methods of one form or another have become the most popular technique for approximating integrals after a series of research on simulation-based estimators (E.g. McFadden 1984, 1989, Pakes and Pollard 1989). These methods are appealing because they are easy to implement and can be justified by familiar, statistical arguments. I show that polynomial-based quadrature methods such as monomial rules [Stroud, 1971] are better in all respects: they are typically a factor of 10 more accurate for a given number of nodes and a factor of 10 less costly to compute for a given accuracy. I use Monte Carlo experiments to show the consequence of using pseudo-Monte Carlo (pMC) integration rules instead of monomial rules in the context of the BLP model. The extra accuracy of the monomial rules produces much more reliable point estimates and larger standard errors when a parameter is not precisely estimated. pMC methods can mask identification problems because pMC rules introduce false local optima with artificially high curvature due to simulation error which produce artificially tight standard errors. In addition, simulation studies show, using the infrastructure from Chapter 3, that polynomial-based rules have much less bias than pMC rules. Finally, polynomial-based rules are almost as easy to use as pMC methods – all that is required is a table lookup or a call to a library function. Consequently, there is no adverse tradeoff to using better quadrature methods.

Chapter 3 computes the finite sample bias of the BLP estimator on a much larger scale than any previous Monte Carlo studies of the BLP model, including Berry et al. [2004b] and Armstrong [2011]. This research shows that modern High Throughput Computing (HTC)[1] enables simulation[2] of complex problems.[3] In order to harness the power of HTC, I develop a robust implementation of the BLP model which uses current best practice for optimization [Su and Judd, 2010] and quadrature rules (Chapter 2). To make the experiments realistic, I generate data from a fully-specified structural model. I compute several measures of bias for different numbers of markets, ranging from 1 to 50, and numbers of products, ranging from 12 to 100. In addition, I compare the performance of two different types of instrumental variables: the characteristics of rival products (the traditional 'BLP' instruments) and supply-side cost shifters. I find that for these experiments, the BLP instruments produce biased point estimates and elasticities. Cost shifter IV provides a small improvement, but the results show that even for 50 markets and 100 products asymptotics have not yet begun to take effect. In short, applied practitioners should exercise considerable care to reduce bias before basing welfare calculations and policy decisions on this model. Further research is required to find better ways to estimate demand for differentiated products.

[1] High Throughput Computing provides massive computational power for jobs which are readily parallelized by gathering many low-powered CPUs – typically no more powerful than a desktop CPU – into nodes which are in turn linked by fast interconnects. This allows a complex problem to be accelerated by breaking it up into chunks which can be run simultaneously on many processors. High Performance Computing (HPC) is similar to HTC but focuses on providing massive computing power on much shorter timescales – i.e., days or weeks instead of months or years.

[2] The term *simulation* is potentially confusing because it could refer either to a variety of pseudo- or quasi-Monte Carlo rules for approximating numerical integration or to the use of Monte Carlo experiments to study problems which are otherwise too costly or complex to study in another manner. The meaning of the term should be clear from the context.

[3] Natural scientists increasingly view simulation as a third method of research with equal merit to theory and experiment [U.S. National Science Foundation, 2007].

Finally, in Chapter 4, continuing the theme of product differentiation, I evaluate the welfare consequences of the 2003 acquisition of Safeway plc by Wm Morrisons plc by modeling how the location of stores both in geographic and characteristic space affects firms' market power. In order to compute the welfare consequences of this merger as well as counter-factual policy experiments – such as preventing the merger or allowing Tesco plc to acquire Safeway – I model the geographic distribution of consumer demand for units of groceries. I construct this distribution by combining a discrete/continuous structural model of demand with disaggregate census data. The demand estimation captures both store choice and conditional expenditure, similar to Smith [2004], however, I use the TNS Worldpanel, a high quality panel of 40,000 households' grocery purchases. I then use the geographic distribution of demand to compute the impact of different (counter-factual) policy decisions on both consumer welfare and firm profits. I find that for most regions and household types, the merger had little impact on consumer welfare because Morrisons' and Safeway's stores rarely compete in the same geographic markets.

Part I

COMPUTATIONAL TOOLS FOR APPLIED ECONOMICS

The first Part of the thesis examines several computational tools which are applicable for a wide class of problems in Economics. In Chapter 2, I demonstrate that polynomial-based quadrature methods outperform traditional simulation methods even for high dimensions. In the context of Berry et al. [1995]'s (BLP) model of differentiated products, I show that monomial rules and sparse grids integration provide both more reliable point estimates and more reliable asymptotic standard errors than Monte Carlo integration. In addition, the polynomial methods provide superior accuracy at much lower cost. Next in Chapter 3, I perform a large scale Monte Carlo experiment to study the finite sample performance (bias, Root Mean Squared Error (RMSE)) of the BLP estimator for data generated from a structural model. I find that the estimator has persistant finite sample bias for several instrumentation strategies and that the estimator's behavior is far from asymptotic for the sample sizes often encountered in practice.

2

HIGH PERFORMANCE QUADRATURE RULES: HOW NUMERICAL INTEGRATION AFFECTS A POPULAR MODEL OF PRODUCT DIFFERENTIATION[1]

Efficient, accurate, multi-dimensional, numerical integration has become an important tool for approximating the integrals which arise in modern economic models built on unobserved heterogeneity, incomplete information, and uncertainty. In this chapter, we examine the interaction between the solver and several quadrature rules. Using a series of Monte Carlo experiments, we demonstrate that polynomial-based rules out-perform pseudo-Monte Carlo rules both in terms of efficiency and accuracy for many types of integrals with standard kernels and domains, such as normal, uniform, or exponential. To show the impact a quadrature method can have on results, we compare the performance of these rules in the context of Berry et al. [1995]'s model of product differentiation (BLP hereafter) and find that Monte Carlo methods introduce considerable numerical error and instability into the computations. We find that the interaction between the solver and the quadrature rule matters: inaccurate rules produce unreliable point estimates, excessively tight standard errors, instability of the inner loop 'contraction' mapping for inverting the market share equation, and poor convergence of several of the best available solvers. Both monomial rules and sparse grids methods lack these problems and provide more accurate, cheaper methods for quadrature. We also show that polynomial-based rules are superior for high dimensional integrals.

[1]This chapter is essentially Skrainka and Judd [2011] with few modifications. Che-Lin Su and JP Dube aided my research for this chapter by graciously sharing their knowledge of BLP and their MATLAB code. This project was supported by the UK Economic and Social Research Council grant (RES-589-28-0001) to the ESRC Centre for Microdata Methods and Practice (Cemmap) and Ian Foster at the Computation Institute at the University of Chicago. Although my co-author was instrumental in developing these ideas, I wrote this paper in its entirety.

In this chapter, we demonstrate that the twin goals of computational efficiency and accuracy can be achieved by using monomial rules instead of simulation, even in higher dimensions. We compare monomial[2] rules to several popular methods of numerical integration – both polynomial-based (Sparse Grids Integration (SGI), and Gaussian product rules) and Monte Carlo – and show that monomial rules are both more accurate and often an order of magnitude cheaper to compute for a variety of integrands, including low and high order polynomials as well as the market share integrals in Berry et al. [1995]'s model of product differentiation. In addition, we also demonstrate that polynomial-based rules produce more stable estimation results because pseudo-Monte Carlo integration introduces numerical error into the BLP model which produces multiple local optima, artificially tight standard errors, instability of the inner loop 'contraction' mapping for inverting market shares, and poor convergence even with some of the best solvers currently available.

A good quadrature rule delivers high accuracy at low computational cost. High accuracy comes from either using more points and/or choosing those points more cleverly. Cost depends on minimizing evaluations of the integrand – i.e. decreasing the number of nodes. A good numerical approximation to an integral should use few nodes yet sacrifice as little accuracy as possible. Fortunately, researchers now have access to a variety of high performance quadrature[3] methods, one or more of which should suit the problem at hand. Monte Carlo methods are the primary option for very high dimensional problems, but for many multi-dimensional problems, the analyst can often obtain a cheaper, more accurate approximation by choosing a polynomial-based quadrature rule – even for ten, 15, or more dimensions.[4] Because most integrals in economics are analytic, polynomial-based methods should provide accurate, efficient numerical approximations.

[2]Monomials are the simplest possible basis for multidimensional polynomials. Each basis function is simply a product of the coordinates, each raised to some power. E.g., $x_1^3 x_2^2 x_5^1$. A formal definition follows below on page 32.

[3]Some authors (e.g. Cools [2002]) use *quadrature* to refer to one dimensional integrals and *cubature* to refer to integrals of dimension ≥ 2. We will always use quadrature to refer to any integration rule, regardless of the dimension.

[4]What actually constitutes a high dimensional problem will depend on the computing resources and numerically properties of the integral.

Our paper builds on Heiss and Winschel [2008] which showed that sparse grids integration outperforms simulation for likelihood estimation of a mixed logit model of multiple alternatives. However, we make several new contributions including that simulation introduces false local optima and excessively tight standard errors; that polynomial-based rules approximate both the level and derivatives of integrals better than pMC; that quadrature rules affect solver convergence; that polynomial rules outperform Monte Carlo for low to moderate degree monomials but both are poor for higher degrees; and that monomial rules provide a low cost alternative to SGI.

We obtain these results by comparing the performance of different quadrature rules (pMC, Gaussian-Hermite product rule, monomial rule, and sparse grids integration.) when estimating the BLP model of differentiated products for several synthetic datasets.[5] BLP's paper developed an innovative method for studying both vertical and horizontal aspects of product differentiation by using a random coefficients multinomial logit with unobserved product-market characteristics. But, the model's results depend heavily on the numerical techniques used to approximate the market share integrals: any errors in computing these integrals – and, more importantly, the gradient of the the GMM objective function – have far reaching consequences, rippling through the model, affecting the point estimates, the standard errors, the convergence of the inner loop mapping, and even the convergence of the solver used to compute the parameter estimates. Although, the original BLP papers use importance sampling [Berry et al., 1995, 2004a], an informal survey of the BLP literature shows, with few exceptions [Conlon, 2010], most BLP practitioners [Nevo, 2000b,a, 2001] use pseudo-Monte Carlo integration without any variance reduction methods. Thus, errors from inaccurate (pMC) quadrature rules could potentially affect much of the BLP literature.

In addition, the failure of a modern solver such as KNITRO Byrd et al. [2006] or SNOPT [Gill et al., 2002] often means that the Hessian is ill-conditioned or that a problem is numerically unstable. The former usually indicates that a model is not precisely estimated because the mapping from data to parameters is nearly singular to working precision. Using an inferior quadrature rule, such as Monte-Carlo, can mask these problems because the noisiness of pMC creates false basins

[5] We often refer to Monte Carlo rules as *pseudo-Monte Carlo* or *pMC* because of the *pseudo* random numbers used to generate these nodes. Quasi-Monte Carlo is an alternative, number-theoretic method. See Section 2.2.2 and Judd [1998].

of attraction where the solver converges and, thus, multiple optima. By combining a high-quality solver and quadrature rule a researcher can get early feedback that such problems may be present.

Furthermore, the logit-class of models is prone to numerical instability because the exponential function quickly becomes large. Consequently, poorly implemented code will suffer from a variety of floating point exceptions, including overflow, underflow, and NaNs[6]. These problems are exacerbated by the long tails of the normal pdf and typically cause a solver to abort. Some researchers attempt to address these issues by setting the objective function to a large constant on overflow or by using a set of draws which avoids the problem regions of parameter space. A better approach is to address the underlying problem with robust code, hand-coded derivatives, and proper box constraints.

We begin the chapter by surveying current best practice for numerical integration, explaining the strengths and weaknesses of several popular methods for computing multi-dimensional integrals. In addition, we compute several metrics to illustrate the superiority of polynomial-based quadrature rules to simulation. Next, we briefly review the BLP model of product differentiation. Then, we estimate the BLP model using monomial, sparse grids integration, and pseudo-Monte Carlo methods. After comparing the point estimates, standard errors, and convergence properties under these different rules, we find that monomial rules provide correct point estimates unlike simulation methods. Finally, we conclude.

2.2 MULTI-DIMENSIONAL NUMERICAL INTEGRATION

All quadrature methods approximate an integral as a weighted sum of the integrand evaluated at a finite set of well-specified points called *nodes*. I.e., a quadrature method approximates the integral

$$I\left[f\right] := \int_{\Omega} w\left(x\right) f\left(x\right) dx,\ \Omega \subset \mathbb{R}^{d},\ w\left(x\right) \geq 0\, \forall x \in \Omega$$

as

[6] *NaN* means 'not a number' and indicates that a floating point computation produced an undefined or unrepresented value such as ∞/∞, $\infty \cdot 0$, or $\infty - \infty$ per the IEEE-754 floating point standard. What happens when a program generates a NaN depends on the platform, typically either the process receives a signal to abort or the operating system silently handles the floating point exception and the computation produces the special floating point value NaN.

$$Q^R[f] := \sum_{k=1}^{R} w_k f(y_k), \; y_k \in \Omega,$$

where $w(x)$ is the weight function such as 1, $\exp(-x)$, or $\exp\left(-x'x\right)$ depending on the problem [Stroud, 1971]. The region of integration, Ω, is also problem dependent. $\{w_k\}$ and $\{y_k\}$ are the quadrature *weights* and *nodes*, respectively, R refers to the number of nodes (or draws), and d is the dimension of the integral. We use N below for the number of replications of each simulation experiment. Thus, $N = 100$ and $R = 1,500$ means that we computed the integral 100 times using a different set of $1,500$ draws for each replication.[7]

A simple Monte Carlo rule would set $w_k = 1/R, \forall k$, and draw y_k from a suitable probability distribution, namely $w(x)$. For example, if the integral is the mixed, or random-coefficients, logit with $\Omega = \mathbb{R}^d$, $f(x)$ the multinomial logit for some taste coefficient, and $w(x) = \exp\left(-x'x\right)$. Then, assuming a pMC rule, y_k is drawn from the normal distribution for the coefficients.[8] To use a polynomial-based rule, a researcher would simple look up the nodes and weights in a reference such as Stroud [1971] or call a library function such as those provided by Heiss and Winschel [2008].

The art of numerical integration lies in choosing these nodes and weights strategically so that the approximation achieves the desired accuracy with few points and, thus, minimal computational expense. A solution is said to be *exact* when $I[f] = Q^R[f]$: i.e., the approximation has no error. Many rules give an exact result for all polynomials below a certain degree. Because polynomials span the vector space of 'well-behaved' functions, any integral of a function which is smooth and differentiable – or better yet analytic – should have a good numerical approximation. More often, though, the approximation will not be exact. The quality of the approximation also depends on other properties of the integrand such as the presence of sharp peaks, kinks, high frequency oscillations, high curvature, symmetry, and the thickness of the tails all of which can lead to non-vanishing, high order

[7] This notation is based on Cools [2002].

[8] If you are integrating over a normal density, $\tilde{w}(u) = (2\pi|\Omega|)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}(u-\overline{u})^T \Omega^{-1}(u-\overline{u})\right)$, you must transform the problem to one with the Gaussian weighting function $w(x) = \exp(-x'x)$ by performing a change of variables based on the Cholesky decomposition. In this case, $x = C^{-1}(u-\overline{u})$ and $CC' = 2\Omega$. See Section 2.2.1.

terms in a Taylor series expansion. A good approximation, as well as minimizing error and the number of (expensive) function evaluations, should converge to the true value of the integral as the number of nodes approaches infinity [Stroud, 1971].

The two primary methods for choosing the quadrature nodes and weights are number theoretic and polynomial-based methods [Cools, 2002]. The former refers to Monte Carlo (or simulation) methods whereas the later includes product rules based on the Gaussian quadrature family of methods as well as monomial rules and sparse grids integration. In general, polynomial-based methods are both more efficient and more accurate. Heiss and Winschel [2008] warn that polynomial-based methods poorly approximate functions with large flat regions or sharp peaks, and, by extension, regions with high frequency oscillations. The later two problems are also likely to affect Monte Carlo (MC) methods as we show in Section 2.2.4. In the BLP example below, monomial rules have no trouble in the tails because of the Gaussian kernel. However, for very high dimensional integration MC rules may be the only option because the 'curse of dimensionality' makes even the most efficient polynomial rule intractable. MC methods can also be superior when integrating over irregularly shaped regions, unless there is a clever change of variables. If the integrand has kinks, jumps, or other singularities more work is usually required, such as performing separate integrations on the different sides of the kink or using an adaptive rule. Many economic applications, however, have ten or fewer dimensions and well-behaved integrands (analytic, smooth, and bounded), making these problems well-suited for polynomial-based rules.

2.2.1 *A One-Dimensional Example*

To illustrate these issues, consider a one-dimensional random coefficients multinomial logit (MNL) model. An agent i chooses the alternative $j \in J$ which yields the highest utility $U_{ij} = \alpha_i \left(\log y_i - \log p_j\right) + z_j^T \beta + \epsilon_{ij}$, where ϵ_{ij} follows a Type 1 Extreme Value distribution and the taste shock is a one dimensional random coefficient on price, $\alpha_i \sim N\left(\bar{\alpha}, \sigma^2\right)$. Because of the distributional assumption on ϵ_{ij}, the market shares conditional on type α_i are

$$s_{ij}\left(\alpha_i\right) = \frac{\exp\left[-\alpha_i \log p_j + z_j^T \beta\right]}{\sum\limits_k \exp\left[-\alpha_i \log p_k + z_k^T \beta\right]}$$

(See Train [2009] for details.). Consequently, the total market share of good j is just the expectation of the conditional market share integral for j:

$$\begin{aligned}
s_j &= \int\limits_{\Omega} s_{ij}\left(\alpha_i\right) g\left(\alpha_i\right) d\alpha_i \\
&= \int\limits_{-\infty}^{\infty} s_{ij}\left(\alpha_i\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(\alpha_i - \bar{\alpha}\right)^2\right) d\alpha_i \\
&= \frac{1}{\sqrt{\pi}} \int\limits_{-\infty}^{\infty} s_{ij}\left(\sqrt{2}\sigma u + \bar{\alpha}\right) \exp\left(-u^2\right) du \\
&\approx \frac{1}{\sqrt{\pi}} \sum_{k=1}^{R} w_k s_{ij}\left(\sqrt{2}\sigma y_k + \bar{\alpha}\right),
\end{aligned}$$

where a one-dimensional 'Cholesky' transformation was used to convert from the economic problem to the mathematical formula and $g\left(\cdot\right)$ is the normal probability density function for α_i. $\{w_k\}$ and $\{y_k\}$ are the R weights and nodes for a Gauss-Hermite quadrature rule, which is the appropriate rule for this weighting function and domain of integration. A nice feature of this type of rule is that the normal density disappears from the sum used to approximate the integral by construction of the weights.

In the following two sections, we survey the two main types of rules: Monte Carlo and polynomial-based.

2.2.2 *Monte Carlo Integration*

Monte Carlo integration is one of the most popular choices for numerical integration because it is easy to implement and conceptually simple. This method computes the integral by taking draws from an appropriate distribution and may include other techniques to increase accuracy and speed, such as importance sampling, Halton draws, and antithetic draws [Train, 2009]. In its simplest form, simulation weights all nodes equally by setting the weights $\omega_k = 1/R$, where $R = |\{y_k\}|$, and the nodes are drawn from a suitable distribution. The weight function is set to $1/R$ because the draws come from the corresponding distribution. Consequently, simulation is easy to implement and also works over irregular-shaped regions or with functions which are not smooth, even if MC methods do not always produce the most accurate approximations.

The Law of Large Numbers is used to justify MC rules: draw enough points and the result must converge to the 'truth' without bias. Unfortunately, accuracy only improves as $\sqrt{R}$ – so the number of nodes must be increased by a factor of 100 for each additional digit of accuracy. Consequently, a more sophisticated quadrature rule will usually outperform Monte Carlo for moderate-sized problems because adding well-chosen nodes improves the integral approximation more quickly than the same number of randomly-chosen points. In practice, Monte Carlo draws are created using an algorithm for generating apparently random numbers, such as Mersenne twister [Matsumoto and Nishimura, 1998], which can pass the statistical tests associated with random numbers. These numbers are known as *pseudo random* and the corresponding Monte Carlo method is known as *pseudo-Monte Carlo* (pMC) integration. Because pseudo-random numbers are not truly random, the Law of Large Numbers only applies to theoretical discussions of MC methods based on true random numbers, not the pseudo-random implementations commonly used for numerical integration. A poor random number generator can compromise results. See Judd [1998] for further discussion of the potential pitfalls.

More sophisticated methods of taking draws – quasi-Monte Carlo methods, importance sampling, and antithetic draws – remedy some of the deficiencies of simple pMC. *Quasi-Monte Carlo* (qMC) rules use a non-random algorithm which will not pass all of the statistical tests of randomness but, instead, provides better coverage of the parameter space by constructing equidistributed nodes, resulting in convergence which is often much faster than pMC methods. The weights, as in the case of pMC, are $w_j = 1/R$. In an earlier draft, we used a qMC quadrature rule with Niederreiter sequences[9] to estimate the BLP model. In theory, it should considerably out-perform a pMC rule. We chose a Niederreiter rule which produces good results for a variety of problems while retaining the simplicity of pMC rules [Judd, 1998]. In practice, we found that using even 5,000 nodes for the 5 dimensional

[9]The current standard in Economics is to use Halton draws and Kenneth Train's code, which is available on his website [Train, 1999]. Niederreiter sequences are an alternative quasi-Monte Carlo (qMC) algorithm [Judd, 1998].

integrals we consider below, qMC was not a significant improvement on pMC. Consequently, we do not discuss qMC further. Nevertheless, qMC is easy to implement and performs at least as well as pMC.

Another common mistake is to use the same set of draws for each integral. For example, in BLP, there are $J \times T$ market share integrals, where J is the number of products per market and T is the number of markets. Instead of taking $J \times T$ sets of draws ($J \times T \times R$ total draws), some researchers take only R draws and use the same R draws for each of the $J \times T$ integrals. By taking a new set of draws for each integral, the simulation errors will cancel to some extent, which can considerably improve the quality of the point estimates because the individual integrals are no longer correlated [McFadden, 1989]. [10]

In summary, the basic problems of simulation remain regardless of the simulation rule: it is noisy and can produce inaccurate results, as Berry et al. [1995] point out:

> ... we are concerned about the variance due to simulation error. Section 6 develops variance reduction techniques that enable us to use relatively efficient simulation techniques for our problem. Even so, we found that with a reasonable number of simulation draws the contribution of the simulation error to the variance in our estimates (V3) is not negligible.

2.2.3 *Polynomial-based Methods*

We compare simulation to three multi-dimensional polynomial-based rules: Gaussian product rules, sparse grids integration, and monomial rules. Often these rules are said to be *exact for degree d* because they integrate any polynomial of degree d or less without error.[11] A common example in one-dimension is the Gaussian-family of rules: with R nodes they exactly integrate any polynomial of degree $2R - 1$ or less. Consequently, polynomial rules require many fewer nodes than pMC, making them both parsimonious and highly accurate for most integrands. The higher the degree for which the rule is exact, the more accurate the approximation of the integral but the greater the cost because of the increase in nodes. The actual choice of nodes

[10] Quantifying the actual benefit of separate draws is an open research question which merits further investigation.

[11] At least, theoretically. With finite precision of arithmetic there may be extremely small errors from truncation and round off.

and weights depends on the the rule and the weighting function in the integral. For smooth functions which are well approximated by polynomials – such as analytic functions – a good quadrature rule should always outperform simulation, except perhaps in extremely high dimensions. Like MC rules, polynomial approximations of integrals converge to the true value of the integral as the number of nodes approaches infinity, i.e. $\lim\limits_{R\to\infty} Q^R[f] = I[f]$.

To use a Gaussian rule, simply determine which rule corresponds to the weighting function and parameter space of the integral in question. Then look up the nodes and weights in a table or use the appropriate algorithm to calculate them. Judd [1998] discusses the most common rules.

2.2.3.1 *Gaussian Product Rule*

The Gaussian product rule uses a straight-forward method to construct nodes and weights: compute nodes by forming all possible tensor products of the nodes and weights of the one dimensional rule which is appropriate for the integral's domain and weighting function. I.e., each of the d-dimensional node z_k's individual coordinates are one of the one-dimensional nodes. The set of nodes, then, is all possible zs which are on the lattice formed from the Kronecker product of the one-dimensional nodes. See Figure 1 in Heiss and Winschel [2008]. The weights are the product of the weights which correspond to the one-dimensional nodes. For example, consider a two-dimensional rule with one dimensional nodes and weights $\{y_1, y_2, y_3\}$ and $\{w_1, w_2, w_3\}$, respectively. Then the product rule has nodes $\mathcal{Y} = \{(y_1, y_1), (y_1, y_2), (y_1, y_3), \ldots, (y_3, y_3)\}$. The corresponding weights are $\mathcal{W} = \{w_1 \cdot w_1, w_1 \cdot w_2, w_1 \cdot w_3, \ldots, w_3 \cdot w_3\}$. And the approximation for the integral is $Q[f] = \sum\limits_{k \in \mathcal{I}} \tilde{w}_k f(\tilde{y}_k)$, where $\mathcal{I}$ indexes $\mathcal{W}$ and $\mathcal{Y}$, and $\tilde{y}_k \in \mathcal{Y}$ and $\tilde{w}_k \in \mathcal{W}$.[12] See the example code in Listing 1 for the actual algorithm.

Consequently, we must evaluate the function at R^d points to approximate a d-dimensional integral, which quickly becomes much larger than $10,000$, often a practical upper limit on the number of

[12] To be more formal, consider a set of one-dimensional nodes and weights, $\{y_k, w_k\}_{k=1}^R$. The d-dimensional product rule is the set of nodes $z_k \in \left\{\times_{m=1}^d y_{i_m}\right\}$. Let $\mathcal{C}(z_k)$ be a function which returns an ordered list of the indexes $(i_1, i_2, \ldots, i_d)$ of the one-dimensional nodes forming the coordinates of the d-dimensional vector z_k. Then each node $z_k = \left(y_{i_1}, y_{i_2}, \ldots, y_{i_d}\right)$ and has weight $w_{i_1} \cdot w_{i_2} \cdot \ldots \cdot w_{i_d}$, the product of the one-dimensional weights corresponding to the one dimensional nodes of z_k.

nodes which are feasible with current computer technology. We use the product of formulas which have the same number of nodes in each dimension – i.e. are exact for the same degree – so that we know roughly what degree polynomial can be integrated exactly [Cools, 1997]. If the formulas are not exact to the same degree, then we know only upper and lower bounds on what polynomial will integrate exactly. One problem with product rules is that some higher degree monomials will be integrated exactly because of the product between the one-dimensional bases. For example, consider a problem with three dimensions and five nodes per dimension. The one-dimensional Gaussian formula is exact for all polynomials of degree $2 * 5 - 1 = 9$, but the corresponding product rule is also exact for terms such as $x_1^9 x_2^3 x_3^1$, where x_i is the variable for the i-th dimension. Thus, product rules provide some indeterminate amount of extra accuracy, but also impose additional computational costs.

2.2.3.2 *Sparse grids Integration*

We also consider SGI which is related to the Gaussian product rules. SGI uses a subset of the nodes from the product rule and rescales the weights appropriately. The advantage of SGI is that it exploits symmetry so that it requires many fewer points, making it more efficient to compute with little or no loss in accuracy. In addition, the nodes and weights for higher levels of exactness are easier to compute for SGI than for monomial rules. We use a Kronrod-Patterson rule for choosing nodes as explained in Heiss and Winschel [2008]. Our experiments show that SGI is very competitive with monomial rules in many cases. However, when the lowest possible computational costs matter, the monomial rule is the best option because it delivers the highest accuracy with fewest nodes.

2.2.3.3 *Monomial Rules*

Monomial rules exploit symmetries even more effectively than SGI and provide very accurate approximations with surprisingly few nodes, even for moderate dimensions [Stroud, 1971, Cools, 2003]. Formally, a *monomial* in $x \in \mathbb{R}^d$ is the product $\prod_{i=1}^{d} x_i^{p_i}$ where $p_i \in \mathbb{W}$ and $\mathbb{W} \equiv \{0, 1, 2, \ldots\}$. Thus, monomials are the simplest possible basis for the set of multi-dimensional polynomials. The *total order* is just the sum of the exponents $\sum_i p_i$. $\mathbf{x}^{\mathbf{p}}$ is a compact notation which refers to the monomial $\prod_{i=1}^{d} x_i^{p_i}$. For example, $x_1 x_2^3 x_3^2$ is a three-dimensional monomial with $d = 3$, $p = (1, 3, 2)$, and total order 6. The monomial rules are constructed so that they will exactly integrate all monomials less than or equal to some total order. Monomial rules are more efficient than product rules because they do not exactly integrate any higher order terms.

The performance gains from monomial rules are clear, but the difficulty lies in computing the rule's nodes and weights. Fortunately, many efficient, accurate rules have already been computed for standard kernels and parameter spaces [Cools, 2003, Stroud, 1971]. A practitioner only needs to look up the appropriate monomial rule in a table [13,14] and can then compute the integral as the weighted sum of the integrand at the nodes. Unfortunately, if the necessary rule doesn't exist you will have to consult a specialist (For example, see Cools [1997]).

Section 2.4 provides an example of how to use Stroud monomial rule 11-1 – which is exact for degree 11 polynomials in five dimensions – to compute the BLP market share integrals.

2.2.4 *Precision and Accuracy*

While verifying that our implementation correctly approximated known integrals of standard monomials, we were surprised by the results.[15] As expected, the polynomial rules correctly integrate all monomials less than or equal to their respective degrees and produce poor results for monomials of higher degree. But despite using 100 replications of a pMC rule with $R = 10,000$ draws, pMC performed poorly for the low order monomials, with error increasing with the degree of the monomial. pMC also produced poor results for high order monomials where we expected it would outperform the polynomial rules. We

[13] Typically, a small amount of computation is required because the table will only provide each unique set of nodes and weights. A researcher must then calculate the appropriate (symmetric) permutations of the unique nodes to generate all possible nodes.

[14] A monomial rule may have several equivalent sets of nodes and weights because the system of equations used to compute the monomial rule may have multiple solutions. For example, Stroud rule 11-1 has two solutions, which we refer to as 'Left' and 'Right' after the two columns in the table which list the different solutions. The performance of these solutions will vary slightly depending on the shape of the problem.

[15] The results should be the same up to the limits of standard arithmetic errors such as truncation and round-off error.

conjecture that pMC works well only when the high-order terms in a Taylor series expansion are very small, something which is explicit in the construction of monomial rules.

These results are summarized in Table 1 which shows the difference between the theoretical value and the value computed with each rule. The first three columns are the results for the Gaussian-Hermite Product rule with 3^5, 5^5, and 7^5 nodes – i.e., 3, 5, and 7 nodes in each of the 5 dimensions; next the sparse grids rule which is exact for degree 11; then, the left and right versions of rule 11-1 in Stroud [1971], also exact for degree 11; and, finally, the two right most columns show the mean absolute error and the standard error for the pMC rule with 100 replications. The monomials are listed by increasing degree. Note that the Gauss-Hermite product rules will exactly integrate any monomial rule as long as the coordinates in each dimension are raised to some power less than or equal to $2R - 1$ where R is the number of one dimensional nodes used in the tensor product. For odd monomials, the difference in performance is even more stark: the polynomial rules are 0 to the limits of numerical precision whereas pMC has significant error, especially as the degree of the monomial increases. These results, in our opinion, considerably strengthen the case for using sparse grids or monomial rules because pMC is never better.[16]

The values for $\overline{|pMC|}$ are surprisingly large. Further investigation revealed that the set of draws in this experiment are particularly bad, despite using 10,000 draws and 100 replications. The Central Limit Theorem governs this behavior: there is always some probability of choosing a bad set of draws. Consequently researchers should check how sensitive their results are to a given set of draws.

2.2.4.1 *Bias and Noise*

When choosing which quadrature rule to use, a researcher should consider how it will affect their results. Simulation methods have become extremely popular because they are easy to implement. However simulation can suffer from both bias as well as noise. The nature of the bias depends on the type of estimator: for Method of Simulated Moments (MSM) the bias is zero unlike Maximum Simulated Likelihood (MSL) and Maximum Simulated Score (MSS). The bias occurs because the bias term is only linear for MSM: consequently, Jensen's

[16] We also computed these tests for Halton draws generated by MATLAB R2010b's qrandstream facility which did not perform significantly better than pMC.

inequality shows that MSL and MSS must be biased. The normalized bias term will approach zero asymptotically if R, the number of draws, approaches infinity faster than $\sqrt{N}$, the number of observations, i.e. $\sqrt{N}/R \rightarrow 0$ as $R, N \rightarrow \infty$. Consequently, researchers who use MSL or MSS should remember to correct for this bias. Noise, however, will disappear regardless of R as the sample size increases provided that different draws are used for each observation. See Train [2009] for further discussion.

Polynomial-rules, on the other hand, only suffer from approximation error and that to a much lesser degree than Monte Carlo methods. Thus, the error is much smaller for these methods, making them much better suited for empirical and other problems than simulation [Heiss and Winschel, 2008]. With polynomial rules, researchers can also consider more efficient econometric methods such as MLE instead of GMM.

2.2.4.2 *Approximation of the Tails of the Normal*

Polynomial-based rules approximate the entire distribution more accurately than pMC rules, especially the tails. Extremal values in the tails often cause numeric problems such as overflow at larger nodes, despite their small quadrature weights. These problems are less common with Monte Carlo methods because draws in the tails are infrequent by definition. The source of the problem is not the correct weighting of the tails, but the modeling decision to use the normal distribution for convenience. Most real-world distributions have bounded support: consequently, the problem could be solved by using a more representative distribution for the random coefficients.

$f(x)$	GH 3^5	GH 5^5	GH 7^5	Sparse Grid	11-1 L	11-1 R	$\overline{\lvert pMC \rvert}$	$\sigma(pMC)$
1	−1.1e−15	−3.3e−15	−2.1e−15	6.7e−15	5.2e−12	−9.9e−14	2e−14	0
x_1^1	−2.1e−17	1e−15	1.3e−15	2.3e−17	7.1e−15	−3.6e−15	0.0075	0.0092
$x_1^1 x_2^1$	0	−1.7e−18	−1.7e−18	2.8e−17	0	−2.2e−16	0.008	0.01
$x_1^1 x_2^1 x_3^1 x_4^1 x_5^1$	0	0	0	0	0	0	0.0086	0.011
x_1^2	−4.4e−16	−2e−15	−1.1e−15	−5.2e−14	9.6e−13	−7.9e−14	0.012	0.014
x_1^4	0	−8.9e−15	4e−15	−1.5e−13	3.9e−13	7.5e−15	0.076	0.096
x_1^6	−6	−3.6e−14	−1.8e−14	−7.6e−13	4e−13	1.9e−13	0.76	0.94
$x_2^6 x_4^4$	−18	−9.9e−14	−2.3e−13	−2.2e−12	−3.1e−13	1.8e−13	7.4	9.4
x_1^{10}	−8.6e+02	−1.2e+02	−3.2e−12	−4.9e−11	2.9e−11	2.4e−11	1.8e+02	2.1e+02
$x_1^5 x_2^4 x_3^2$	5e−16	3.6e−15	−8.9e−15	−6.9e−16	1.8e−15	0	3.9	6.8
x_1^{12}	−1e+04	−3.7e+03	−4.4e−11	−5e−10	−7.2e+02	−7.2e+02	3.2e+03	3.9e+03
x_1^{13}	0	3.1e−11	2.8e−10	−1e−11	−2.9e−11	5.8e−11	1.3e+04	2e+04
x_1^{14}	−1.3e+05	−8.1e+04	−5e+03	−7.2e−09	−3.1e+04	−3.1e+04	6.2e+04	8.2e+04
x_1^{15}	9.9e−14	−3.3e−10	−1.2e−09	−5.8e−11	2.3e−10	−4.7e−10	2.4e+05	4.3e+05
x_1^{16}	−2e+06	−1.6e+06	−2.9e+05	−3.4e+04	−8.9e+05	−8.9e+05	1.3e+06	1.8e+06
$x_1^6 x_2^6 x_3^4 x_4^2 x_5^2$	−4.3e+02	−1.4e−12	−1.1e−12	−4.3e+02	1.6e+05	1.6e+05	8.5e+02	2.6e+03
$x_1^8 x_2^6 x_3^4 x_4^2 x_5^2$	−4e+03	−6.4e−12	−1.5e−11	−4e+03	1.8e+06	1.8e+06	6.7e+03	1.9e+04
$x_1^{10} x_2^5 x_3^4 x_4^2 x_5^2$	0	1.2e−13	−1.7e−13	0	5.8e−11	5.8e−11	1.9e+04	6.5e+04
$x_1^{10} x_2^{10} x_3^6 x_4^4 x_5^2$	−4e+07	−9.6e+06	−8.9e−08	−4e+07	3e+11	3e+11	7.4e+07	2.6e+08
$x_1^{16} x_2^{12} x_3^4 x_4^4 x_5^2$	−1.9e+11	−1.6e+11	−2.7e+10	−1.9e+11	4e+14	4e+14	2.1e+11	3.3e+11

Table 1: Monomial Integration Errors: Polynomial Rules vs. pMC ($R = 10,000$)

We now quickly review the features and notation of Berry et al. [1995]'s model before examining how different quadrature rules affect estimation. BLP has become one of the most popular structural models of product differentiation because it fits empirical data well by using a flexible form which combines both random coefficients and unobserved product-market characteristics, ξ_{jt}, enabling the model to explain consumers' tastes for both horizontal and vertical product differentiation. The model produces more realistic substitution patterns:[17] the random coefficients can handle correlations between different choices, overcoming the Independence from Irrelevant Alternatives (IIA) problem that is a feature of logit models, and ξ_{jt} captures unobserved heterogeneity in product quality, preventing bias in parameter estimates from product traits which the econometrician cannot observe. Nevo [2000b] provides a detailed and accessible explanation of the model. BLP is now sufficiently established that the several recent textbooks [Train, 2009, Davis and Garcés, 2009] also cover it.

Throughout this paper, we base our notation on a simplified version of the notation in Dubé et al. [2011]. Thus, we consider T markets which each have J products plus an outside good. Each product $j \in J$ in market $t \in T$ has K characteristics, x_{jt}, price, p_{jt}, and an unobserved, product-market shock, ξ_{jt}. A market can be a time period, as in the original BLP papers on automobiles, or a city, as in Nevo's papers on ready-to-eat breakfast cereal. The shock ξ_{jt} is observed by consumers and firms but not by the econometrician. This shock captures vertical aspects of product differentiation whereas the random coefficients model horizontal differentiation: all consumers value a larger ξ_{jt} but rank product characteristics differently according to their type. Lastly, y_i is consumer i's expenditure and drops out of the model because it is not interacted with any product-specific characteristics.

BLP assume consumers are rational, utility maximizers who choose the good which maximizes their utility. Let consumer i's utility from purchasing product j in market t be[18]

[17] The substitution patterns will be incorrect if congestion in product space matters. See Berry and Pakes [2007] and, for an application where congestion matters, Nosko [2010].

[18] Some researchers specify $\log\left(y_i - p_{jt}\right)$ instead of $\left(y_i - p_{jt}\right)$ to capture income effects [Petrin, 2002].

$$U_{ijt} = V_{ijt} + \epsilon_{ijt}$$

with

$$V_{ijt} = \alpha_i \left(y_i - p_{jt}\right) + x'_{jt}\beta_i + \xi_{jt}.$$

ϵ_{ijt} is an IID, Type I Extreme value shock, which leads to a simple closed form solution for market shares, conditional on consumer types, (α_i, β_i). In practice, y_i and p_{jt} are often the logarithm of the respective quantities which ensures that the utility is homogeneous of degree zero. Similarly, the utility of choosing the outside, 'no purchase' option ($j = 0$ by convention) is[19]

$$U_{i0t} = \alpha_i y_i + \epsilon_{i0t}.$$

The coefficients α_i and β_i are type-specific 'random coefficients' – i.e., they depend on a consumer's type and are drawn from some distribution in order to capture unobserved differences in consumers' tastes:[20]

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \bar{\alpha} \\ \bar{\beta} \end{pmatrix} + \Sigma \nu_i$$

where all consumers have the same mean taste preferences $\bar{\alpha}$ and $\bar{\beta}$. The unobserved taste shock ν_i is a $K+1$ column vector (because there are K product characteristics plus price) and has distribution $\nu_i \sim P_\nu$. The Cholesky factor of the variance of the taste shock is Σ, a $(K+1) \times (K+1)$ matrix. P_ν is usually assumed to be multivariate normal. Following convention, we refer to the model's parameters as θ where $\theta = (\theta_1, \theta_2)$, $\theta_1 = \left(\bar{\alpha}, \bar{\beta}\right)$, the parameters for mean utility, and $\theta_2 = \mathtt{vech}\,(\Sigma)$, the parameters for the standard deviation of the random coefficients. Thus, θ refers to all of the parameters to be estimated.

It is convenient to partition the utility into the mean utility

[19] Berry et al. [1995] specify $U_{i0t} = \alpha_i y_i + \xi_{0t} + \sigma_{0t}\nu_{i0t} + \epsilon_{i0t}$.

[20] Some researchers also include demographic information in the random coefficients. We ignore demographics in order to focus on the numerical properties of the model.

$$\delta_{jt}\left(\xi_{jt};\theta_{1}\right) \quad = \quad x_{jt}^{'}\bar{\beta}-\bar{\alpha}p_{jt}+\xi_{jt},$$

which is the utility that any consumer type gains from choosing product j in market t, regardless of type, and a type-specific preference shock

$$\mu_{ijt} \quad = \quad \left[-p_{jt}\ x_{jt}^{'}\right]\left(\Sigma\nu_{i}\right).$$

μ_{ijt} has mean zero and captures individual heterogeneity. Some researchers permit $\alpha_i < 0$ which can produce a positive price coefficient for some consumer types. A possible solution is to assume that α_i is log-normally distributed. In other applications $\alpha_i < 0$ may make sense if price is a signal of quality.

Researchers typically assume that $\epsilon_{ijt} \sim$ Type I Extreme Value so the market shares, conditional on consumer type ν,[21] have a closed-form analytic solution, the multinomial logit:

$$s_{jt}\left(\delta\left(\xi,\theta_{1}\right)|\nu\ ;\theta_{2}\right) \quad = \quad \frac{\exp\left[\delta_{jt}+\mu_{ijt}\left(\nu\right)\right]}{\sum\limits_{k}\exp\left[\delta_{kt}+\mu_{ikt}\left(\nu\right)\right]}.$$

Then the unconditional share integrals are the just the expectation of the regular MNL choice probabilities with respect to ν:

$$s_{jt}\left(\delta\left(\xi,\theta_{1}\right);\theta_{2}\right) \quad = \quad \int\limits_{\mathbb{R}^{K+1}}\frac{\exp\left[\delta_{jt}+\mu_{ijt}\left(\nu\right)\right]}{\sum\exp\left[\delta_{kt}+\mu_{ikt}\left(\nu\right)\right]}\phi\left(\nu\right)d\nu.$$

Here $\phi\left(\nu\right)$ is the standard multivariate normal probability density function. We restrict Σ to be diagonal as in the original BLP papers.[22] The random coefficients logit can in theory model any choice probabilities given a suitable mixing distribution [McFadden and Train,

[21] Note: the consumer type, ν, is scaled by Σ, the Cholesky decomposition of the variance matrix of the random coefficients.

[22] Nevo [2001] estimates the off-diagonal elements. We expect that the advantages of monomial rules would be even greater when estimating a model with off-diagonal elements.

2000, Train, 2009].[23,24] In practice, researchers choose a normal distribution for the random coefficients because it is tractable. Fiebig et al. [2010] find that for several empirical datasets the mixing distribution is significantly non-normal. Nevertheless, logit + normal should work well as long as the real-world mixing distribution is smooth and single-peaked because the tails will not contribute much to the integral.

Historically, a nested fixed point (NFP) algorithm based on Rust [1987] is used to estimate the model: the outer loop computes the point estimates of $\hat{\theta}$ by minimizing a GMM objective function whose moments are constructed from ξ_{jt}; the inner loop solves the nonlinear system of equations equating predicted and observed shares for the mean utilities, $\delta_{jt} = \delta_{jt}\left(S_{jt}, x_{jt}, p_{jt}, \theta_2\right)$, and, hence, ξ_{jt}. The original implementation [Berry et al., 1995, Nevo, 2000b] uses a contraction mapping to perform this inversion. Consequently, $\xi_{jt} = \xi_{jt}\left(\delta_{jt}, x_{jt}, p_{jt}, \theta\right)$ is a function of the mean utilities, covariates, and θ. Thus, the researcher codes an outer loop to solve the program

$$\hat{\theta} \quad = \quad \arg\max_{\theta} \left(Z'\xi\left(\theta\right)\right)' W \left(Z'\xi\left(\theta\right)\right)$$

and an inner loop to recover δ by inverting the market share equation via the contraction mapping

$$\exp\left(\delta_{jt}^{n+1}\right) \quad = \quad \exp\left(\delta_{jt}^{n}\right) \times S_{jt}/s_{jt}\left(\delta_{jt}^{n}; \theta_2\right),$$

where S_{jt} are the observed market shares, s_{jt} the predicted market shares, and δ_{jt}^{n} the n-th iterate in a sequence which hopefully converges to the true mean utilities. Given the mean utilities, the product market shock is simply

$$\xi_{jt} \quad = \quad \delta_{jt} - \left[-p_{jt} \; x_{jt}\right] \theta_1.$$

[23] McFadden and Train [2000] is a very general result and applies to elasticities and moments as well as choice probabilities. Consequently, the mixed logit can approximate general substitution patterns to arbitrary accuracy.

[24] It is worth emphasizing that you must use the correct mixing distribution for McFadden and Train [2000]'s result to hold. Keane and Wasi [2009] provide evidence from several empirical datasets that this is almost never the case.

Berry et al. [1995] proves that this mapping is a contraction and Nevo [2000b] advocates using this exponential form to improve performance by avoiding computing logarithms which are more costly than exponentiation.[25] In addition, Gandhi [2010] shows that the market share equations are invertible for a random utility model which satisfies certain monotonicity and substitutability conditions. Reynaerts et al. [2010] demonstrates that other methods for inverting the market share equations are faster and more robust. They also discuss some of the convergence problems of the contraction mapping.

The choice of rule to compute the market share integrals affects both choice probabilities and the inversion of the market share equation because numerical errors can propagate through both of these channels. With the above GMM specification, the gradient of the GMM objective function depends on the gradient of δ, which in turn depends on the gradient of the inverse of the market share equation, $s^{-1}(S;\theta)$. Consequently, numerical errors in computing the gradient of the market share integrals also propagate, affecting both the point estimates and the standard errors. As discussed below, one big advantage of polynomial-based rules over pMC is that they provide a more accurate approximation for both an integral and its gradient.

When estimating the BLP model below, we use the same moment conditions as Dubé et al. [2011]. These moment conditions depend on the product of the unobserved product-market shock, ξ, and a matrix of instruments. The matrix of instruments consists of various products of product attributes and a set of synthetic instrumental variables which are correlated with price but not ξ. See Dubé et al. [2011]'s code for details.

In this paper, we use Mathematical Programming with Equilibrium Constraints (MPEC) to estimate the BLP model because it is faster and more robust than Nested Fixed Point (NFP) [Su and Judd, 2010]. MPEC relies on a modern, high quality solver such as KNITRO or SNOPT to solve the model in a single loop as a constrained optimization problem:

[25] In simple heuristic tests, we find that the contraction mapping has poor convergence properties, fails to satisfy the sufficiency conditions of the Berry et al. [1995]'s theorem 10% of the time, and often has a contraction rate close to or exceeding 1.

$$\begin{aligned}\max_{\theta,\delta,\eta} \quad & \eta^{'} W \eta \\ \text{s.t.} \quad & s\left(\delta\left(\xi\right);\theta\right) = S \\ & \eta = Z^{'}\xi.\end{aligned}$$

By adding the extra variable η, we improve the sparseness pattern which makes the problem easier to solve and more stable numerically.[26] Furthermore, MPEC solves for the mean utilities implicitly via the constraint that observed market shares equal predicted market shares, increasing both speed and stability.

The asymptotic and finite sample properties of BLP are still not well understood. Berry et al. [2004b] prove asymptotic normality as $J \to \infty$. Berry and Haile [2010] show the model is identified under the 'Large Support' assumption. Chapter 3 uses large scale simulations to characterize finite sample performance.

2.3.1 *Example: Computing BLP Product-Market Shares*

Given the above assumptions, the monomial (or Gauss-Hermite or sparse grid) approximation for the integral is

$$s_{jt} \quad \approx \quad \frac{1}{\pi^{(K+1)/2}} \sum_k \left\{ \frac{\exp\left[\delta_{jt} + \mu_{ijt}\left(\psi_k\right)\right]}{\sum\limits_m \exp\left[\delta_{mt} + \mu_{imt}\left(\psi_k\right)\right]} \omega_k \right\}$$

where $\left(\psi_k, \omega_k\right)$ are the nodes and weights for a suitable quadrature rule with a Gaussian kernel and $K+1$ is the dimension of ν_k.[27] The factor $\pi^{-(K+1)/2}$ comes from the normalization of the normal density. The choice of monomial rule depends on the number of dimensions of the integral, desired level of exactness (accuracy), the domain of integration, and the mixing distribution a.k.a. weighting function.

For a Monte Carlo method, the approximation is

$$s_{jt} \quad \approx \quad \frac{1}{R} \sum_k \frac{\exp\left[\delta_{jt} + \mu_{ijt}\left(\psi_k\right)\right]}{\sum\limits_m \exp\left[\delta_{mt} + \mu_{imt}\left(\psi_k\right)\right]}$$

[26] With modern solvers, the sparseness pattern and type of non-linearities are more important than the number of variables.

[27] $K+1$ for the K product characteristics plus price.

for R nodes ψ_k drawn from the normal distribution. Note that these two formulae have the same structure: a weighted sum of the integrand evaluated at a set of nodes. For a Monte Carlo method, the weight $\omega_k = 1/R$.

2.4 THE EXPERIMENTS: SIMULATION VS. QUADRATURE

We compare how pMC, Gaussian product, sparse grid, and monomial quadrature rules affect the computation of several key quantities in the BLP model. We compare how these rules perform when computing the market share integrals, point estimates, standard errors, and the unobserved heterogeneity ξ_{jt}, all of which are critical components of the model. We use a high quality solver (KNITRO or SNOPT) and algorithm (MPEC) for these experiments.[28]

2.4.1 *Experimental Setup*

Our experiments use five different synthetic datasets which we generated from unique seeds and the parameters shown in Table 2 using MATLAB's rand and randn functions. We use the same values as Dubé et al. [2011] (DFS hereafter), except that we use fewer products and markets and chose different seeds.[29] We refer to these datasets via their seeds, which we label 1 to 5. To ensure that there is some noise in each dataset, as in real-world data, we compute the 'observed' market share integrals using a pMC rule with $R = 100$ nodes.[30] Currently, these parameter values result in a market share of about 90% for the outside good, which seems reasonable for a differentiated, durable good such as an automobile. That many of the observed market shares are exceedingly small could lead to inaccuracies in the corresponding computed market shares because both types of quadrature rules can be unreliable in large regions of flatness. We only consider diagonal

[28] Historically, point estimates were computed with BLP's nested, fixed point algorithm (NFP). We use MPEC to compute our point estimates because it is much more robust and, in theory, both algorithms produce equivalent values at the global optimum [Su and Judd, 2010, Dubé et al., 2011].

[29] Their code provided the starting point for the code which we developed to explore the impact of quadrature rules on BLP. We downloaded the code from JP Dubé's website (http://faculty.chicagobooth.edu/jean-pierre.dube/vita/MPEC%20code.htm), Fall 2009.

[30] For the rest of the paper, we use R to refer to the number of draws in a Monte Carlo simulation and N as the number of replications.

PARAMETER	VALUE
J	25
T	50
$\theta_{1\cdot} \equiv \left(\bar{\beta}', \bar{\alpha}\right)$	$\begin{pmatrix} 2 & 1.5 & 1.5 & 0.5 & -3 \end{pmatrix}'$
$\theta_{2\cdot} \equiv \mathtt{diag}\,(\Sigma)^{1/2}$	$\begin{pmatrix} \sqrt{0.5} & \sqrt{0.5} & \sqrt{0.5} & \sqrt{0.5} & \sqrt{0.2} \end{pmatrix}'$
R	100

Table 2: Parameters Used to Generate Monte Carlo Datasets.

Σ to facilitate validation and to maintain consistency with most BLP papers and DFS.

Although this approach of generating data with an inaccurate quadrature rule is often the practice (See, for example, Dubé et al. [2011]), this approach introduces false 'sampling' error into the synthetic data and potentially confounds results. A better design would use the most accurate possible integration method to ensure that errors in the data and results are related to the data generating process instead of this pseudo-sampling error. Berry et al. [1995] explains how sampling error can affect results and shows that asymptotically it does not compromise their results for their data. In Chapter 3, however, we generate the synthetic data using a very accurate, SGI rule.[31]

We focus on how the interaction between the solver and each quadrature rule affects the point estimates – i.e. whether the solver could consistently and efficiently find a unique global optimum. For each dataset and quadrature rule, we compute the optimum for the following experiments: five randomly choose starts near the two-stage least squares (2SLS) logit estimate, multiple starts taken about the average of the best point estimates for $\hat{\theta}$ for the 5^5 Gauss-Hermite product rule, and multiple Monte Carlo draws of nodes for the same starting value (pMC rules only). In all cases, we compute standard errors using the standard GMM sandwich formula $\mathrm{Var}\left(\hat{\theta}\right) = \left(\hat{G}'W\hat{G}\right)^{-1}\hat{G}'W\hat{\Lambda}W\hat{G}\left(\hat{G}'W\hat{G}\right)$ where $\hat{G}$ is the gradient of the moment conditions, W the weighting matrix formed from the instruments, $\left(Z'Z\right)^{-1}$, and $\hat{\Lambda}$ the covariance of the moment conditions, $\sum\limits_{j\in J}\sum\limits_{t\in T} z_{jt}z_{jt}'\xi_{jt}^2$. In addition, we compare the level of the market share integrals calculated by the different rules.

[31]We only realized this error after completing the research for this chapter and plan to rerun these experiments on a larger scale post-thesis using the infrastructure of Chapter 3.

Future research should also examine how quadrature rules affect the approximation of the gradient and Hessian of the objective function, which are more important than the level of the market shares in determining the point estimates and standard errors.

In our computations, we use the following numerical integration techniques: pseudo-Monte Carlo (pMC), Gaussian Hermite product rule, SGI [Heiss and Winschel, 2008], and Stroud monomial rule 11-1 [Stroud, 1971]. Because we have assumed that the mixing distribution of the random coefficients is normal, we compute the pMC nodes by drawing between $1,000$ and $10,000$ nodes from a standard normal distribution using MATLAB's randn function. We use the same draws for each market share integral, s_{jt}, as in DFS. Current 'best practice' seems to be $5,000$ points so $10,000$ nodes will enable us to put reasonable bounds on the accuracy of simulation-based BLP results. Berry et al. [1995] use pMC with importance sampling in an attempt to reduce variance, but importance sampling is just a non-linear change of variables and should not significantly improve the performance of pMC. Although using different draws for each market share integral would improve the point estimates, we use the same set of draws for each integral because this appears to be common practice for estimating BLP models.

For polynomial-based rules, we use quadrature rules which are designed for a Gaussian kernel, $\exp^{-x^2}$, because the mixing distribution is normal. Consequently, the correct one-dimensional rule to generate the nodes and weights for the multi-dimensional product rule is Gaussian-Hermite. The product rule consists of all Kronecker products of the one-dimensional nodes and the weights are the products of the corresponding one dimensional weights. The algorithm is shown in Listing 1:[32]

Listing 1: MATLAB code to generate a multi-dimensional rule from the tensor product of one dimensional nodes and weights.

```
function [ Q_NODES, Q_WEIGHTS ] = GHQuadInit( nDim_, nNodes_ )

  % Get one-dimensional Gauss-Hermite nodes and weights using
  % algorithm in Numerical Recipies by Press et al (1992).
  tmp = gauher( nDim_ ) ;
```

[32]The function `gauher` uses the algorithm in Press et al. [2007] to compute one dimensional Gaussian-Hermite nodes and weights.

```
% extract quadrature information for one dimension
vNodes = tmp( :, 1 ) ;
vWeights = tmp( :, 2 ) ;

% calculate three dimensional nodes and weights
Q_WEIGHTS = vWeights ;
for ix = 2 : nDim_
  Q_WEIGHTS = kron( vWeights, Q_WEIGHTS ) ;
end

% Make sure that the right-most dimension (ixDim = nDim_)
% varies most quickly and the left-most (ixDim = 1) most slowly
Q_NODES = zeros( nDim_, nNodes_^nDim_ ) ;
for ixDim = 1 : nDim_
  Q_NODES( ixDim, : ) =                          ...
    kron( ones( nNodes_^(ixDim - 1), 1 ),        ...
      kron( vNodes, ones( nNodes_^(nDim_ - ixDim), 1 ) ) ) ;
end

% Correct for Gaussian kernel versus normal density
Q_WEIGHTS = Q_WEIGHTS / ( pi ^ ( nDim_ / 2 ) ) ;
Q_NODES = Q_NODES * sqrt( 2 ) ;
```

Note that the normal density requires renormalization of the nodes and weights because the Gaussian kernel, unlike the normal density, lacks a factor of $1/2$ in the exponent as well as the factors of $\pi^{-1/2}$ used for normalization. We experimented with product rules for five dimensions[33] which used 3, 4, 5, 7, or 9 nodes in each dimension. We found that using more nodes than 7 in each dimension did not improve accuracy but greatly increased computational cost because of the curse of dimensionality: for a five dimensional shock the product rule with 7 nodes per dimension requires $7^5 = 16,807$ nodes to compute a share integral (whereas 9 nodes per dimension would require $9^5 = 59,049$.).

Sparse grids integration rules function similarly to product rules but exploit symmetry so that fewer points are required. We use the Kronrod-Patterson algorithm for a Gaussian kernel, as described in Heiss and Winschel [2008], and compute nodes and weights for a five-dimensional problem which is exact for polynomials of total order 11 or less using their MATLAB code.[34] We chose this configuration so that SGI is exact for the same total order as the monomial rule. For this level

[33]The dimension is five because the synthetic data has four product characteristics plus price.

[34]The code can be downloaded from http://www.sparse-grids.de/.

of accuracy, 993 nodes are required, a substantial improvement on the product rule and only 10 more than the monomial rule. However even a small increase in accuracy requires a rapid increase in the number of nodes, e.g. exactness for total order 13 requires $2,033$ nodes. See their paper for the details.

Lastly, we use Stroud [1971]'s monomial rule 11-1 for a Gaussian kernel. Stroud [1971] provides two solutions,[35] both of which provide comparable performance and integrate all five-dimensional monomials of total order 11 or less exactly using only 983 nodes. To implement the rule, we wrote a function which computed the nodes and weights from the data in Stroud's text. [36] This simply involves a lot of book-keeping to compute the correct permutations of the node elements using Stroud's data.

Now we briefly discuss our choice of the SNOPT solver, how we configured it, and numerical stability.

2.4.1.1 *Solver Choice and Configuration*

Because different solvers work better on different problems, we tried both the KNITRO and SNOPT solvers on BLP. Both solvers use efficient, modern algorithms: KNITRO supports active set and interior point algorithms [Byrd et al., 2006] whereas SNOPT uses a Sequential Quadratic Programming (SQP) method [Gill et al., 2002]. For details about these algorithms see Nocedal and Wright [2000]. Although KNITRO out performs MATLAB's fmincon non-linear solver, we found that SNOPT could often find an optimum when KNITRO would not converge. We suspect SNOPT is more reliable because it uses a sequential quadratic programming algorithm which is more robust than KNITRO's interior point method when the objective function or constraints are non-convex. In addition, SNOPT 7 was recently upgraded to handle rank deficient systems. Chapter 3 develops a C++ implementation of BLP which further exploits the robustness of SNOPT when solving BLP models by enabling SNOPT's LU rook pivoting option.[37] This is another indication of (near) rank deficiency and ill-conditioning. Consequently, if the objective function is nearly flat – i.e.,

[35]We label the two versions of rule 11-1 as 'Left' or 'Right', according to whether we use the set of nodes and weights in the left or right column of his Table $E_n^{r^2}$: 11-1 on pp. 322-323.

[36]Our monomial code is available at www.ucl.ac.uk/$\sim$uctpbss/public/code/HighPerfQuad.

[37]The LU rook pivoting option uses a more stable LU decomposition scheme which is roughly a factor of two slower.

poorly identified – SNOPT should be more stable. In addition, interior point methods, such as those used by KNITRO, do not work well on nonconvex problems. SNOPT uses an SQP method which can handle the local non-convexities caused by simulation for almost all of the datasets which we generated.

To get the most out of the solver, we fine-tuned the solver's options. In addition, for both solvers we specified the sparseness pattern and supplied hand-coded derivatives (gradient and Hessian of the objective function; Jacobian of the constraints) in order to increase numerical stability and performance. We also set box constraints to prevent the solver from searching bad regions of parameter space, as discussed below in 2.4.1.2. Lastly, we set the tolerance to $1e-6$ which is the default for SNOPT [Gill et al., 2002] and is a typical outer loop tolerance for BLP.

2.4.1.2 *Numerical Stability Considerations: Overflow and Underflow*

During our initial experiments we soon became concerned that the BLP model, despite some support for identification [Berry et al., 2004b, Berry and Haile, 2010], was not identified – or at least could not be precisely estimated given the limits of current computers. SNOPT often terminated with error code EXIT=10 and INFORM=72. These codes did not mean that the solver had failed to converge but that it had encountered market shares which were indeterminate, i.e. the computations produced a NaN.[38] In some cases, the constraint that $\log S_{jt} = \log s_{jt}$, i.e. that the logs of the observed and calculated market shares are equal, diverged to $-\infty$ when the shares were nearly zero. This problem occurs because the market share calculation is numerically unstable when the utility from the chosen alternative is extremely large.

This problem is common with logit-based models because the exponential function diverges quickly to infinity for even moderately-sized arguments. Consider the typical straight-forward implementation of the logit where $f(V;j) = \dfrac{\exp(V_j)}{\sum\limits_k \exp(V_k)}$, for some vector of utilities, V, and choice j. This implementation is unstable when $V_j \to \infty$ because

[38] Many higher-level languages such as MATLAB treat these error conditions by setting a variable's value to *Inf* or *NaN*. However, the researcher must explicitly check for these conditions using `isinf()` and `isnan()`. In general, the CPU generates a floating point exception when these conditions occur. The operating system will then raise the signal SIGFPE to the process and the process can either catch the signal by installing a signal handler or ignore it, which is often the default.

then $f(V;j) \rightarrow \dfrac{\infty}{\infty} \equiv \text{NaN}$. This situation can happen when evaluating quadrature nodes which are in the tail of the distribution. However, this formulation does allow one to compute a vector $w_k = \exp(V_k)$ and then compute choice probabilities from w, which greatly speeds up computation because it decreases the number of evaluations of $\exp(\cdot)$, which is an expensive function to compute. We found that the code was more than $10\times$ slower without this optimization on a 2.53 GHz dual-core MacBook Pro with 4 GB of 1067 MHz DDR3 RAM.[39]

By re-expressing the logit as the difference in utilities, $\tilde{f}(V;j) = \dfrac{1}{\sum\limits_k \exp(V_k - V_j)}$, we can solve the stability problem. This specification is equivalent to $f(V;j)$ but much more stable: the difference in utilities are typically small whereas utility itself can be large and lead to overflow. The cost is that we are no longer able to work in terms of $w = \exp(V)$ and must perform more operations. See the code for details. This is a common example of the engineering trade-off between speed and robustness.

Because the BLP model uses an outside good with $V_0 = 0$, the choice probability is now $f(V;j) = \dfrac{\exp(V_j)}{1 + \sum\limits_k \exp(V_k)}$ and, consequently, this trick no longer works. Instead, we impose box constraints which are tight enough to prevent the solver from examining regions of parameter space which lead to overflow yet loose enough to usually include the global optimum. Typically, the box constraints are $\pm 15 \times \|\hat{\theta}\|$ for θ_1 and θ_2 and $\pm 10^8$ for the other variables, δ and $g = Z'\xi$.

Nevertheless, this does not fully address the underlying problem of exceeding the limits of MATLAB's numerical precision. Recently, we developed a fast, robust implementation of BLP in C++ which solves these issues: see Chapter 3 for more information.[40] This implementation uses MPEC, high performance quadrature rules, and a high quality solver (SNOPT). In addition the code uses higher precision arithmetic, which has twice the precision of MATLAB, to overcome problems with overflow and underflow. Initial investigations show that higher precision completely solves these overflow and underflow problems. This C++ implementation also computes the same huge standard errors for all polynomial rules and resolves the difficulty in

[39]Test runs on an 8 core Mac Pro with 32 GB of RAM were considerably faster although MATLAB used only two of the cores. Consequently, the bottleneck appears to be swapping and not CPU cycles.

[40]This code is available upon request.

reliably calculating standard errors which we report in 2.5.2.3: higher precision arithmetic allows the solver to find a better optimum when the objective function is nearly flat.

Lastly, we start the solver at multiple different points which are randomly chosen about the average of the initial point estimates. If the solver converges to the same point for all starts, then it is likely to be the global optimum. On the other hand, if the solver converges to many different points, there are multiple local optima.

2.5 RESULTS

The polynomial rules out-perform simulation in all respects: they produce more accurate results at much lower computational cost. Of all the rules, the Gauss-Hermite product rule with 7^5 nodes should be considered the 'gold standard' and serves as our benchmark for the other rules because it is exact for degree 13 monomials as well as many higher moments. We obtained broadly similar results for all five Monte Carlo datasets. All rules performed consistently well on dataset 3. Estimates using datasets 4 and 5 varied considerably based on the quadrature choice, especially the standard errors. Datasets 1 and 2 performed between these two extremes.

We now discuss how the different rules affect computed market shares, point estimates, and standard errors.

2.5.1 *Computation of Predicted Market Shares*

One of the first experiments we performed was to compute the predicted BLP market share integrals for $T = 50$ markets and $J = 25$ products with each quadrature rule. These results provided both a quick check that our code was performing correctly and a visual comparison of the rules. We computed the market share integrals for each dataset at ten different parameter values near the MPEC point estimates, $\hat{\theta}^{MPEC}$. We selected these parameter values by first computing the GMM estimates using MPEC and then computing an additional nine parameter values where θ is drawn from a normal distribution with $\theta \sim N\left(\hat{\theta}_{MPEC}, \mathtt{diag}\left[(0.25)^2 \left\|\hat{\theta}_{MPEC}\right\|\right]\right)$, i.e. the standard errors are 25% of the magnitude of the point estimates. Figure 1 plots relative error of the market share integrals at $\hat{\theta}^{MPEC}$ versus mean market share. These computations show a vertical cloud of points for the $N = 100$

different pMC calculations of each integral ($R = 1,000$ draws) with the polynomial rules centered in the middle, as we would expect: pMC is unbiased so the polynomial results should be near the average of the Monte Carlo values. The relative error is with respect to the mean pMC market share, $\bar{s}_{jt}^{pMC} = \sum_{n=1}^{N} s_{jt}^{pMC(n)}$, where $s_{jt}^{pMC(n)}$ is the n-th replication of (j,t) market share integral computed using pMC. The green circles represent the pMC cloud of values calculated for different replications of a specific (j,t) product-market pair; the magenta pentagon the 7^5 node Gaussian-Hermite product rule; the red and blue triangles the 'Left' and 'Right' Stroud rules; and the yellow diamond the SGI. We only show this figure for dataset 1 at $\hat{\theta}^{MPEC}$ because the story is essentially the same for other datasets.[41] These plots clearly show the simulation noise in the computation of market shares s_{jt}: the different MC share values form a green 'pMC cloud' in which the polynomial based rules are located in the center of the cloud. This is exactly where you would expect the true share value to be located because pMC is unbiased as $N \to \infty$. Often, it was necessary to magnify the figures many times in order to detect any difference between the polynomial-based rules. This figure demonstrates the much higher accuracy of the monomial and sparse grids rules. An example of a close up for a market share integral is plotted in Figure 2, in this case the largest computed share value, the the story is the same for the other integrals. The noise in the pMC calculations and the consistency of the polynomial rules are clearly evident.

The 'Right' Stroud rule did produce one share value which was far outside the MC cloud and also far from the values for the other polynomial rules. In addition, this rule also produced a negative share value, as discussed in 2.5.1.2, and was more likely to generate numerically undefined results during optimization.

To quantify the performance of the different rules, we computed the maximum and average absolute deviation for the predicted share values from the product rule with 7^5 nodes at $\hat{\theta}^{MPEC}$ over all shares.[42] Although, it may not be the best representation of the 'truth', the Gaussian product rule is the most precise rule which we compute because of the additional, higher order terms. Consequently, we use it as our benchmark. Table 3 shows that the maximum absolute errors

[41] One exception is dataset 5 which has extremely low variance for the integrals computed with pMC.

[42] I.e., the maximum absolute error is $\max_{j,t}\left[\left|s_{jt}\left(\hat{\theta}^{MPEC}\right) - s_{jt}^{Product\ Rule}\left(\hat{\theta}^{MPEC}\right)\right|\right]$.

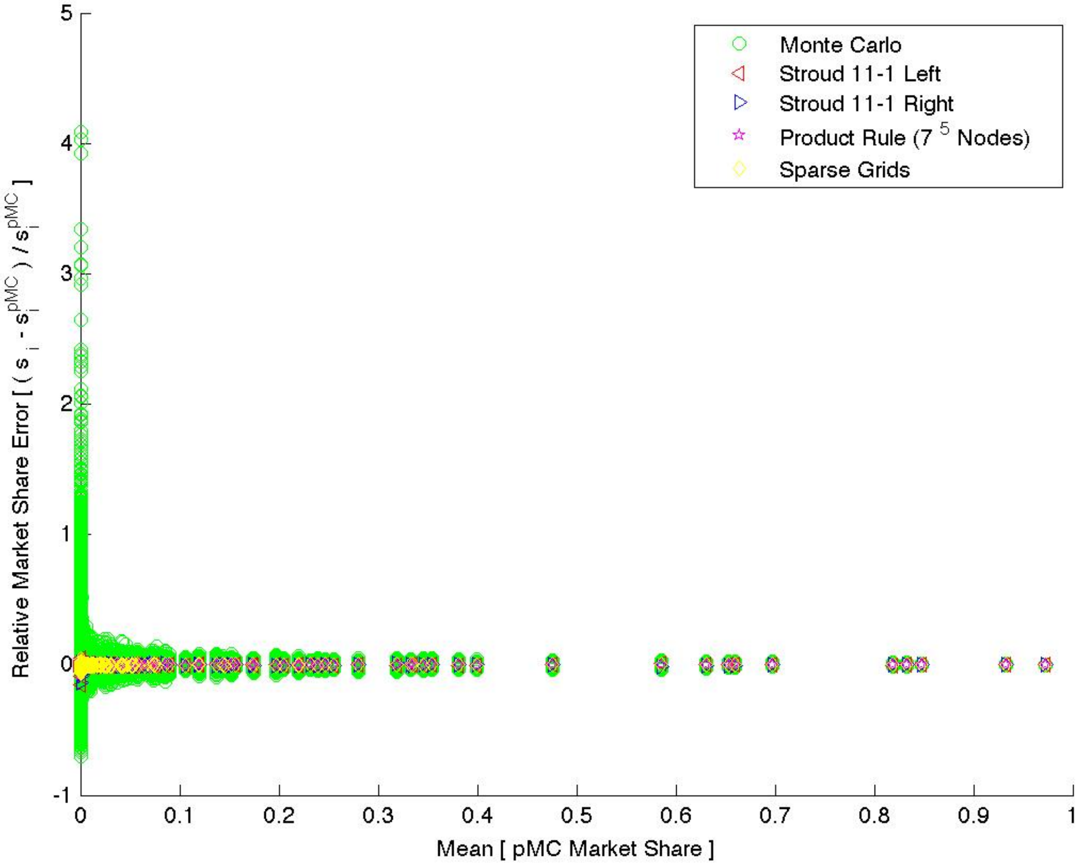

Figure 1: Relative Share Error for Different Quadrature Rules

Figure 1 shows the relative error of the product-market share integrals, s_{jt}, computed using different quadrature rules versus market share, s_{jt}. The relative error is calculated with respect to the mean pMC share value for each integral. The pMC rule is computed with $R = 1,000$ nodes and $N = 100$ replications.

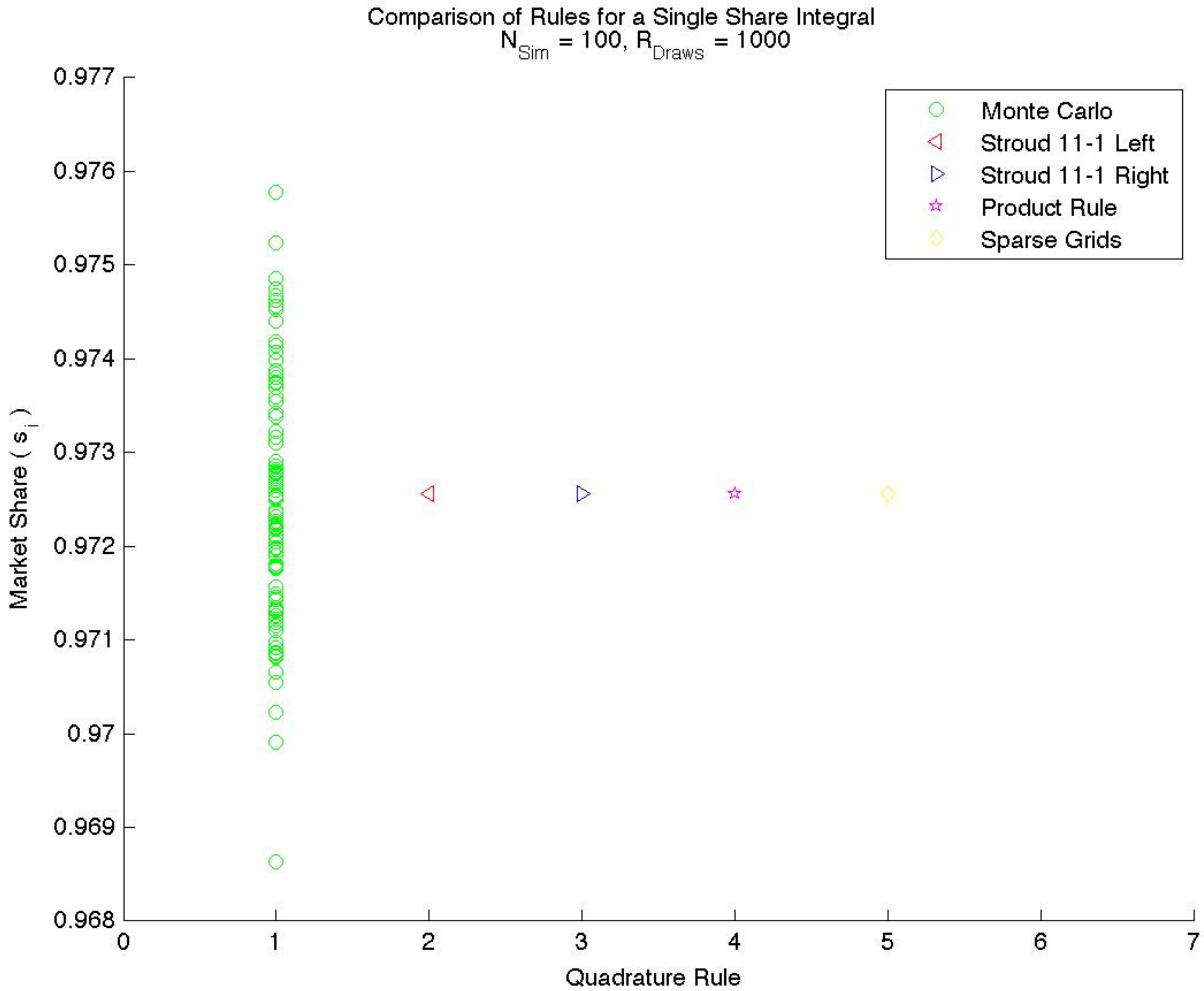

Figure 2: Close-Up of a Market Share Integral.

Figure 2 shows a close-up of the different computed values for a specific product-market share integral.

of polynomial rules are at least an order of magnitude more accurate than pMC when comparing rules with a similar number of nodes. Even with a large numbers of nodes by contemporary standards, such as $10,000$ draws, pMC produces much less accurate results. SGI's results may differ less from the product rule than the monomial rule's because SGI uses a subset of the nodes in the Gaussian product rule, whereas the monomial rule uses entirely different nodes and weights. Because the sparse grids rule drops extremal product rule nodes which have very small weights, SGI should provide extra numerically stability.

The results in Table 3 only tell part of the story. The biggest differences between the Gauss-Hermite product rule with 7^5 nodes and the other quadrature rules occur for the largest share values. For larger shares an error of 10% or so appears as a huge maximum absolute error whereas the maximum absolute error for smaller shares may appear small even if a rule differs from the benchmark by several orders of magnitude because the share value is essentially zero. In these cases, relative error is a better measure of performance. Examining the histograms for the maximum absolute error shows that for the polynomial rules there are only a few integrals with significant differences from the benchmark 7^5 node product rule whereas for pMC there is a fat tail of shares which differ considerably. In addition, the Gaussian product rule with 7^5 requires about ten times more nodes than the other polynomial-based rules for insignificant gains in accuracy.

Tables 4 through 8 show the computational costs (seconds) of computing the point estimates for the different rules.[43] The CPU Time column refers to the total time in seconds the solver took to compute an optimum and hence depends on both the speed of the quadrature rule and how quickly the solver converged. We see that the most efficient polynomial rules – SGI and monomial rule 11-1 – are more than a factor of ten faster than the pMC rule with $R = 10,000$ draws and also more accurate. pMC with $R = 10,000$ draws and the Gauss-Hermite product rule with 7^5 nodes are both much slower than the monomial and sparse grids rules because they use many more nodes, which primarily determines the increase in computational costs. We discuss the other columns below in 2.5.2.

Increasing the number of simulation draws from 100 to $10,000$ does little to improve the accuracy of the integral because pMC convergence

[43]Quoted CPU times are for a 2.53 GHz Intel Core 2 Duo MacBook Pro running OS/X 10.6.4 in 64-bit mode with 4 GB of 1067 MHz DDR3 RAM, 6MB L2 cache, and 1.07 GHz bus.

RULE	TYPE	n_{nodes}	MAX ABS ERROR	AVE ABS ERROR
Pseudo-	Simple Random	100	7.29e-02	6.73e-04
Monte	Draws	1,000	2.60e-02	2.19e-04
Carlo		10,000	7.05e-03	6.83e-05
Product	Gauss-Hermite	$3^5 = 243$	1.21e-03	1.60e-05
Rule		$4^5 = 1,024$	2.51e-04	2.51e-06
		$5^5 = 3,125$	4.94e-05	5.43e-07
		$7^5 = 16,807$	0	0
Monomial	Left Column	983	1.36e-02	2.80e-05
Rule 11-1	Right Column	983	1.14e-02	4.02e-05
Sparse Grids	Kronrod-Patterson	993	4.98e-04	4.09e-06

Table 3: Comparison of Integration Rules

The columns titled MAX ABS ERROR and AVE ABS ERROR refer to the maximum and average absolute error observed for the market shares computed for each rule with respect to the benchmark Gaussian-Hermite product rule with 7 nodes in each of the 5 dimensions. The Monte Carlo rules use $N = 1$ replication. All values are computed at $\hat{\theta}_{MPEC}$ based on $R = 1,000$ draws.

improves as $\sqrt{R}$. Because most of the products have very small market share – for the five synthetic datasets, about 90% of predicted shares are less than 0.01 – we conjecture that only a few products in each market determine the parameter values and that estimating these market shares correctly is necessary in order to obtain accurate point estimates for $\hat{\theta}$. The larger predicted market shares also have larger standard error, where standard error is computed over the N different pMC share replications . The small market shares have smaller errors not because they are calculated more accurately but because they are essentially zero. This effect becomes starker with more simulation draws. Another issue is that the parameter value used to compute the shares will affect which combinations of product and market produce the largest shares. Simple experiments show that 10% or more of shares can move into or out of the top decile of predicted share values.

When comparing the own-price elasticities computed with pMC ($R = 1,000$) and SGI, the values appear very similar (See Figure 3), with most of the difference in elasticities clumped at zero. But, most market share integrals are extremely close to zero. Consequently, we expect elasticities of small shares to be nearly the same for both rules, based on the following argument. With linear utility and a simple logit, the own price elasticity is $e_{jt} = -\alpha p_{jt} \left(1 - s_{jt}\right)$. If $s_{jt} \approx 0$ then $e_{jt} \approx -\alpha p_{jt}$ and the residual should be close to zero. Using this intuition for the mixed logit, even with random coefficients, if the market shares are small then the elasticities are likely to agree. For the larger product-market shares, the deviations in elasticity can be 10% or more, showing that pMC does not approximate the derivatives of the integrals as well as SGI. Results for the monomial rule are identical.

2.5.1.1 *Simulation Error and Bias*

Numerical integration is an approximation and like all approximations has error. The quality of a quadrature rule depends on how quickly the rule converges as $R \to \infty$ and the bounds on its error. Because pMC rules converge as $R^{-1/2}$ [Judd, 1998], you must increase the number of nodes R by a factor of 100 to gain an extra decimal place with pMC. For polynomial-based rules, if the Riemann–Stieltjes integral exists, the product rule will converge [Stroud, 1971]. Multi-dimensional error bounds formulas do exist but they are sufficiently complicated that Stroud [1971] only states very simplified versions of the theorems. The key point is that the polynomial rules should converge more quickly

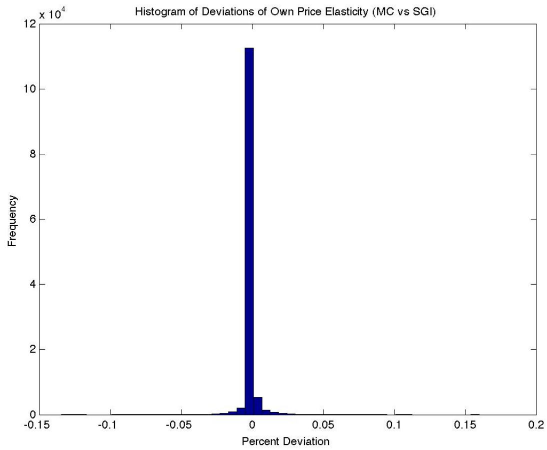

Figure 3: Comparison of Computed Own-Price Elasticities for Sparse Grids and pMC

Figure 3 shows the distribution of residuals which are the difference between the elasticities calculated with polynomial rules and the mean of the pMC share computations for $N = 100$ replications with $R = 1,000$ draws.

and have much tighter error bounds than MC methods because their error depends on higher order terms in a Taylor series expansion. Initially, we thought that pMC could out perform polynomial rules when high order terms of the Taylor series of the integrand did not vanish quickly. As we discussed in 2.2.4, simulation does not preform significantly better than polynomial rules when these high order terms are significant.

Section 2.3 explained how integration errors can propagate through the model either via the choice probabilities or the gradient of δ (i.e. the inverse of the market shares, $s^{-1}(S;\theta)$. Error enters the share integrals from integrating over the distribution of the random coefficients which affect the BLP model via the mean zero, type-specific preference shock μ_{ijt}. Errors in computing this shock propagate through the model and are further distorted by the multinomial logit transformation which can be flat, concave, or convex depending on parameter values. From Jensen's inequality we know that the expectation of a concave (convex) function is less (more) than the function of the expectation. Consequently, simulation error percolates through the multinomial logit form to produce either positive or negative error. Two facts suggest that pMC causes much more error and bias than the monomial rule: (1) the expectation of μ_{ijt} with a pMC rule is on the order of 10^{-3} even with $R = 10,000$ draws but about 10^{-17} for the monomial rule; and (2) the correlation coefficient of μ_{ijt} and the simulation error, $e_{jt} = s_{jt}^{MC} - s_{jt}^{Monomial}$, is about -0.2 conditional on $|e_{jt}| > 10^{-4}$.[44]

2.5.1.2 *Negative Market Shares*

Because some weights for monomial and sparse grids rules are negative, the approximation for a market share integral could be negative. However, this is only likely for extremely small shares where the polynomial approximation is poor in a region of parameter space where the integral is essential zero everywhere, i.e. flat. We only observed one negative value out of the $625,000$ integrals calculated.[45] This value was approximately -10^{-9} (i.e. effectively zero) and occurred with the 'Right' version of the monomial rule.

2.5.2 *Robustness of Point Estimates and Standard Errors*

We computed the point estimates and standard errors for each synthetic dataset at five starting values.[46] Tables 4-8 show f_k, the value of the GMM objective function at the optimum, as well as the CPU time in seconds required for SNOPT to converge to the point estimate.[47] If the value of f_k is the same for all starts, then the solver has likely found a unique global optimum. For the most accurate rule, the Gauss-Hermite product rule with 7^5 nodes, the solver always finds the same f_k for each start. For SGI and the monomial rule, the solver always found the same optimum for every starting value and dataset except for one start for dataset 5. Furthermore, both SGI and the monomial rule had the same problematic starting value. pMC, however, typically finds two or three different optima for each dataset, even when $R = 10,000$ draws, because Monte Carlo noise creates spurious local optima. In addition these tables show that sparse grids

[44] Here, mean $\left(\mu_{ijt}\right) \equiv \frac{1}{R}\sum_i \mu_{ijt}$.

[45] Five Monte Carlo datasets, each with $R = 100$ replications of the $J \times T = 1250$ share integrals results in $5 \times 100 \times 1,250 = 625,000$.

[46] Initially, we simply computed these optima for the same five starting values for each rule and dataset. However, the solver often aborted with numerical problems. Imposing box constraints which were sufficiently loose to include a large region of parameter space yet rule out extremal regions of parameter space solved this problem for most quadrature rules and datasets. Many of these numerical problems are caused by floating point underflow/overflow. Ultimately, we resolved the problem by rewriting our code in C++ and using higher precision arithmetic. See 2.4.1.2.

[47] Quoted CPU times are for a 2.53 GHz Intel Core 2 Duo MacBook Pro running OS/X 10.6.4 in 64-bit mode with 4 GB of 1067 MHz DDR3 RAM, 6MB L2 cache, and 1.07 GHz bus.

integration (993 nodes) and the monomial rule (983 nodes) require the same amount of CPU time as pMC with $R = 1,000$ despite being more accurate than pMC with $R = 10,000$. Both of these polynomial rules are also a factor of ten faster than pMC with $R = 10,000$ draws.

Dataset	EXIT	INFORM	f_k	CPU Time
1	0	1	33.23675	366.80
1	0	1	33.23675	753.15
1	0	1	33.85679	632.36
1	0	1	33.23681	687.31
1	0	1	38.53239	740.08
2	0	1	26.05084	457.01
2	0	1	24.60745	444.16
2	0	1	26.05084	526.40
2	0	1	26.05084	802.80
2	0	1	23.27163	855.66
3	0	1	19.76525	1071.80
3	0	1	19.76526	420.09
3	0	1	19.76524	783.48
3	0	1	19.76528	641.23
3	0	1	19.76524	635.87
4	0	1	28.19951	654.80
4	0	1	28.19951	1081.98
4	0	1	28.19951	820.40
4	0	1	28.19951	810.95
4	0	1	28.19951	796.42
5	0	1	203.50784	668.71
5	0	1	213.97591	503.92
5	0	1	203.50784	626.74
5	0	1	208.76144	489.06
5	0	1	208.76144	696.81

Table 4: Point Estimates: pMC with first 5 good starts and $R = 1,000$ draws.

The point estimates[48] (Tables 9-13) also indicate that pMC rules cause false local maxima: by comparing $\hat{\theta}$ for different starts for a given dataset, the estimates which have the same value for f_k agree to three or more digits while those with different f_k do not agree at all. On the other hand, the polynomial rules – with the exception of dataset 5's fifth start – agree to many decimal places. pMC point estimates also suffer from increased variation in $\hat{\theta}$, excessively tight standard errors (See 2.5.2.3), and confidence intervals which do not contain the point estimates from the polynomial rules. In general, the point estimates for $\hat{\theta}_1$ are more often significant than those for $\hat{\theta}_2$, the square root of the diagonal elements of the variance of the random coefficients.

[48]Note: sometimes the solver finds negative values for θ_2, which is the square root of the diagonal elements of the variance matrix, Σ, for the random coefficients. In our code θ_2 acts as the scale on the quadrature nodes because we have assumed that Σ is diagonal. The symmetry of the Gaussian kernel means that only the magnitude of θ_2 matters, not the sign. To avoid this confusion, we report $\left|\hat{\theta}_2\right|$ for the point estimates of $\hat{\theta}_2$.

Dataset	EXIT	INFORM	f_k	CPU Time
1	0	1	34.75771	8035.81
1	0	1	34.75772	5744.15
1	0	1	34.75771	3308.05
1	0	1	33.71660	3341.52
1	0	1	34.75776	7782.77
2	0	1	23.33186	7666.86
2	0	1	23.13213	6793.78
2	0	1	22.66818	7161.72
2	0	1	23.24129	7053.06
2	0	1	23.76645	8901.79
3	0	1	21.64541	8376.50
3	0	1	21.58369	8265.87
3	10	72	294.95326	178.32
3	0	1	21.69790	6567.52
3	0	1	21.95653	7835.28
4	0	1	22.49406	7955.48
4	0	1	22.49407	5446.51
4	0	1	26.12617	6544.76
4	0	1	26.12617	7427.27
4	0	1	26.22725	6852.28
5	0	1	260.57447	5450.45
5	0	1	279.95232	6514.08
5	0	1	299.22156	5555.86
5	0	1	299.22156	5444.99
5	0	1	279.95232	4403.82

Table 5: Point Estimates: pMC with $R = 10,000$ draws.

Dataset	EXIT	INFORM	f_k	CPU Time
1	0	1	35.05646	7083.95
1	0	1	35.05639	9142.16
1	0	1	35.05644	4940.91
1	0	1	35.05639	6184.56
1	0	1	35.05651	5952.06
2	0	1	22.98928	15317.16
2	0	1	22.98929	14141.56
2	0	1	22.98927	14354.17
2	0	1	22.98928	9736.57
2	0	1	22.98928	10742.86
3	0	1	21.77869	7306.40
3	0	1	21.77873	6992.33
3	0	1	21.77872	5968.52
3	0	1	21.77869	5154.03
3	0	1	21.77870	6979.46
4	0	1	25.63232	7653.30
4	0	1	25.63232	6574.78
4	0	1	25.63232	8695.48
4	0	1	25.63232	6739.00
4	0	1	25.63232	9277.51
5	0	1	288.69920	6334.33
5	0	1	288.69920	7553.43
5	0	1	288.69920	7164.02
5	0	1	288.69920	8156.16
5	0	1	288.69920	5521.13

Table 6: Point Estimates: Gauss-Hermite with first 5 good starts and 7^5 nodes.

Dataset	EXIT	INFORM	f_k	CPU Time
1	0	1	35.27236	616.63
1	0	1	35.27217	549.22
1	0	1	35.27212	269.22
1	0	1	35.27216	414.71
1	0	1	35.27212	432.32
2	0	1	22.97539	980.88
2	0	1	22.97541	910.89
2	0	1	22.97539	724.68
2	0	1	22.97539	865.45
2	0	1	22.97540	1026.54
3	0	1	21.78433	433.50
3	0	1	21.78430	557.89
3	0	1	21.78432	610.45
3	0	1	21.78437	352.71
3	0	1	21.78434	604.79
4	0	1	25.59501	515.58
4	0	1	25.59501	388.67
4	0	1	25.59501	496.07
4	0	1	25.59501	439.85
4	0	1	25.59501	586.94
5	0	1	293.89029	494.45
5	0	1	293.89029	571.11
5	0	1	293.89029	481.82
5	0	1	293.89029	556.80
5	0	1	487.40742	6535.28

Table 7: Point Estimates: SGI with first 5 good starts and 993 nodes (exact for degree ≤ 11).

Dataset	EXIT	INFORM	f_k	CPU Time
1	0	1	34.63546	644.52
1	0	1	34.63550	578.42
1	0	1	34.63556	449.30
1	0	1	34.63552	294.48
1	0	1	34.63548	443.86
2	0	1	23.24928	1174.66
2	0	1	23.24928	660.97
2	0	1	23.24928	922.52
2	0	1	23.24928	1150.00
2	0	1	23.24928	1022.48
3	0	1	21.79928	688.71
3	0	1	21.79931	373.36
3	0	1	21.79926	669.28
3	0	1	21.79926	483.89
3	0	1	21.79926	573.57
4	0	1	24.72862	435.54
4	0	1	24.72862	587.55
4	0	1	24.72862	739.98
4	0	1	24.72862	613.63
4	0	1	24.72862	657.03
5	0	1	277.89463	441.45
5	0	1	278.03790	441.77
5	0	1	277.89463	548.75
5	0	1	277.89463	1134.53
5	0	1	278.03790	656.13

Table 8: Point Estimates: Monomial with first 5 good starts and 983 nodes (exact for degree ≤ 11).

Data	INFORM	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{21}	θ_{22}	θ_{23}	θ_{24}	θ_{25}
1	1	1.0546	1.577	1.6490	0.09924	-2.594	1.0244	0.5553	0.5823	0.53126	0.2210
		(0.3494)	(0.1204)	(0.1337)	(0.09090)	(0.1465)	(0.4602)	(0.1198)	(0.09144)	(0.09601)	(0.06035)
1	1	1.0545	1.577	1.6489	0.09922	-2.594	1.0243	0.5553	0.5823	0.53125	0.2210
		(0.3494)	(0.1204)	(0.1337)	(0.09090)	(0.1465)	(0.4602)	(0.1198)	(0.09144)	(0.09602)	(0.06035)
1	1	0.9578	1.551	1.5827	0.11983	-2.551	1.0862	0.6529	0.3769	0.47002	0.2003
		(0.3249)	(0.1205)	(0.1260)	(0.09065)	(0.1332)	(0.4235)	(0.07828)	(0.1053)	(0.1095)	(0.05769)
1	1	1.0546	1.577	1.6489	0.09924	-2.594	1.0244	0.5553	0.5822	0.53122	0.2210
		(0.3494)	(0.1204)	(0.1337)	(0.09091)	(0.1465)	(0.4602)	(0.1198)	(0.09145)	(0.09604)	(0.06036)
1	1	0.9948	1.534	1.6011	0.07623	-2.564	0.9913	0.6167	0.4642	0.50064	0.2080
		(0.3667)	(0.1216)	(0.1282)	(0.09236)	(0.1457)	(0.4531)	(0.07804)	(0.08660)	(0.1206)	(0.06271)
2	1	0.5957	1.179	0.9327	0.25733	-2.311	0.6642	0.4303	0.8534	0.70568	0.1183
		(0.3597)	(0.1895)	(0.1781)	(0.1204)	(0.1278)	(0.7046)	(0.1577)	(0.06573)	(0.04496)	(0.09680)
2	1	0.7719	1.221	0.9475	0.26271	-2.396	0.7866	0.3935	0.8336	0.70480	0.1750
		(0.4212)	(0.1944)	(0.1788)	(0.1247)	(0.1807)	(0.6731)	(0.1573)	(0.07146)	(0.03602)	(0.1015)
2	1	0.5957	1.179	0.9328	0.25735	-2.311	0.6641	0.4303	0.8534	0.70568	0.1183
		(0.3597)	(0.1895)	(0.1781)	(0.1204)	(0.1278)	(0.7046)	(0.1577)	(0.06573)	(0.04496)	(0.09679)
2	1	0.5956	1.179	0.9327	0.25733	-2.311	0.6642	0.4303	0.8534	0.70568	0.1183

		(0.3597)	(0.1895)	(0.1781)	(0.1204)	(0.1278)	(0.7046)	(0.1577)	(0.06573)	(0.04496)	(0.09680)
2	1	0.8388	1.253	1.0184	0.26314	-2.439	0.8765	0.4185	0.8932	0.67489	0.1859
		(0.4100)	(0.1976)	(0.1862)	(0.1227)	(0.1733)	(0.5198)	(0.1464)	(0.06769)	(0.03969)	(0.08145)
3	1	1.3104	1.136	1.3428	0.66329	-2.437	0.6895	0.2544	0.9054	0.40733	0.1883
		(0.3115)	(0.1497)	(0.1706)	(0.1209)	(0.1270)	(0.7110)	(0.1604)	(0.04894)	(0.1104)	(0.05616)
3	1	1.3105	1.136	1.3428	0.66332	-2.437	0.6892	0.2544	0.9055	0.40733	0.1883
		(0.3115)	(0.1497)	(0.1706)	(0.1209)	(0.1270)	(0.7113)	(0.1604)	(0.04894)	(0.1104)	(0.05615)
3	1	1.3104	1.136	1.3428	0.66331	-2.437	0.6894	0.2544	0.9054	0.40732	0.1883
		(0.3115)	(0.1497)	(0.1706)	(0.1209)	(0.1270)	(0.7111)	(0.1604)	(0.04895)	(0.1104)	(0.05616)
3	1	1.3105	1.136	1.3428	0.66336	-2.437	0.6890	0.2544	0.9054	0.40730	0.1883
		(0.3115)	(0.1497)	(0.1706)	(0.1209)	(0.1270)	(0.7115)	(0.1605)	(0.04895)	(0.1105)	(0.05616)
3	1	1.3104	1.136	1.3428	0.66331	-2.437	0.6893	0.2544	0.9054	0.40731	0.1883
		(0.3115)	(0.1497)	(0.1706)	(0.1209)	(0.1270)	(0.7112)	(0.1604)	(0.04895)	(0.1104)	(0.05616)
4	1	1.6379	1.832	1.6840	0.26823	-3.021	0.7794	1.0241	0.6984	0.08211	0.6708
		(0.2955)	(0.2845)	(0.2787)	(0.2073)	(0.2661)	(1.070)	(0.1224)	(0.1609)	(0.3069)	(0.1024)
4	1	1.6379	1.832	1.6840	0.26823	-3.021	0.7794	1.0241	0.6984	0.08212	0.6708
		(0.2955)	(0.2845)	(0.2787)	(0.2073)	(0.2661)	(1.070)	(0.1224)	(0.1609)	(0.3069)	(0.1024)
4	1	1.6379	1.832	1.6840	0.26823	-3.021	0.7795	1.0241	0.6984	0.08213	0.6708
		(0.2955)	(0.2845)	(0.2787)	(0.2073)	(0.2661)	(1.070)	(0.1224)	(0.1609)	(0.3069)	(0.1024)

4	1	1.6379	1.832	1.6840	0.26822	-3.021	0.7794	1.0241	0.6984	0.08211	0.6708
		(0.2955)	(0.2845)	(0.2787)	(0.2073)	(0.2661)	(1.070)	(0.1224)	(0.1609)	(0.3069)	(0.1024)
4	1	1.6379	1.832	1.6840	0.26824	-3.021	0.7794	1.0241	0.6984	0.08212	0.6708
		(0.2955)	(0.2845)	(0.2787)	(0.2073)	(0.2661)	(1.070)	(0.1224)	(0.1609)	(0.3069)	(0.1024)
5	1	3.8847	2.941	1.7169	0.75841	-5.028	1.4322	1.2450	0.8917	1.21940	1.2161
		(0.7081)	(0.2525)	(0.2558)	(0.1694)	(0.5138)	(1.019)	(0.3340)	(0.1194)	(0.1908)	(0.1532)
5	1	0.1228	2.608	1.6139	0.24177	-2.575	1.5673	0.9745	0.7653	0.59988	0.4921
		(0.4817)	(0.1932)	(0.2031)	(0.1285)	(0.1604)	(0.6346)	(0.1237)	(0.1238)	(0.1533)	(0.04807)
5	1	3.8847	2.941	1.7169	0.75842	-5.028	1.4323	1.2450	0.8917	1.21941	1.2161
		(0.7081)	(0.2525)	(0.2558)	(0.1694)	(0.5138)	(1.019)	(0.3340)	(0.1194)	(0.1908)	(0.1532)
5	1	1.9287	2.693	1.5499	0.73407	-3.911	2.3156	1.4281	0.7147	1.01829	0.8986
		(0.6686)	(0.2155)	(0.2226)	(0.1501)	(0.3918)	(0.5851)	(0.1760)	(0.1347)	(0.1880)	(0.1172)
5	1	1.9287	2.693	1.5499	0.73408	-3.911	2.3156	1.4281	0.7147	1.01830	0.8986
		(0.6686)	(0.2155)	(0.2226)	(0.1501)	(0.3918)	(0.5851)	(0.1760)	(0.1347)	(0.1880)	(0.1172)

Table 9: Point Estimates: pMC with first 5 good starts and $R = 1,000$ draws.

Data	INFORM	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{21}	θ_{22}	θ_{23}	θ_{24}	θ_{25}
1	1	1.1427	1.5961	1.6348	0.08060	-2.607	1.00107	0.568936	0.5400	0.52351	0.2280
		(0.3011)	(0.1204)	(0.1260)	(0.09181)	(0.1185)	(0.4420)	(0.08408)	(0.08639)	(0.1115)	(0.04491)
1	1	1.1429	1.5962	1.6349	0.08061	-2.607	1.00077	0.568898	0.5401	0.52355	0.2280
		(0.2924)	(0.1192)	(0.1255)	(0.08935)	(0.1148)	(0.4597)	(0.08346)	(0.1001)	(0.09312)	(0.04868)
1	1	1.1429	1.5961	1.6349	0.08062	-2.607	1.00091	0.568914	0.5400	0.52354	0.2280
		(0.3278)	(0.1214)	(0.1296)	(0.09147)	(0.1353)	(0.4421)	(0.08909)	(0.09554)	(0.1115)	(0.04851)
1	1	1.1990	1.5906	1.6269	0.12819	-2.640	1.03337	0.613779	0.5108	0.50231	0.2434
		(0.2736)	(0.1199)	(0.1254)	(0.09026)	(0.1007)	(0.4469)	(0.08230)	(0.09002)	(0.1109)	(0.03897)
1	1	1.1428	1.5961	1.6348	0.08065	-2.607	1.00089	0.568973	0.5400	0.52354	0.2280
		(0.2781)	(0.1205)	(0.1264)	(0.09103)	(0.1078)	(0.4450)	(0.08310)	(0.09000)	(0.1100)	(0.04192)
2	1	0.6241	1.2062	1.0216	0.24133	-2.362	0.93166	0.374587	0.9034	0.71462	0.1442
		(0.3252)	(0.1928)	(0.1763)	(0.1199)	(0.1183)	(0.6059)	(0.1373)	(0.06899)	(0.03736)	(0.06459)
2	1	0.6640	1.2148	1.0039	0.24668	-2.375	0.96484	0.361393	0.8878	0.70441	0.1501
		(0.3252)	(0.1928)	(0.1763)	(0.1199)	(0.1183)	(0.6058)	(0.1373)	(0.06899)	(0.03736)	(0.06458)
2	1	0.7631	1.2212	0.9776	0.30353	-2.421	1.00189	0.403325	0.8816	0.69759	0.1685
		(0.3443)	(0.1944)	(0.1753)	(0.1222)	(0.1261)	(0.5700)	(0.1431)	(0.07513)	(0.03758)	(0.05836)
2	1	0.6603	1.2091	1.0048	0.24541	-2.376	0.95143	0.389741	0.8906	0.72306	0.1521

		(0.3370)	(0.1928)	(0.1763)	(0.1212)	(0.1245)	(0.6021)	(0.1416)	(0.06842)	(0.03572)	(0.06375)
2	1	0.7000	1.2057	1.0241	0.28422	-2.392	0.90694	0.400220	0.9065	0.70525	0.1626
		(0.3370)	(0.1928)	(0.1763)	(0.1212)	(0.1245)	(0.6022)	(0.1416)	(0.06842)	(0.03572)	(0.06373)
3	1	1.3384	1.1197	1.3602	0.70373	-2.432	0.53518	0.242688	0.8573	0.39160	0.1934
		(0.3173)	(0.1539)	(0.1649)	(0.1235)	(0.1298)	(0.6335)	(0.1424)	(0.04434)	(0.1097)	(0.05803)
3	1	1.2696	1.1225	1.3489	0.69837	-2.412	0.56292	0.278950	0.8384	0.39886	0.1824
		(0.3173)	(0.1539)	(0.1649)	(0.1235)	(0.1298)	(0.6336)	(0.1424)	(0.04434)	(0.1097)	(0.05803)
3	72	5.5266	0.9599	1.3286	0.98834	-4.695	0.04864	0.009567	0.2000	0.22681	0.8772
		(0.3177)	(0.1501)	(0.1634)	(0.1217)	(0.1285)	(0.8505)	(0.1510)	(0.04468)	(0.1048)	(0.06012)
3	1	1.2687	1.1307	1.3555	0.69063	-2.400	0.47493	0.284326	0.8512	0.40623	0.1730
		(9.844)	(4.984)	(5.185)	(3.981)	(3.441)	(9.323)	(1.500)	(4.053)	(2.119)	(1.143)
3	1	1.2232	1.1134	1.3305	0.69086	-2.406	0.68129	0.291198	0.8282	0.41123	0.1803
		(0.3169)	(0.1503)	(0.1632)	(0.1221)	(0.1356)	(0.7218)	(0.1247)	(0.04699)	(0.1032)	(0.06299)
4	1	1.8402	1.8307	1.7406	0.27790	-2.779	0.27336	0.814492	0.6361	0.04740	0.5035
		(0.2571)	(0.2980)	(0.2964)	(0.1870)	(0.1182)	(0.7049)	(0.09593)	(0.1270)	(0.3194)	(0.03201)
4	1	1.8403	1.8307	1.7405	0.27789	-2.779	0.27333	0.814509	0.6361	0.04740	0.5035
		(0.2571)	(0.2980)	(0.2964)	(0.1870)	(0.1182)	(0.7050)	(0.09593)	(0.1270)	(0.3194)	(0.03201)
4	1	1.7598	1.8252	1.7178	0.26449	-2.731	0.53581	0.898076	0.4578	0.08526	0.4867
		(0.2571)	(0.2980)	(0.2964)	(0.1870)	(0.1182)	(0.7049)	(0.09593)	(0.1270)	(0.3195)	(0.03201)

4	1	1.7598	1.8252	1.7178	0.26449	-2.731	0.53586	0.898080	0.4579	0.08527	0.4867
		(0.2571)	(0.2980)	(0.2964)	(0.1870)	(0.1182)	(0.7049)	(0.09593)	(0.1270)	(0.3195)	(0.03201)
4	1	1.7619	1.8359	1.7187	0.26380	-2.715	0.34868	0.892344	0.4476	0.10942	0.4802
		(0.2571)	(0.2980)	(0.2964)	(0.1870)	(0.1182)	(0.7051)	(0.09593)	(0.1270)	(0.3195)	(0.03201)
5	1	-0.2568	2.3403	1.1956	0.32084	-2.217	0.61026	1.101380	0.2332	0.56325	0.3884
		(0.4224)	(0.1931)	(0.1949)	(0.1170)	(0.1624)	(2.189)	(0.09419)	(0.2608)	(0.1058)	(0.04790)
5	1	0.3066	2.5059	1.2591	0.43423	-2.707	1.39306	1.193434	0.3408	0.52833	0.5531
		(0.5113)	(0.1932)	(0.1962)	(0.1176)	(0.1620)	(0.6929)	(0.08841)	(0.1647)	(0.1012)	(0.04609)
5	1	-0.7372	2.3060	1.1838	0.34910	-2.073	0.79228	0.992318	0.4481	0.77180	0.3472
		(0.4224)	(0.1931)	(0.1949)	(0.1170)	(0.1624)	(2.189)	(0.09419)	(0.2608)	(0.1058)	(0.04790)
5	1	-0.7372	2.3060	1.1838	0.34910	-2.073	0.79228	0.992316	0.4481	0.77180	0.3472
		(0.4224)	(0.1931)	(0.1949)	(0.1170)	(0.1624)	(2.189)	(0.09419)	(0.2608)	(0.1058)	(0.04790)
5	1	0.3066	2.5059	1.2591	0.43423	-2.707	1.39306	1.193433	0.3408	0.52833	0.5531
		(0.5113)	(0.1932)	(0.1962)	(0.1176)	(0.1620)	(0.6929)	(0.08841)	(0.1647)	(0.1012)	(0.04609)

Table 10: Point Estimates: pMC with $R = 10,000$ draws.

Data	INFORM	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{21}	θ_{22}	θ_{23}	θ_{24}	θ_{25}
1	1	1.1304	1.596	1.6074	0.1013	-2.610	1.062e+00	0.5942	4.613e-01	5.178e-01	0.2307
		(0.3030)	(0.1197)	(0.1268)	(0.09048)	(0.1160)	(0.4376)	(0.08412)	(0.09132)	(0.1031)	(0.04396)
1	1	1.1304	1.596	1.6074	0.1014	-2.610	1.062e+00	0.5942	4.613e-01	5.178e-01	0.2306
		(0.3030)	(0.1197)	(0.1268)	(0.09049)	(0.1160)	(0.4376)	(0.08413)	(0.09131)	(0.1031)	(0.04396)
1	1	1.1304	1.596	1.6075	0.1013	-2.610	1.062e+00	0.5942	4.613e-01	5.178e-01	0.2306
		(0.3030)	(0.1197)	(0.1268)	(0.09050)	(0.1160)	(0.4377)	(0.08414)	(0.09134)	(0.1032)	(0.04398)
1	1	1.1306	1.596	1.6074	0.1013	-2.610	1.061e+00	0.5943	4.612e-01	5.178e-01	0.2306
		(0.3030)	(0.1197)	(0.1268)	(0.09049)	(0.1160)	(0.4377)	(0.08413)	(0.09133)	(0.1031)	(0.04397)
1	1	1.1305	1.596	1.6074	0.1013	-2.610	1.062e+00	0.5942	4.613e-01	5.178e-01	0.2306
		(0.3030)	(0.1197)	(0.1268)	(0.09051)	(0.1160)	(0.4377)	(0.08414)	(0.09134)	(0.1032)	(0.04398)
2	1	0.6704	1.201	0.9877	0.2705	-2.367	9.119e-01	0.3934	8.900e-01	6.823e-01	0.1440
		(0.3062)	(0.1918)	(0.1757)	(0.1212)	(0.09235)	(0.5672)	(0.1509)	(0.07302)	(0.04742)	(0.04844)
2	1	0.6705	1.201	0.9876	0.2705	-2.367	9.119e-01	0.3934	8.901e-01	6.823e-01	0.1440
		(0.3062)	(0.1918)	(0.1757)	(0.1212)	(0.09234)	(0.5672)	(0.1509)	(0.07302)	(0.04742)	(0.04844)
2	1	0.6704	1.201	0.9876	0.2705	-2.367	9.119e-01	0.3934	8.900e-01	6.823e-01	0.1439
		(0.3062)	(0.1918)	(0.1757)	(0.1212)	(0.09236)	(0.5672)	(0.1509)	(0.07302)	(0.04742)	(0.04846)
2	1	0.6705	1.201	0.9877	0.2705	-2.367	9.119e-01	0.3934	8.900e-01	6.823e-01	0.1440

		(0.3062)	(0.1918)	(0.1757)	(0.1212)	(0.09236)	(0.5673)	(0.1509)	(0.07303)	(0.04743)	(0.04845)
2	1	0.6704	1.201	0.9877	0.2705	-2.367	9.119e-01	0.3934	8.900e-01	6.823e-01	0.1439
		(0.3062)	(0.1918)	(0.1757)	(0.1212)	(0.09235)	(0.5673)	(0.1509)	(0.07302)	(0.04742)	(0.04845)
3	1	1.2715	1.116	1.3518	0.6985	-2.410	5.479e-01	0.2943	8.441e-01	4.045e-01	0.1785
		(0.3088)	(0.1511)	(0.1639)	(0.1213)	(0.1175)	(0.7818)	(0.1381)	(0.04395)	(0.1043)	(0.05457)
3	1	1.2715	1.116	1.3519	0.6985	-2.410	5.473e-01	0.2943	8.441e-01	4.044e-01	0.1785
		(0.3088)	(0.1511)	(0.1640)	(0.1213)	(0.1175)	(0.7826)	(0.1381)	(0.04395)	(0.1043)	(0.05459)
3	1	1.2714	1.116	1.3519	0.6984	-2.410	5.482e-01	0.2942	8.442e-01	4.045e-01	0.1786
		(0.3088)	(0.1511)	(0.1639)	(0.1212)	(0.1175)	(0.7814)	(0.1381)	(0.04395)	(0.1043)	(0.05456)
3	1	1.2715	1.116	1.3519	0.6984	-2.410	5.480e-01	0.2943	8.441e-01	4.045e-01	0.1785
		(0.3088)	(0.1511)	(0.1639)	(0.1212)	(0.1175)	(0.7816)	(0.1381)	(0.04395)	(0.1043)	(0.05456)
3	1	1.2715	1.117	1.3519	0.6984	-2.410	5.482e-01	0.2943	8.441e-01	4.045e-01	0.1786
		(0.3088)	(0.1511)	(0.1639)	(0.1212)	(0.1175)	(0.7813)	(0.1381)	(0.04395)	(0.1043)	(0.05455)
4	1	1.7607	1.851	1.7089	0.2778	-2.644	5.448e-07	0.8719	4.834e-01	1.553e-07	0.4462
		(0.2752)	(0.2991)	(0.2975)	(0.1918)	(0.1297)	(9.654E+05)	(0.09196)	(0.1107)	(2.315E+05)	(0.03429)
4	1	1.7607	1.851	1.7089	0.2778	-2.644	9.581e-05	0.8718	4.834e-01	2.261e-06	0.4462
		(0.2751)	(0.2991)	(0.2975)	(0.1918)	(0.1297)	(5489.)	(0.09196)	(0.1107)	(1.589E+04)	(0.03429)
4	1	1.7607	1.851	1.7089	0.2778	-2.644	1.092e-05	0.8718	4.834e-01	1.218e-07	0.4462
		(0.2752)	(0.2991)	(0.2975)	(0.1918)	(0.1297)	(4.818E+04)	(0.09196)	(0.1107)	(2.951E+05)	(0.03430)

4	1	1.7607	1.851	1.7089	0.2778	-2.644	7.800e-05	0.8718	4.835e-01	6.351e-07	0.4462
		(0.2751)	(0.2991)	(0.2975)	(0.1918)	(0.1297)	(6743.)	(0.09196)	(0.1107)	(5.658E+04)	(0.03429)
4	1	1.7607	1.851	1.7089	0.2778	-2.644	1.157e-06	0.8719	4.834e-01	2.014e-07	0.4462
		(0.2752)	(0.2991)	(0.2975)	(0.1918)	(0.1297)	(4.544E+05)	(0.09196)	(0.1107)	(1.784E+05)	(0.03429)
5	1	-0.5510	2.243	1.1337	0.3711	-2.066	1.250e-07	1.0547	1.578e-06	7.183e-01	0.3442
		(0.4717)	(0.2058)	(0.2005)	(0.1191)	(0.1520)	(5.628E+06)	(0.07972)	(4.830E+04)	(0.1011)	(0.04931)
5	1	-0.5510	2.243	1.1337	0.3711	-2.066	1.639e-07	1.0547	1.072e-06	7.183e-01	0.3442
		(0.4717)	(0.2058)	(0.2005)	(0.1191)	(0.1520)	(4.293E+06)	(0.07972)	(7.111E+04)	(0.1011)	(0.04931)
5	1	-0.5510	2.243	1.1337	0.3712	-2.066	1.819e-06	1.0547	5.292e-07	7.183e-01	0.3442
		(0.4717)	(0.2058)	(0.2005)	(0.1191)	(0.1520)	(3.868E+05)	(0.07972)	(1.440E+05)	(0.1011)	(0.04931)
5	1	-0.5510	2.243	1.1337	0.3712	-2.066	1.852e-07	1.0547	7.546e-07	7.183e-01	0.3442
		(0.4717)	(0.2058)	(0.2005)	(0.1191)	(0.1520)	(3.800E+06)	(0.07972)	(1.010E+05)	(0.1011)	(0.04931)
5	1	-0.5510	2.243	1.1337	0.3712	-2.066	3.086e-06	1.0547	2.230e-06	7.183e-01	0.3442
		(0.4717)	(0.2058)	(0.2005)	(0.1191)	(0.1520)	(2.280E+05)	(0.07972)	(3.417E+04)	(0.1011)	(0.04931)

Table 11: Point Estimates: Gauss-Hermite with first 5 good starts and 7^5 nodes.

Data	INFORM	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{21}	θ_{22}	θ_{23}	θ_{24}	θ_{25}
1	1	1.0960	1.595	1.6064	0.09740	-2.592	1.043e+00	0.5915	0.4620	5.214e-01	0.2236
		(0.2930)	(0.1196)	(0.1266)	(0.08999)	(0.1088)	(0.4373)	(0.08350)	(0.09019)	(0.1024)	(0.04203)
1	1	1.0958	1.595	1.6065	0.09747	-2.592	1.044e+00	0.5916	0.4619	5.214e-01	0.2235
		(0.2930)	(0.1196)	(0.1267)	(0.09001)	(0.1088)	(0.4371)	(0.08350)	(0.09022)	(0.1024)	(0.04204)
1	1	1.0959	1.595	1.6064	0.09745	-2.592	1.043e+00	0.5915	0.4619	5.214e-01	0.2235
		(0.2931)	(0.1196)	(0.1267)	(0.09003)	(0.1088)	(0.4373)	(0.08352)	(0.09025)	(0.1024)	(0.04206)
1	1	1.0958	1.595	1.6065	0.09744	-2.592	1.043e+00	0.5916	0.4619	5.214e-01	0.2235
		(0.2931)	(0.1196)	(0.1267)	(0.09001)	(0.1088)	(0.4372)	(0.08351)	(0.09022)	(0.1024)	(0.04205)
1	1	1.0958	1.595	1.6064	0.09749	-2.592	1.044e+00	0.5915	0.4619	5.214e-01	0.2235
		(0.2931)	(0.1196)	(0.1267)	(0.09003)	(0.1088)	(0.4372)	(0.08352)	(0.09024)	(0.1024)	(0.04207)
2	1	0.6712	1.200	0.9866	0.27148	-2.367	9.160e-01	0.3967	0.8891	6.792e-01	0.1442
		(0.3063)	(0.1919)	(0.1759)	(0.1209)	(0.09201)	(0.5634)	(0.1496)	(0.07312)	(0.04634)	(0.04810)
2	1	0.6712	1.200	0.9866	0.27147	-2.367	9.160e-01	0.3967	0.8891	6.792e-01	0.1442
		(0.3063)	(0.1919)	(0.1759)	(0.1208)	(0.09200)	(0.5634)	(0.1496)	(0.07312)	(0.04634)	(0.04810)
2	1	0.6712	1.200	0.9866	0.27150	-2.367	9.160e-01	0.3967	0.8891	6.792e-01	0.1442
		(0.3063)	(0.1919)	(0.1759)	(0.1209)	(0.09201)	(0.5634)	(0.1496)	(0.07312)	(0.04634)	(0.04810)
2	1	0.6712	1.200	0.9866	0.27149	-2.367	9.160e-01	0.3967	0.8891	6.792e-01	0.1442

		(0.3063)	(0.1919)	(0.1759)	(0.1209)	(0.09202)	(0.5634)	(0.1496)	(0.07312)	(0.04634)	(0.04811)
2	1	0.6712	1.200	0.9866	0.27149	-2.367	9.160e-01	0.3967	0.8891	6.792e-01	0.1442
		(0.3063)	(0.1919)	(0.1759)	(0.1208)	(0.09200)	(0.5634)	(0.1496)	(0.07312)	(0.04634)	(0.04809)
3	1	1.2691	1.116	1.3525	0.69822	-2.406	5.179e-01	0.2928	0.8440	4.056e-01	0.1770
		(0.3068)	(0.1508)	(0.1636)	(0.1211)	(0.1153)	(0.8268)	(0.1381)	(0.04395)	(0.1037)	(0.05372)
3	1	1.2693	1.116	1.3525	0.69836	-2.406	5.175e-01	0.2929	0.8440	4.055e-01	0.1769
		(0.3069)	(0.1508)	(0.1636)	(0.1211)	(0.1153)	(0.8275)	(0.1380)	(0.04396)	(0.1037)	(0.05374)
3	1	1.2691	1.116	1.3524	0.69829	-2.407	5.181e-01	0.2928	0.8440	4.056e-01	0.1770
		(0.3068)	(0.1508)	(0.1636)	(0.1211)	(0.1153)	(0.8265)	(0.1381)	(0.04395)	(0.1037)	(0.05371)
3	1	1.2690	1.116	1.3526	0.69823	-2.406	5.182e-01	0.2927	0.8440	4.056e-01	0.1770
		(0.3068)	(0.1508)	(0.1636)	(0.1211)	(0.1153)	(0.8263)	(0.1381)	(0.04395)	(0.1037)	(0.05371)
3	1	1.2691	1.116	1.3525	0.69826	-2.406	5.180e-01	0.2928	0.8440	4.056e-01	0.1770
		(0.3068)	(0.1508)	(0.1636)	(0.1211)	(0.1153)	(0.8266)	(0.1380)	(0.04395)	(0.1037)	(0.05371)
4	1	1.7555	1.843	1.7052	0.28229	-2.642	1.005e-06	0.8666	0.5007	1.449e-07	0.4468
		(0.2687)	(0.3005)	(0.2979)	(0.1899)	(0.1247)	(4.801E+05)	(0.09294)	(0.1156)	(2.516E+05)	(0.03291)
4	1	1.7554	1.843	1.7051	0.28229	-2.642	5.060e-05	0.8666	0.5007	6.030e-07	0.4468
		(0.2687)	(0.3005)	(0.2979)	(0.1899)	(0.1247)	(9535.)	(0.09294)	(0.1156)	(6.045E+04)	(0.03291)
4	1	1.7555	1.843	1.7052	0.28229	-2.642	4.291e-06	0.8666	0.5007	3.843e-07	0.4468
		(0.2687)	(0.3005)	(0.2979)	(0.1899)	(0.1247)	(1.125E+05)	(0.09294)	(0.1156)	(9.484E+04)	(0.03291)

4	1	1.7555	1.843	1.7052	0.28229	-2.642	3.661e-06	0.8666	0.5007	1.571e-06	0.4468
		(0.2687)	(0.3005)	(0.2979)	(0.1899)	(0.1247)	(1.318E+05)	(0.09294)	(0.1156)	(2.321E+04)	(0.03291)
4	1	1.7555	1.843	1.7052	0.28229	-2.642	1.007e-05	0.8666	0.5007	5.783e-07	0.4468
		(0.2687)	(0.3005)	(0.2979)	(0.1899)	(0.1247)	(4.793E+04)	(0.09294)	(0.1156)	(6.303E+04)	(0.03291)
5	1	-0.5386	2.263	1.1507	0.37200	-2.075	2.704e-06	1.0143	0.3028	7.257e-01	0.3472
		(0.4448)	(0.1969)	(0.1966)	(0.1182)	(0.1555)	(2.479E+05)	(0.08670)	(0.2649)	(0.1042)	(0.04988)
5	1	-0.5386	2.263	1.1507	0.37200	-2.075	3.862e-09	1.0143	0.3028	7.257e-01	0.3472
		(0.4448)	(0.1969)	(0.1966)	(0.1182)	(0.1555)	(1.736E+08)	(0.08670)	(0.2649)	(0.1042)	(0.04988)
5	1	-0.5386	2.263	1.1507	0.37200	-2.075	1.490e-06	1.0143	0.3028	7.257e-01	0.3472
		(0.4448)	(0.1969)	(0.1966)	(0.1182)	(0.1555)	(4.498E+05)	(0.08670)	(0.2649)	(0.1042)	(0.04988)
5	1	-0.5386	2.263	1.1507	0.37200	-2.075	4.880e-06	1.0143	0.3028	7.257e-01	0.3472
		(0.4448)	(0.1969)	(0.1966)	(0.1182)	(0.1555)	(1.374E+05)	(0.08670)	(0.2649)	(0.1042)	(0.04988)
5	1	-1.5668	2.804	1.3896	0.16099	-3.108	4.535e+00	0.9027	1.1143	1.209e+00	0.6888
		(0.3552)	(0.3670)	(0.3737)	(0.2443)	(0.08393)	(0.4089)	(0.2448)	(0.2046)	(0.1781)	(0.02397)

Table 12: Point Estimates: SGI with first 5 good starts and 993 nodes (exact for degree ≤ 11).

Data	INFORM	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{21}	θ_{22}	θ_{23}	θ_{24}	θ_{25}
1	1	1.1450	1.597	1.6074	0.1021	-2.618	1.058e+00	0.5974	0.4649	5.236e-01	0.2344
		(0.3107)	(0.1193)	(0.1266)	(0.09039)	(0.1209)	(0.4442)	(0.08484)	(0.09264)	(0.1033)	(0.04551)
1	1	1.1449	1.597	1.6074	0.1021	-2.618	1.058e+00	0.5974	0.4649	5.237e-01	0.2344
		(0.3107)	(0.1193)	(0.1266)	(0.09039)	(0.1209)	(0.4442)	(0.08484)	(0.09263)	(0.1032)	(0.04551)
1	1	1.1451	1.596	1.6073	0.1021	-2.618	1.058e+00	0.5975	0.4648	5.237e-01	0.2344
		(0.3107)	(0.1194)	(0.1266)	(0.09041)	(0.1209)	(0.4442)	(0.08484)	(0.09268)	(0.1033)	(0.04552)
1	1	1.1450	1.597	1.6073	0.1022	-2.618	1.058e+00	0.5974	0.4649	5.237e-01	0.2344
		(0.3106)	(0.1193)	(0.1266)	(0.09039)	(0.1209)	(0.4442)	(0.08484)	(0.09264)	(0.1032)	(0.04551)
1	1	1.1450	1.597	1.6074	0.1022	-2.618	1.058e+00	0.5974	0.4649	5.236e-01	0.2344
		(0.3107)	(0.1193)	(0.1266)	(0.09040)	(0.1209)	(0.4442)	(0.08484)	(0.09265)	(0.1033)	(0.04552)
2	1	0.7407	1.203	0.9848	0.2777	-2.375	7.587e-01	0.3400	0.9026	6.870e-01	0.1487
		(0.2769)	(0.1909)	(0.1754)	(0.1211)	(0.09287)	(0.6337)	(0.1723)	(0.07105)	(0.04652)	(0.04783)
2	1	0.7407	1.203	0.9848	0.2777	-2.375	7.587e-01	0.3400	0.9026	6.870e-01	0.1487
		(0.2769)	(0.1909)	(0.1754)	(0.1211)	(0.09287)	(0.6337)	(0.1723)	(0.07105)	(0.04652)	(0.04783)
2	1	0.7407	1.203	0.9848	0.2777	-2.375	7.587e-01	0.3400	0.9026	6.870e-01	0.1487
		(0.2769)	(0.1909)	(0.1754)	(0.1211)	(0.09287)	(0.6337)	(0.1723)	(0.07105)	(0.04652)	(0.04783)
2	1	0.7407	1.203	0.9848	0.2777	-2.375	7.587e-01	0.3400	0.9026	6.870e-01	0.1487

		(0.2769)	(0.1909)	(0.1754)	(0.1211)	(0.09287)	(0.6337)	(0.1723)	(0.07105)	(0.04652)	(0.04783)
2	1	0.7407	1.203	0.9848	0.2777	-2.375	7.588e-01	0.3400	0.9026	6.870e-01	0.1487
		(0.2769)	(0.1909)	(0.1754)	(0.1211)	(0.09287)	(0.6337)	(0.1723)	(0.07105)	(0.04652)	(0.04783)
3	1	1.2592	1.119	1.3545	0.6938	-2.406	5.522e-01	0.2969	0.8445	4.096e-01	0.1774
		(0.3125)	(0.1512)	(0.1640)	(0.1213)	(0.1207)	(0.7791)	(0.1367)	(0.04399)	(0.1030)	(0.05693)
3	1	1.2594	1.119	1.3545	0.6940	-2.406	5.511e-01	0.2970	0.8445	4.096e-01	0.1773
		(0.3125)	(0.1513)	(0.1640)	(0.1213)	(0.1208)	(0.7806)	(0.1367)	(0.04400)	(0.1031)	(0.05695)
3	1	1.2592	1.119	1.3545	0.6939	-2.406	5.520e-01	0.2969	0.8445	4.096e-01	0.1774
		(0.3125)	(0.1512)	(0.1640)	(0.1213)	(0.1208)	(0.7794)	(0.1367)	(0.04399)	(0.1031)	(0.05693)
3	1	1.2592	1.119	1.3545	0.6939	-2.406	5.518e-01	0.2969	0.8445	4.096e-01	0.1774
		(0.3125)	(0.1512)	(0.1640)	(0.1213)	(0.1208)	(0.7796)	(0.1367)	(0.04399)	(0.1031)	(0.05694)
3	1	1.2592	1.119	1.3545	0.6939	-2.406	5.518e-01	0.2969	0.8445	4.096e-01	0.1774
		(0.3125)	(0.1512)	(0.1640)	(0.1213)	(0.1208)	(0.7796)	(0.1367)	(0.04399)	(0.1030)	(0.05693)
4	1	1.8109	1.863	1.7192	0.2741	-2.708	2.421e-01	0.8711	0.5005	8.894e-07	0.4690
		(0.2819)	(0.2974)	(0.2946)	(0.1895)	(0.1510)	(2.351)	(0.09297)	(0.1068)	(3.940E+04)	(0.04176)
4	1	1.8109	1.863	1.7192	0.2740	-2.708	2.418e-01	0.8711	0.5005	2.742e-06	0.4690
		(0.2819)	(0.2974)	(0.2946)	(0.1895)	(0.1510)	(2.354)	(0.09297)	(0.1068)	(1.278E+04)	(0.04176)
4	1	1.8109	1.863	1.7192	0.2741	-2.708	2.421e-01	0.8711	0.5005	4.146e-06	0.4690
		(0.2819)	(0.2974)	(0.2946)	(0.1895)	(0.1510)	(2.351)	(0.09297)	(0.1068)	(8453.)	(0.04176)

4	1	1.8109	1.863	1.7192	0.2740	-2.708	2.423e-01	0.8711	0.5005	2.813e-06	0.4690
		(0.2819)	(0.2974)	(0.2946)	(0.1894)	(0.1510)	(2.349)	(0.09297)	(0.1068)	(1.246E+04)	(0.04176)
4	1	1.8109	1.863	1.7192	0.2741	-2.708	2.421e-01	0.8711	0.5005	1.145e-06	0.4690
		(0.2819)	(0.2974)	(0.2946)	(0.1895)	(0.1510)	(2.350)	(0.09297)	(0.1068)	(3.061E+04)	(0.04176)
5	1	-0.4660	2.258	1.1551	0.3580	-2.115	1.550e-06	1.0433	0.2330	7.726e-01	0.3592
		(0.5278)	(0.2026)	(0.1986)	(0.1179)	(0.1670)	(4.379E+05)	(0.07440)	(0.3373)	(0.1017)	(0.05318)
5	1	-0.4920	2.317	1.2141	0.3340	-2.123	1.174e-07	0.9852	0.5095	7.826e-01	0.3627
		(0.5487)	(0.1999)	(0.1967)	(0.1192)	(0.1653)	(5.881E+06)	(0.07564)	(0.1469)	(0.1019)	(0.05232)
5	1	-0.4660	2.258	1.1551	0.3580	-2.115	3.531e-07	1.0433	0.2331	7.726e-01	0.3592
		(0.5278)	(0.2026)	(0.1986)	(0.1179)	(0.1670)	(1.922E+06)	(0.07440)	(0.3372)	(0.1017)	(0.05318)
5	1	-0.4660	2.258	1.1551	0.3580	-2.115	1.848e-06	1.0433	0.2331	7.726e-01	0.3592
		(0.5278)	(0.2026)	(0.1986)	(0.1179)	(0.1670)	(3.671E+05)	(0.07440)	(0.3372)	(0.1017)	(0.05318)
5	1	-0.4920	2.317	1.2141	0.3340	-2.123	1.785e-06	0.9852	0.5095	7.826e-01	0.3627
		(0.5487)	(0.1999)	(0.1967)	(0.1192)	(0.1653)	(3.870E+05)	(0.07564)	(0.1469)	(0.1019)	(0.05232)

Table 13: Point Estimates: Monomial with first 5 good starts and 983 nodes (exact for degree ≤ 11).

Furthermore, for each rule there are several datasets where the true data generating process (Table 2.) is not within the confidence interval formed from the point estimates and standard errors. For example, the true value of θ_{11} is 2, but the point estimates for datasets 2, 3, and 5 are never close for any of the rules. The point estimates for $\hat{\theta}_2$ are further from the 'truth' more often than those for $\hat{\theta}_1$. Consequently, the BLP model appears to suffer from finite sample bias, as shown in Chapter 3. This problem could also be exacerbated because we estimate the model without a pricing equation. Increasing the number of markets, T, will only improve the identification of θ_2 if product characteristics vary across markets. It would also be useful to compare the GMM standard errors to bootstrap standard errors.

We now look at these issues in more detail in 2.5.2.1 and 2.5.2.3.

2.5.2.1 *Impact of Quadrature on Optimization*

One encouraging result of our experiments is that the point estimates computed via MPEC + SNOPT for the polynomial-based rules are always the same for all starting values when the solver found a valid solution, unlike the pMC rules whose point estimates varied widely depending on the starting value. In general, SNOPT and KNITRO encountered numerical difficulties – typically an undefined computation for a conditional logit share of the form ∞/∞ – more often with the polynomial-based rules than pMC because the polynomial rules provide better coverage of extreme areas of the parameter space, even though the weights are quite small.[49] That the pMC point estimates vary widely depending on starting values indicates that the solver is finding false local maxima because of the inaccuracies of the pseudo-Monte Carlo approximation to the integral.

Another important issue is that approximating the share integrals is less important than accurately computing the gradient and Hessian of the GMM objective function and the Jacobian of the constraints which the solver uses to find a local optimum. The product rule and SGI affect solver convergence similarly, which is unsurprising because the SGI nodes, as mentioned in 2.2.3.2, are a subset of the product rule nodes. By omitting the nodes in the corners of the product rule

[49] Dubé et al. [2011] side-step these issues to some extent by using the MC draws which were used to generated their synthetic data to compute the market share integrals. Clearly, in a real world problem these shocks would not be observed by the econometrician. When we redraw these shocks, their code produces NaNs for some starting values. In addition, they use the same set of draws for each market share integral, s_{jt}.

lattice, SGI is less likely to evaluate the objective function, constraints, or the gradients at extremal points which produce NaNs and cause the solver to abort.[50]

Note that increasing the number of draws to $R = 10,000$ does not significantly improve the optimum found by SNOPT with a pMC rule (Tables 9 and 10). Many of the point estimates and values of the optimum at the solution still vary considerably depending on the starting value even when the solver reports that it has found a local optimum. Clearly, even with $10,000$ draws, pMC still introduces spurious local optima.

In the MPEC formulation of BLP, quadrature only affects the problem via the constraint equating observed and predicted market shares. With simulation, this constraint will be noisier and have local areas where simulation errors make it possible to find a local optima. Gandhi [2010]'s proof that the market share equation is invertible requires monotonicity which fails in this case. Furthermore, the optimizer adjusts parameters so that the spectrum of the mapping is less singular and has local basins of attraction. The different sizes of these basins affect how often solver finds them when searching for a local minimum of the GMM objective function.

2.5.2.2 *Differences in Objective Function Values*

We were initially surprised to discover in Tables 4-8 that the objective function values, f_k, do not agree at the optimal point estimates. This occurs because the value of the objective function in the MPEC formulation is $g^{'}Wg$ where $g = Z^{'}\xi$ are the moment conditions. Using MPEC, we solve for g as part of the optimization program: consequently, differences in g across quadrature rules at the local optimum found by the solver produce different values of the objective function. Initial investigations with a new C++ implementation using quad precision arithmetic and LU rook pivoting to increase solver stability appear to eliminate these differences so that only about 10 out 1302 values of the solver solution $(\theta_1, \theta_2, \delta, g)$ differ by more than 2%. However when using MATLAB, $\hat{\xi}$ varies considerably at the optima found by SNOPT even when $\hat{\theta}$ does not.

[50] If SNOPT encounters a NaN it will abort with EXIT=10 and INFO=72.

2.5.2.3 *Simulation, Identification, and Standard Errors*

When computing standard errors, we found that simulation – unlike the polynomial rules – produces excessively tight values and will lead researchers to think that some parameters are estimated precisely when they are not. Examining the standard errors (Tables 9-13) shows that, in general, the pMC standard errors are much smaller than those computed with polynomial rules. For some datasets, such as datasets 4 and 5, the polynomial rules produce standard errors on the order of 10^4 or larger vs. pMC errors of 10^{-1} even with $R = 10,000$ draws. For example, compare the results for $\hat{\theta}_{21}$ and $\hat{\theta}_{24}$ for dataset 4 and $\hat{\theta}_{21}$ and $\hat{\theta}_{23}$ for dataset 5. Standard errors computed using pMC show apparently precise estimates when in fact the Hessian is ill-conditioned. Because pMC introduces spurious local optima and, concomitantly, pockets of higher curvature it produces standard errors which are too tight. Consequently, pMC can mask poor identification and practitioners will miss an important diagnostic that the objective function is nearly flat. In fact, as the order of differentiation increases, pMC performs increasingly poorly. Polynomial-based rules do not suffer from this problem because they approximate the level, gradient, and Hessian correctly: if a parameter had huge standard errors for one dataset and rule, then it had huge standard errors for all rules and starts. Nevertheless, the quadrature rules did not reliably detect large standard errors: the Gauss-Hermite product rule with 7^5 nodes detected 4 cases out of $10 \times 5 = 50$ parameters estimated;[51] SGI and the monomial rule found 3 of the 4 found by the product rule; and, pMC, even with $R = 10,000$ draws, failed to find any. Recently, we began replicating these results using the higher precision implementation discussed in Chapter 3. Our initial results show that when using higher precision arithmetic, we can reliably compute the same large standard errors for the same datasets and parameters, θ, when using any of the polynomial rules. Even with this BLP implementation, the pMC rules still produce anomalously tight standard errors. Consequently, pMC quadrature rules will cause a downward bias in standard errors and mask identification problems.

In BLP, the substitution patterns are 'diffuse' and all goods are substitutes for each other as opposed to the 'local' substitution patterns in pure characteristics models, where cross-price elasticities are non-zero for only a finite set of products (E.g., Berry and Pakes [2007],

[51] I.e., we estimated ten parameters, $\hat{\theta}$, for five starts for each rule and dataset

Shaked and Sutton [1982]). Consequently, the model is very sensitive to both sampling error in the observed market shares, S_{jt}, and simulation error in the computation of predicted market shares, s_{jt} [Berry et al., 2004b]. Particularly as J increases (And, Berry et al. [2004b] require $J \to \infty$ to prove that BLP is consistent and asymptotically normal) small simulation or sampling errors considerably affect the value of ξ which is computed in the traditional NFP BLP implementations.

The small sample properties of GMM is another potential source of difficulty in estimating θ_2 parameters well. Altonji and Segal [1996] show that optimal minimum distance (OMD) estimators – i.e. GMM with the optimal weighting matrix – perform poorly in small sample estimates of the variance because the shocks which make the variance large also tend to increase the variance of the variance. Consequently, because θ_2 measures the standard error of the random coefficients it is probably estimated with downward bias. This correlation could also explain why $\text{Var}\left(\theta_2\right)$ is often surprisingly large: when estimation uses more accurate, polynomial rules, the correlation caused by simulation is no longer an issue.

2.5.3 *Bias of Point Estimates*

Subsequent to completing the research for this chapter, I developed infrastructure in Chapter 3 to compute a variety of statistics to characterize the finite sample performance of the BLP estimator. Using these tools, I computed these metrics for both traditional BLP instruments, which use the characteristics of rival products, and supply-side cost shifters. Table 33 compares the results using pMC and SGI for the Data Generating Process (DGP) described in Chapter 3 for 2 markets and 24 products. The SGI rule is accurate for monomials of degree 11 or less. To ensure that comparisons were fair – i.e. the computational cost was roughly the same for both rules – I used the same number of nodes for SGI and pMC. SGI is clearly superior to pMC for all parameter estimates and metrics: bias, mean absolute deviation, median absolute deviation, and RMSE. This result shows that finite sample bias is actually worse for pMC than SGI.

2.5.4 *Computational Burden*

Figure 4 shows the costs of the different quadrature rules and how they vary across replications and experiments. Execution time scales as the number of nodes, although the accuracy of the approximation of the integrals can affect how many major and minor iterations SNOPT requires to find an optimum. Note that the Gaussian-Hermite product rule uses $7^5 = 16,807$ nodes but have only slightly larger median and mean run times compared to the pMC rule with $10,000$ nodes because better rules seem to facilitate solver convergence.

2.6 CONCLUSION

A head-to-head comparison of Monte Carlo and polynomial-based quadrature rules for approximating multi-dimensional integrals of moderate size shows that monomial rules provide superior accuracy at a computational cost which is at least an order of magnitude smaller. Monomial rules are marginally more difficult to implement than pMC, requiring a few well-defined permutations of the nodes and weights found in a table look-up. An even easier option is to use sparse grids integration which can generate a set of nodes and weights that provide comparable accuracy, often with only a few more nodes. An important area for future research is to develop monomial rules which exploit the fact that many functional forms in Economics are well-behaved.

When we applied these quadrature methods to BLP, it became clear that the choice of quadrature rule affects the model's results, including the computed value of product-market share integrals, the values of the point estimates, and the standard errors. In particular, pseudo-Monte Carlo rules produce very different point estimates for different starting values – even with very large numbers of draws – unlike the polynomial rules which always produce the same optimum. pMC rules also generate excessively tight standard errors, hiding potential identification problems in the local basins of attraction created by simulation noise.

Using a high-quality quadrature rule, then, provides an easy way to improve the accuracy and efficiency of many numerical projects.

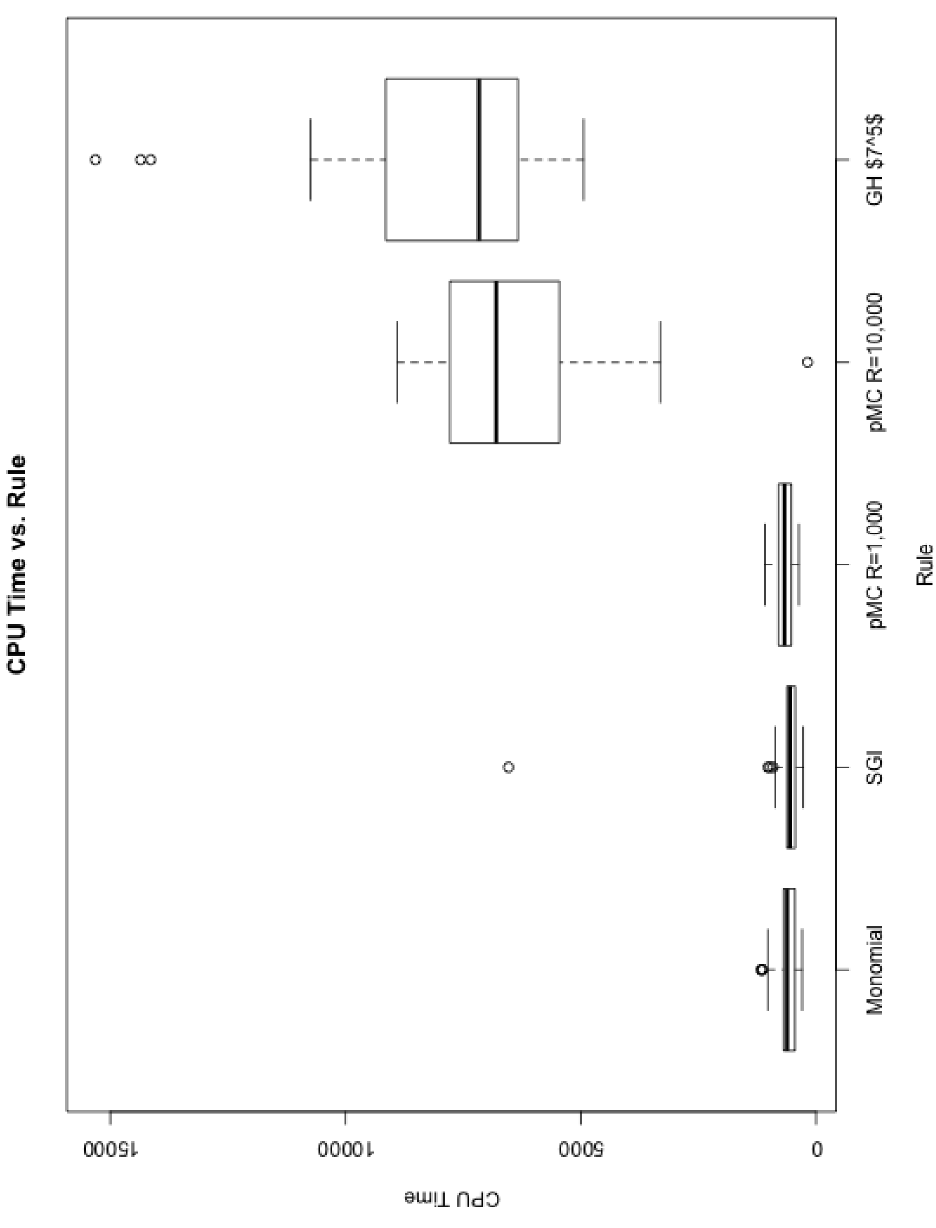

Figure 4: Computational Cost by Rule

3

A LARGE-SCALE STUDY OF THE SMALL SAMPLE PERFORMANCE OF RANDOM COEFFICIENT MODELS OF DEMAND

Berry et al. [1995]'s model of demand for differentiated products (BLP, hereafter) developed a flexible approach to modeling demand by allowing for consumer heterogeneity as well as unobserved product-market quality – especially while working with aggregate, market share data. Despite BLP's importance, there are few results about its finite sample behavior. Simulation experiments, in theory, provide a tool to answer such questions but the computational and numerical difficulties have prevented researchers from performing any realistic studies. Nevertheless, by utilizing recent advances in High Throughput Computer (HTC) systems[1] and a fast, reliable implementation of BLP, I show that a large-scale simulation approach is now feasible and compute the finite sample bias for a variety of scenarios.

Recent computational results on optimization [Su and Judd, 2010, Dubé et al., 2011] and numerical integration (See Chapter 2.) show that the model has subtle numerical problems and that considerable care is required to obtain reliable estimates. A careful practitioner must worry about optimization with many non-linear constraints, accurate numerical approximation of high dimensional integrals, solving non-linear systems of equations, and numerical overflow/underflow. When estimation is scaled up for a series of Monte Carlo experiments these problems can become prohibitively difficult – both in terms of numerical stability and computational cost – even when computing a limited number of replications.[2] Those Monte Carlo studies that exist focus on only one market and often take computational short-cuts.[3]

[1]Computer scientists refer to computer clusters designed for moderate or large-scale parallel computing as either *High Performance Computing* (HPC) or *High Throughput Computing* (HTC). HPC systems typically run jobs of up to several hours which are often tightly coupled (synchronous) whereas HTC clusters run much longer serial jobs which are independent of each other.

[2]For example, this study estimated BLP over 320,000 times and used 94,325 CPU-hours – not counting jobs which failed – once all the experiments, replications, multiple starts, and instrumentation strategies were accounted for. See Section 3.5.5 for a discussion of computational cost.

[3]For example, Armstrong [2011] uses only 10 pMC quadrature nodes and fixes the scale of the random coefficients.

Consequently, a large-scale, numerically rigorous simulation study with data generated from a structural model has appeared infeasible, especially given that estimation can push the limits of desktop computing.[4]

My implementation overcomes previous numerical difficulties – especially those of the traditional Nested Fixed Point (NFP) estimation strategy – by exploiting current best practice in optimization [Su and Judd, 2010] and multi-dimensional numerical integration (See Chapter 2.). I conduct a series of simulation experiments to quantify the finite sample performance of the BLP model for realistic numbers of markets, numbers of products, and instrumentation strategies. This chapter, then, has two objectives: to demonstrate the power of modern computational technology for solving previously intractable problems in Economics via massive parallelization and to characterize the finite sample behavior of the BLP estimator.

3.1 OVERVIEW

Berry et al. [1995] has influenced Industrial Organization and applied Economics considerably. Their insight was to combine a flexible model of heterogeneity and unobserved product-market characteristics with an GMM-based estimation strategy which controlled for price endogeneity. Despite the model's ubiquity, there are few results on BLP's finite sample behavior or the impact of instrument type and quality. Furthermore, the complexity of the model has made it difficult to implement reliably [Dubé et al., 2011], let alone study via simulation. Consequently, there are few papers which examine the theoretical properties of the model [Berry et al., 2004b, Armstrong, 2011] and previous Monte Carlo experiments only consider one market.[5]

In this paper, I develop an implementation of BLP which makes a large-scale simulation study possible and quantify how the BLP model performs for a variety of market sizes, numbers of products, and instrument choices. I examine both how market size affects bias and the effectiveness of different instrumentation strategies.

---

[4] Generating synthetic data from a structural model faces similar numerical issues and, consequently, is at least as challenging as estimating the BLP model.

[5] Researchers focus on the case with one market because for most BLP datasets the variation is primarily with-in market. Consequently, identification must come from $J \to \infty$.

Only a few asymptotic results currently exist. Berry et al. [2004b] proves consistency and asymptotic normality for one market as the number of products $J \to \infty$, given some high level regularity conditions. But, these conditions do not prove exactly which price setting mechanisms produce consistent and asymptotically normal (CAN) estimators. Armstrong [2011] solves this problem for Bertrand-Nash price equilibrium using a similar setup plus a fully-specified structural model of supply and demand. He proves that the BLP model cannot be identified from the traditional product characteristics instruments, but that the model is identified with exogenous, supply-side cost shifters.[6] The intuition for Armstrong's result comes from the Bertrand-Nash first order conditions (FOCs):

$$0 = s_{jt} + \left(p_{jt} - mc_{jt}\right) \frac{\partial s_{jt}}{\partial p_{jt}}$$

for market shares s_{jt}, price p_{jt}, marginal cost mc_{jt}, product $j \in J$, and market $t \in T$. For logit demand with linear utility this equation simplifies to

$$\begin{aligned} p_{jt} &= mc_{jt} - \frac{s_{jt}}{\partial s_{jt} / \partial p_{jt}} \\ &= mc_{jt} - \frac{1}{\beta_{price}\left(1 - s_{jt}\right)} \end{aligned}$$

where $\beta_{price} < 0$ is the coefficient on price in the utility. As the number of products goes to infinity, market shares approach zero (assuming there are no mass points), and the price equation becomes indistinguishable from marginal cost plus a constant markup. Without the non-linearities of the markup term, the matrix of product characteristics instruments becomes perfectly collinear – i.e. the model is unidentified [Armstrong, 2011]. Armstrong [2011] also proves that there is less bias when a firm produces more products. He closes by arguing that simulation studies must use a structural model to generate data instead of the more common reduced form approach or the data will lack the correct statistical properties.

I would expect that increasing the number of markets could help offset the lack of identification as $J \to \infty$ if there is sufficient variation

[6] The traditional 'BLP' instruments use either characteristics of other goods produced by the same firm or characteristics of goods produced by rival firms.

across markets. In practice, product characteristics for differentiated consumer products are often fixed for all markets, in which case results on dual asymptotics in panel data, such as Phillips and Moon [1999] and Phillips and Moon [2000], are unlikely to provide useful intuition for the properties of the BLP estimator.

For some time researchers have been concerned about finite sample bias in GMM (E.g. Altonji and Segal [1996]). Recent developments such as Empirical Likelihood [Kitamura, 2006] have made some progress. Conlon [2010], for example, uses empirical likelihood (EL) to estimate a dynamic BLP-style model of demand and reports that it performs much better for small samples than the traditional BLP Nested Fixed Point GMM estimator.

Simulation experiments of the scope and scale needed to study the BLP model require fast, reliable hardware and software. Using recent advances in software algorithms, software engineering, and computer hardware, I develop a suitable implementation which can run multiple replications in parallel on a modern cluster. I use Mathematical Programming with Equilibrium Constraints (MPEC) [Su and Judd, 2010] and SNOPT, one of the best solvers currently available, to overcome the limitations of the NFP approach, especially the slow and unreliable convergence of the inner loop contraction mapping.[7] In addition, I also use a polynomial-based quadrature rule to correct the deficiencies of pMC integration: this quadrature rule produces more accurate point estimates and standard errors as well as much faster software (See Chapter 2.). Finally, modern High Throughput Computing infrastructure enabled me to estimate the model thousands of times in parallel on the Petascale Active Data Store (PADS) cluster at the Computation Institute at the University of Chicago.

Like many simulation studies, I rely on synthetic data and, following Armstrong [2011], generate data from a structural model. The data generating process includes exogenous cost shifters as well as correlation between the unobserved product-market shock, ξ_{jt}, and the marginal cost shock, η_{jt}. Generating this data is non-trivial because the technical challenges are the same as estimating the BLP model; in addition, it is necessary to solve for the Bertrand-Nash price equilibrium.

[7] The traditional NFP implementation of the model relies on an outer loop to perform GMM and an inner loop 'contraction mapping' to recover parameters by inverting the market share equation.

This chapter proceeds as follows. First, I quickly review the Berry et al. [1995] model of demand and how to estimate it using MPEC; then, I explain the details of my implementation of BLP; Section 3.3 describes the synthetic data set; in Section 3.4, I describe the Monte Carlo experiments and results; finally, I conclude.

3.2 REVIEW OF THE BLP MODEL

To provide a common ground for readers and introduce notation, I provide a quick review of Berry et al. [1995]'s model of demand for differentiated products. Their model augments a mixed logit (random coefficient) functional form with an unobserved product-market shock, ξ_{jt}, which makes it possible to fit predicted shares to observed shares. The original estimation strategy uses a GMM-based, NFP algorithm to estimate the model from aggregate market share data for T markets and J products. For a more thorough introduction, see Nevo [2000b] or Train [2009].

3.2.1 *Demand*

The demand side of the model assumes that consumer i receives utility u_{ijt} when he purchases product $j \in J$ in market $t \in T$ with characteristics x_{jt}:

$$\begin{aligned} u_{ijt} &= v_{ijt} + \epsilon_{ijt} \\ v_{ijt}\left(\beta_i\right) &= x_{jt}^{'}\beta_i + \xi_{jt} \end{aligned}$$

where the coefficients β_i depend on each consumer's type. The product characteristics, x_{jt}, include price p_{jt} to simplify the notation. Researchers in Economics typically assume that β_i is normally distributed as $N\left(0, \Omega\right)$ and that ϵ_{ijt} is Type I Extreme value.[8] ξ_{jt}, the product-market shock, is observed by consumers and firms but not the econometrician. A consumer's taste parameters $\beta_i = \overline{\beta} + \Sigma \nu_i$, where $\overline{\beta}$ is the mean preference parameter and $\nu_i \sim N\left(0, 1\right)$ is i's taste shock, scaled by Σ, the Cholesky factor of $\Omega = \Sigma\Sigma^{'}$. BLP researchers often define

[8] To focus on the key features of the model, I ignore more general formulations where β_i may depend on consumer characteristics. See Berry et al. [2004a], Nevo [2000a], and Nevo [2001] for examples with microdata.

$\theta_1 = \overline{\beta}$, $\theta_2 = \text{vech}(\Sigma)$, and $\theta = (\theta_1, \theta_2)$. The parameters θ are what we as econometricians hope to recover.

By convention, utility is repackaged into mean utility

$$\delta_{jt}(\xi; \theta_1) = x'_{jt}\overline{\beta} + \xi_{jt}$$

which is constant for all consumers and the total type-specific shock

$$\mu_{ijt} = x'_{jt}\Sigma\nu_i.$$

Then utility is $u_{ijt} = \delta_{jt}(\xi; \theta_1) + \mu_{ijt} + \epsilon_{ijt}$.

These assumptions produce a convenient, closed form solution for market shares conditional on consumer type:

$$s_{jt}(x\,|\beta) = \frac{\exp(v_{jt}(\beta))}{1 + \sum_{k \in J} \exp(v_{kt}(\beta))},$$

where I have assumed that the utility of the outside good is 0: i.e., $u_{i0t} = 0 + \epsilon_{i0t}$. Note, I have dropped the suffix i which is implied because the conditional shares are a function of the taste parameters, β. Also, β_{price} is $-\alpha$ in standard BLP notation. Thus, aggregate demand is

$$s_{jt}(x\,|\theta) = \int s_{jt}(x\,|\beta)\, dF(\beta\,|\theta)$$

where θ parametrizes the distribution of individual preferences. Assuming a normal distribution for preferences,

$$\begin{aligned} s_{jt}(x\,|\theta) &= \int s_{jt}(x\,|\beta)\, d\Phi(\beta\,|\theta) \\ s_{jt}(x\,|\theta) &= \int s_{jt}(x\,|\beta)\, \frac{1}{\det(\Sigma)} \phi\left(\Sigma^{-1}\left(\beta - \overline{\beta}\right)\right) d\beta. \end{aligned}$$

Berry [1994] observes that the mean utilities, and hence ξ_{jt}, can be obtained by inverting the market share equation

$$S_{jt} = s_{jt}(\delta; \theta_2),$$

which equates observed market shares, S_{jt}, and predicted market shares, s_{jt}. BLP recover ξ_{jt} from the mean utilities, δ_{jt}, and use it to estimate the model's parameters via GMM.

Sampling error is a concern in these models. BLP provide statistical arguments to support their claim that they observe market shares with sufficient accuracy to ignore sampling error. In this paper, I assume that there is no sampling error in order to focus on the core issues of their model. See Section 3.4.3 for further discussion.

3.2.2 *Supply Side*

Following Berry et al. [1995], I assume that marginal cost is linear in cost shifters z_{jt}:

$$\log\left(mc_{jt}\right) = z_{jt}^{'}\gamma + \eta_{jt}.$$

In practice, the cost shifters include product characteristics, x_{jt}, as in BLP, as well as other factors affecting cost such as the wages and commercial rental rates. Because the profit first order conditions show that price is marginal cost plus a mark-up,

$$p_{jt} = mc_{jt} - \frac{s_{jt}}{\partial s_{jt}/\partial p_{jt}},$$

cost shifters should help identify prices, which are surely endogenous because of correlation between p_{jt} and ξ_{jt}, especially when the mark-up term approaches 0 as Armstrong [2011] proved. Another important benefit of specifying a supply side is that including it in the GMM moments helps estimate structural parameters more precisely. Of course the risk of using a full structural model is that the supply-side could be misspecified.[9]

[9] Which supply-side specifications satisfy Berry et al. [2004b] is an open question.

3.3 IMPLEMENTING RELIABLE, HIGH PERFORMANCE ESTIMATION SOFTWARE

Implementing the BLP model is a formidable engineering challenge. Estimation results are quite sensitive to optimization software,[10] coding the NFP algorithm correctly, and using tight inner loop tolerances [Dubé et al., 2011]. An implementation must also handle the numerical instabilities inherent in the underlying logit form, approximate multidimensional integrals accurately, and deal with (a few) weak instruments. Furthermore, running hundreds of thousands of simulation experiments requires fast and reliable estimation software because the scale of the experiments ensures that the estimation software will inevitably encounter cases which expose defects in the code. To overcome these issues, I utilize the current best practices for estimation, numerical integration, and parallelization. In addition, I use modern software engineering techniques to produce code which is fast, robust, and easy to write as well as to verify. The following sections describe these issues in more detail.

3.3.1 *Estimation of the BLP Model using MPEC*

My implementation uses the MPEC algorithm advocated by Su and Judd [2010] and further refined in Dubé et al. [2011] (DFS hereafter). MPEC is easier to code and more robust than the traditional nested fixed point strategy because MPEC expresses the entire problem in more intuitive terms than NFP, i.e. as the minimization of a GMM objective function subject to constraints.[11] Dubé et al. [2011] prove that these two approaches are equivalent; consequently, there are no adverse tradeoffs in using MPEC.

Accordingly, I solve

$$\begin{aligned} \max_{\theta,\delta,\psi} \quad & \psi^{'} W \psi \\ \text{s.t.} \quad & s\left(\delta\left(\xi\right);\theta\right) = S \\ & \psi = Z^{'}\xi. \end{aligned}$$

[10] Currently, the best constrained optimizers for non-linear problems include CONOPT, KNITRO, and SNOPT, but this is an area of active research.

[11] The NFP algorithm consists of an outer loop to perform GMM and an inner loop to invert the market share equation.

To get trustworthy point estimates, researchers must use good quadrature rules (See Chapter 2) and a high-quality solver. Although Dubé et al. [2011] use the KNITRO solver [Byrd et al., 2006], I prefer SNOPT 7.0 [Gill et al., 2002] because it is more robust than KNITRO for BLP-type problems (See Chapter 2.).[12] With its default settings, SNOPT still struggles with some rank deficient regions. Enabling LU rook pivoting, a slower but more reliable pivoting algorithm for factoring the basis matrix, considerably improves SNOPT's ability to find a local optimum.[13] Researchers should weigh these issues carefully in order to choose a solver that is appropriate for their problem.[14]

The above comments are based on a prior version of DFS's code which used KNITRO 6.0.0. DFS report that the latest version of their code, which now uses an analytic Hessian, is considerably faster than the previous version. KNITRO 7.0.0, the current version, also has improved support for dense linear algebra. As the number of markets and products grows, as discussed in Section 3.5.3, SNOPT has increasing difficulty finding a feasible point. KNITRO may be more successful at finding points which satisfy constraints prior to commencing optimization. Consequently, KNITRO may now be a better choice than SNOPT for BLP, but more research is required.

3.3.2 *Quadrature*

Poor numerical integration adds instability and unreliability to the BLP model even when using MPEC. Chapter 2, building on work by Heiss and Winschel [2008], discusses how pMC integration is inferior to polynomial-based methods such as monomial rules and sparse grids. Polynomial rules are much faster and approximate the market share integrals and their gradients more accurately. pMC produces multiple local optima and, hence, point estimates as well as artificially tight standard errors, masking identification problems. Polynomial based rules produce the same point estimates – even when passing different starting values to the solver – more often than pMC. In these experiments, I use a 165 node SGI rule which is exact for all three-dimensional polynomials whose total order (sum of the exponents of

[12]SNOPT uses an SQP algorithm which can overcome some non-convexity unlike KNITRO's Interior Point (IP) algorithm. In addition, SNOPT 7.0 has support for rank deficient systems.

[13]I found that LU rook pivoting increased computation time by a factor of 2.

[14]In practice, it is worth experimenting with different algorithms as well as the tuning parameters of the solver.

each term) is less than degree 11. This rule provides a fast, accurate approximation for the BLP market share integrals.

Many applied studies fail to take enough draws when using pMC methods to approximate integrals. Using the machinery of this project, I actually compute the bias of using pMC instead of a sparse grids rule in Section 3.5.4. These results show that sloppy quadrature can induce considerable bias into estimates and that SGI is much less biased than pMC by all metrics.

3.3.3 *Numerical Stability and Software Engineering*

The shape of the BLP optimization problem and functional form choices often produce numerical instabilities including overflow and underflow, although the extent of these problems depends on the actual data. Consequently, I use the SNOPT solver (See Section 3.3.1, above.), careful implementation, and higher precision arithmetic to minimize numerical problems.

Using the logit distribution for the market shares conditional on consumer type makes the model susceptible to arithmetic overflow and underflow, when software generates numbers which are, respectively, too big or small in magnitude to represent.[15] The IEEE-754 floating-point specification requires that overflow produce the value `Inf`. Unfortunately, many researchers follow the common practice of setting the objective function to some very large value such as 10^{20} when overflow occurs which makes the objective function non-differentiable and flat for certain values and can prevent the solver from making progress. In some cases, it is possible to transform the logit by renormalizing utility with respect to the largest utility in the choice set. When there is an outside good this approach often converts overflow into underflow, leading to problems with zero share values.

To solve problems with overflow yet guarantee code which executed quickly, I implemented the BLP model in C++ and used higher precision floating point arithmetic (`long double` instead of `double`, i.e. 128-bit instead of the default 64-bit representation). In most cases, I found that this was sufficient, but it is possible to work in even higher precision if necessary. The extra robustness from switching to the 128-bit representation only cost a factor of two in execution speed.

[15] Goldberg [1991] provides a review of essential knowledge about floating-point arithmetic for scientists.

The difficulty in identifying the scale parameters for the random coefficients provides another example of the numerical challenges in developing a reliable implementation. In simulations for both this project and for Chapter 2, high quality quadrature rules lead to huge asymptotic standard errors, indicating that these parameters cannot be estimated precisely. Furthermore, the results of the experiments in this Chapter show that asymptotic inference is not yet valid for the experiments I consider. Armstrong [2011] sidesteps this issue by fixing the variance matrix:

> For both tables, I treat the variance of the random coefficients, σ, as known and solve for the IV estimates of the other parameters with σ fixed at its true value. Since the resulting IV estimator is the closed form solution to a system of linear equations, this eliminates potential concerns that negative results may come from a failure to minimize a nonlinear GMM objective function. Estimators that perform poorly can be expected to do even worse when σ needs to be estimated as well.

When the BLP instruments are highly collinear – which is often the case – the weighting matrix has a higher condition number and the solver finds many more local optima. Furthermore, as the number of products increases, this problem becomes worse.[16] With reduced form data, such as that used in Dubé et al. [2011] or Chapter 2 and accurate quadrature rules, SNOPT usually finds the same local optima.

Despite these considerable engineering precautions, the BLP model remains numerical unstable because small share values produce even smaller derivatives. For many datasets the Jacobian elements of the market share constraints consist of ones along the diagonal and nearly zero off diagonal elements. In practice, many of these elements are so small that most optimizers set them equal to zero.[17] Consequently, the solver can have difficulty making progress. I address these problems by using Dubé et al. [2011]'s trick of specifying the market share constraint in terms of the logarithm of the market shares. I also impose

[16] SNOPT"s memory requirements also explode.

[17] For example, when I examined a synthetic data set which produced a variety of valid point estimates for different starting values, I found that the largest off-diagonal element had magnitude 8.5×10^{-7} and the median magnitude was 1.6×10^{-15}. Another dataset with the same number of markets and products always produced the same point estimates regardless of starts but had a median off-diagonal element of 5.3×10^{-5}.

sensible box constraints and compute feasible starting guesses, as described in Section 3.3.3.1, to facilitate solver convergence. Nevertheless, I find that for many data sets the point estimates can vary considerably for different starting values, even when the solver reports that the optimality conditions are satisfied.

3.3.3.1 *Computing a Good Starting Guess*

Computing a good starting guess can considerably accelerate a solver's convergence by avoiding regions of the parameter space which are infeasible. Many optimization algorithms fail or are considerably slower if the initial guess does not satisfy the constraints because the solver must first attempt to find a feasible point. To facilitate finding an optimum and accelerate solver convergence, I compute each initial guess so that the optimization constraints are approximately satisfied. My procedure for computing each start is based on the method in Dubé et al. [2011]:

1. Estimate mean utilities from a simple logit model of demand

$$\begin{aligned}
\log\left(S_{jt}\right) - \log\left(S_{0t}\right) &= \log\left(s_{jt}\right) - \log\left(s_{0t}\right) \\
&= \delta_{jt} \\
&= x_{jt}^{'}\theta_1 + \xi_{jt}
\end{aligned}$$

   exploiting the fact that the denominators in the logit expression for the predicted shares of the inside and outside goods are identical.

2. Estimate $\widehat{\theta_1}$ using 2SLS with either BLP or cost-shifter instruments.

3. Draw additional starting values for the scale parameter, θ_2^{guess}, about $1/2 \times |\theta_1|$.

4. Invert the market share equations for each market individually by iterating Berry's contraction mapping [Berry, 1994] 1,000 times for each market.[18] I invert each market separately to ex-

[18] In practice, Berry's mapping is very inefficient, unreliable, and can require more than 1,000 iterations to converge if it converges at all (See Chapter 2.). I used it because it was easy to implement, although better algorithms exist [Reynaerts et al., 2010]. Nevertheless, in some cases these points still did not satisfy the constraints, probably because the contraction mapping failed to converge.

ploit the block diagonal structure, which avoids problems where blocks require different scaling or stopping conditions. This approach produces a more robust approximation for the mean utility, δ_{jt}.

5. Recover guesses for θ_1^{guess} and ξ_{jt} from δ_{jt} using 2SLS.

Despite using DFS's procedure to generate good starting guesses, I found that some starting values produced different optima, each of which satisfied the solver's optimality criteria because the BLP problem is often nearly flat, rank deficient, or poorly identified. Consequently, I generate 50 different starts for each replication and then restart the solver at each of the optima found for the original starts. This strategy improves the likelihood of finding the true global optimum for each replication. In most cases the difference between original and restarted optima are much less than 10^{-4}.[19]

3.3.3.2 *Verification*

Credible numerical work rests on intensive verification because there is always another bug lurking, no matter how much you have tested and debugged. Because I based my implementation on the relatively stable code in Dubé et al. [2011] I had the luxury of comparing my results to those from their MATLAB implementation. In addition, I verified derivatives using finite difference, checked quadrature rules by comparing computed moments to closed form results, and computed other checks to validate that data was loaded and processed correctly. I also took advantage of engineering tools including valgrind [Nethercote and Seward, 2007a,b] to check for memory leaks and the GDB debugger to ensure that the software executed as expected. C++'s strong typing system ensured that all expressions were conformable and used appropriate types – i.e. I disabled all implicit type conversions (or casts, to use C-language terminology). Lastly, I loosely adhered to the Test Driven Development philosophy and wrote functions which did one thing well to facilitate testing. Of course, one must stop writing unit tests at some point.

[19] In a few cases the restart failed to solve, probably because SNOPT builds up an approximation to the Hessian as optimization proceeds.

3.3.4 *Hardware and Parallelization*

After prototyping the estimation and data generation code on a MacBook Pro, I migrated the software to the Petascale Active Data Store (PADS) cluster at the University of Chicago. I ran the simulations on PADS's analysis cluster which consists of 48 nodes, each with eight Nehalem 2.66 GHz cores and 24 GB RAM for a total of ~4.25 teraflops. Like most clusters, PADS has a resource manager and scheduler which facilitate running multiple jobs in parallel. I use a parameter sweep architecture to parallelize the estimation which works extremely well when running many independent jobs, such as estimating a model repeatedly as part of a Monte Carlo experiment. The PADS cluster's Torque/Portable Batch System (PBS) job manager facilitates running many jobs which differ only in their input because PBS passes a unique, user-defined ID to each instance which can be used to determine the input. See Listing 2 in Section A.2 for an example PBS script and further discussion.

3.4 Synthetic Data Generating Process

Much of the power of simulation studies comes from the researcher's knowledge of the true data generating process. Following Armstrong [2011], I generate synthetic data from a fully-specified model of demand and supply to ensure that the data has the correct statistical properties, such as product characteristics losing their power to identify the model as $J \to \infty$. I tuned parameter values to approximate real world data generating processes to the extent that it was reasonable and possible to do so. I also allow for correlation between the product-market unobservable, ξ_{jt}, and the unobserved marginal cost shock, η_{jt}. Finally, I generated product characteristics and marginal costs so that the product characteristics and shocks for experiments with more markets or products add more products or markets to the existing data. For example, the product characteristics, cost shifters, and corresponding shocks for the first 12 products in the first market of an experiment with $T = 50$ and $J = 100$ are identical to those for an experiment with one market and 12 products. Of course, prices and market shares will differ because of the different Bertrand-Nash price equilibria. This approach ensures that asymptotics drive results instead of random changes in the data.

Generating synthetic data encounters the same numerical problems as estimating BLP (See Section 3.3). In addition, it is necessary to solve for Bertrand-Nash price equilibria, find realistic parameter values, and deal with weak instruments. I discuss these issues below.

3.4.1 *Simulation of Data*

I perform experiments to estimate BLP models for several different combinations of markets and products, where the number of markets $T \in \{1, 10, 25, 50\}$ and the number of products $J \in \{12, 24, 48, 100\}$. Each firm produces two products to ensure that the traditional BLP instruments formed from characteristics of rival goods are not perfectly collinear.

I generate the data as follows:

1. Draw observed product characteristics for each product, market, and replication. I use the same observed product characteristics for each market in a replication. For each experiment, draw the data so that experiments with larger numbers of products or markets contain data from smaller experiments.

2. Draw unobserved product characteristics ξ_{jt} and unobserved costs η_{jt} where

$$\begin{pmatrix} \xi_{jt} \\ \eta_{jt} \end{pmatrix} = \begin{pmatrix} \mathrm{U}(0,1) \\ \mathrm{U}(0,1) \end{pmatrix} + (\mathrm{U}(0,1) - 1) \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

   This process ensures that $\mathbb{E}\left[\xi_{jt}\right] = \mathbb{E}\left[\eta_{jt}\right] = 0$ and that these shocks are correlated, which can be an important source of endogeneity in real data.

3. Compute marginal costs. I specify

$$\log c_{jt} = \begin{pmatrix} 1 & x_{jt} & z_{jt} \end{pmatrix} \gamma + \eta_{jt},$$

   where z_{jt} is an exogenous supply-side cost shifter.

4. Solve for Bertrand-Nash equilibrium prices using the Path 5.0.00 solver (Dirkse and Ferris 1995, Ferris et al. 1999, Ferris and Munson 2000), as described further in Section 3.4.2.

5. Compute observed shares at equilibrium using a sparse grids integration rule, which approximates the share integrals efficiently and accurately.

6. Compute 50 starting guesses to estimate the model for each replication, as discussed in 3.3.3.1.

I do not draw unobserved consumer characteristics (i.e. β_i) because I assume that there is no sampling error, as discussed in 3.4.3. Also, I draw some data, such as product characteristics, from a uniform distribution to ensure that the data is bounded. I generate the data from the following 'true' parameter values:

$$\begin{aligned}
\theta_1 &= \begin{pmatrix} 3 \\ 2 \\ -5 \end{pmatrix} \\
\theta_2 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
\gamma &= \begin{pmatrix} -2.0 \\ 0.5 \\ 0.5 \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}\left[\begin{pmatrix} \xi_{jt} \\ \eta_{jt} \end{pmatrix}\right] &= \begin{pmatrix} 2.0 & 0.1 \\ 0.1 & 2.0 \end{pmatrix} \\
\mathrm{Var}\left[x_{jt}\right] &= 6.0 \\
\mathrm{Var}\left[z_{jt}\right] &= 0.4
\end{aligned}$$

3.4.1.1 *Issues Involved in Choosing Parameter Values*

The numerical problems inherent in the BLP model made it difficult to choose parameter values which were realistic yet reliably produced a price equilibrium. These problems became more acute as the number of products increased because increasing the number of products decreases share values, produces even smaller derivatives, and causes off-diagonal elements in the constraint Jacobian or first order conditions to vanish. Also, the instruments become more collinear as T and

J increase, compounding the numerical difficulties. In addition large parameter values can produce numeric overflow, especially for data drawn from the normal distribution, which has long tails. The use of polynomial-based quadrature rules exacerbates this problem by approximating the tails of the distribution more accurately; with a pMC rule, researchers often avoid this problem by taking draws until they find a set of nodes which are more stable. Finally, β_{price} must be large enough in magnitude so that demand is elastic, otherwise the Bertrand-Nash first order conditions will not have an interior solution.

I was able to mitigate these problems by drawing data from a uniform distribution to bound the data and by choosing variances for product characteristics and random coefficients which were large enough to provide sufficient variation for identification but small enough to avoid overflow. Nevertheless, the weakness of the instruments often produced weighting matrices, $W = \left(Z'Z\right)^{-1}$, with high condition numbers.

3.4.1.2 *Some Observations on Instruments*

The BLP model suffers from price endogeneity.[20] Researchers have pursued several identification strategies, including characteristics of other products [Bresnahan et al., 1997] and cost shifters, both of which I evaluate in this paper.[21] BLP motivate their choice of product characteristics instruments from a functional form argument that price should be exchangeable in the characteristics of other goods. Let each firm f produce some set of products $\mathcal{F}_f$. Then, BLP's candidate instruments are x_{jk}, $\sum\limits_{r \neq j, r \in \mathcal{F}_f} x_{rk}$, and $\sum\limits_{r \neq j, r \notin \mathcal{F}_f} x_{rk}$, which are own-product characteristics, characteristics of other goods the firm produces, and characteristics of rival goods, respectively. Because these instruments are often collinear, they choose a subset [Berry et al., 1995].

While generating synthetic data, I discovered that this collinearity can be quite bad and often leads to GMM weighting matrices with large condition numbers. Furthermore, these instruments are linearly dependent if each firm produces only one product. Consequently, I assume that firms produce at least two products. Even for multi-product firms, collinearity persists. Note that the instruments formed from the

[20]Some argue that product characteristics are also endogenous [Mazzeo, 2002]. In this paper, I treat them as exogenous.

[21]Nevo [2001] uses prices from other markets as instruments which he justifies via Hausman [1997]'s argument that there are common cost shocks but no nation-wide demand shocks.

characteristics of rival goods $\sum_{r \neq j, r \notin \mathcal{F}_f} x_{rk}$ are just $\sum_{r \in J} x_{rk} - \sum_{r \neq j, r \in \mathcal{F}_f} x_{rk} - x_{jk}$ which is a linear combination of the columns of instrument matrix (1, $\sum_{r \neq j, r \in \mathcal{F}_f} x_{rk}$, and x_{jk}). Consequently, I omit either the sums of the characteristics of a firm's other products or the sum of rival products' characteristics to ensure that the instrument matrix has full rank. Armstrong [2011]'s simulations show that increasing the number of products per firm does decrease bias to some extent, perhaps because the instruments are less collinear.

In addition to product characteristic instruments, I also use cost shifters because BLP is asymptotically unidentified for one market without cost shifters [Armstrong, 2011]. To check that the generated data were reasonable, I regressed price on the various sets of instruments. For large J, R^2 was typically less than 0.01 with product characteristics instruments but $R^2 \approx 0.5$ with cost-shifter IV. I tried using higher powers of these instruments as in Dubé et al. [2011] but found that higher moments often increased the condition number and, thus, collinearity of the weighting matrix. Because correlation between many of these higher powers and the simple instruments was greater than 0.98, I decided to work only with the linear terms. As shown in Section 3.5, we need a better instrumentation strategy to estimate BLP reliably. Furthermore, given that the optimal instruments could be calculated in theory if we knew the functional relationship between price and the unobserved product market shock, it may make sense to use a different basis combined with a high dimensional model and a penalization method [Belloni and Chernozhukov, 2011] to choose the best practical set of instruments.

3.4.2 *Solving for Price Equilibrium*

Finding a Bertrand-Nash price equilibrium for BLP-style demand presents both econometric and technical challenges.[22] Armstrong [2011] proves the necessity of generating data from a structural model in order to avoid obscuring asymptotic properties of the data generating process. In addition BLP's instruments require that firms produce multiple products which adds further complexity and non-linearity to the system of FOCs. This non-linearity makes finding a reliable,

[22]If the equilibrium is not Bertrand-Nash, then it is unclear whether or not the model is identified.

interior optimum difficult: if the solver overshoots the solution, the shape of the problem is such that the solver will diverge to either zero or infinite price, both of which are clearly wrong.

I now derive the system of FOCs and discuss a novel transform I developed to correct most of the non-linearities. This solution facilitates quadratic convergence to a point which satisfies the optimality conditions for over 90% of the markets I generated.

3.4.2.1 *Multi-Product Firms*

Because the characteristics of rival goods are perfectly collinear when each firm produces only a single product, I consider multi-product firms where each firm f produces a set of goods $\mathcal{F}_f$. This modeling decision is also more realistic, because firms – especially consumer goods manufacturers – often produce several products which are somewhat substitutable. For multi-product firms, profits in each market t of size M are

$$\begin{aligned} \pi_t^f &= \sum_{k \in \mathcal{F}_f} \pi_{kt} \\ \pi_t^f &= \sum_{k \in \mathcal{F}_f} \left(p_{kt} M s_{kt} - C\left(M s_{kt} \right) \right). \end{aligned}$$

Without loss of generality, I normalize the market size to $M = 1$. Then, the FOCs for each firm are

$$\begin{aligned} \frac{\partial \pi_t^f}{\partial p_{jt}} &= s_{jt} + \sum_{k \in \mathcal{F}_f} \left(p_{kt} - mc_{kt} \right) \frac{\partial s_{kt}}{\partial p_{jt}} \\ \frac{\partial \pi_t^f}{\partial p_{jt}} &= s_{jt} + \sum_{k \in \mathcal{F}_f} \left(p_{kt} - mc_{kt} \right) \frac{\partial s_{kt}}{\partial p_{jt}} \end{aligned}$$

for all $j \in \mathcal{F}_f$. Now the FOCs include the impact of a change in price on inframarginal units and marginal units as before as well as an additional strategic term $\sum\limits_{k \neq j, k \in \mathcal{F}_f} \left(p_{kt} - mc_{kt} \right) M \dfrac{\partial s_{kt}}{\partial p_{jt}}$ for the impact of the price change on the firm's other products.

The Jacobian of this system of FOCs is

$$\frac{\partial^2 \pi_t^f}{\partial p_{kt} \partial p_{jt}} = \frac{\partial}{\partial p_{kt}} \left(s_{jt} + \sum_{i \in \mathcal{F}_f} \left(p_{it} - mc_{it} \right) \frac{\partial s_{it}}{\partial p_{jt}} \right)$$

$$\frac{\partial^2 \pi_t^f}{\partial p_{kt} \partial p_{jt}} = \frac{\partial s_{jt}}{\partial p_{kt}} + \mathbb{I} \left[k \in \mathcal{F}_f \right] \frac{\partial s_{kt}}{\partial p_{jt}} + \sum_{i \in \mathcal{F}_f} \left(p_{it} - mc_{it} \right) \frac{\partial^2 s_{it}}{\partial p_{jt} \partial p_{kt}}$$

where

$$\begin{aligned} \frac{\partial^2 s_{it}}{\partial p_{jt} \partial p_{kt}} &= \frac{1}{\det\left(\Sigma\right)} \int \beta_{price}^2 \left\{ \mathbb{I}\left[i=j\right] \left(\mathbb{I}\left[i=k\right] - s_{kt} \right) s_{jt} \right. \\ &- \left. \left(\mathbb{I}\left[j=k\right] + \mathbb{I}\left[i=k\right] - 2 s_{kt} \right) s_{it} s_{jt} \right\} \phi \left(\Sigma^{-1} \left(\beta - \overline{\beta} \right) \right) d\beta. \end{aligned}$$

3.4.2.2 *Evaluation*

To evaluate the integrals for market shares and their gradients, I use an SGI rule with nodes and weights $\{y_r, w_r\}$. I compute the scaled nodes from the unweighted SGI nodes using the relationship $\Sigma^{-1} \left(\beta_r - \overline{\beta} \right) = y_r$. Then, the share integrals become

$$s_{jt} = \sum_r w_r s_{jt} \left(\beta_r \right)$$

$$\frac{\partial s_{jt}}{\partial p_k} = \begin{cases} \sum_r \beta_r \left(1 - s_{jt}\left(\beta_r\right) \right) s_{jt}\left(\beta_r\right) w_r & \text{if } k = j \\ \sum_r \beta_r \left\{ - s_{kt}\left(\beta_r\right) s_{jt}\left(\beta_r\right) \right\} w_r & \text{if } k \neq j \end{cases}$$

$$\frac{\partial^2 s_{jt}}{\partial p_{jt} \partial p_k} = \begin{cases} \sum_r \beta_r^2 \left(1 - s_{jt}\left(\beta_r\right) \right) s_{jt}\left(\beta_r\right) \left(1 - 2 s_{jt}\left(\beta_r\right) \right) w_r & \text{if } k = j \\ \sum_r \beta_r^2 \left\{ - s_{kt}\left(\beta_r\right) s_{jt}\left(\beta_r\right) \left(1 - 2 s_{jt}\left(\beta_r\right) \right) \right\} w_r & \text{if } k \neq j \end{cases}$$

where the functions are evaluated at all the rescaled nodes $\beta_r = \overline{\beta} + \Sigma y_r$ which are the type-specific tastes for price.

In addition to the above terms, the second derivative due to the multi-product term is

$$\begin{aligned}\frac{\partial^2 s_{it}}{\partial p_{jt}\partial p_{kt}} &= \sum_r \beta_r^2 \left\{ \mathbb{I}\left[i=j\right]\left(\mathbb{I}\left[j=k\right]-s_{kt}\left(\beta_r\right)\right)s_{jt}\left(\beta_r\right) \right.\\ &- \left.\left(\mathbb{I}\left[j=k\right]+\mathbb{I}\left[i=k\right]-2s_{kt}\left(\beta_r\right)\right)s_{it}\left(\beta_r\right)s_{jt}\left(\beta_r\right)\right\} w_r.\end{aligned}$$

I have exploited the fact that the product of $\mathbb{I}[i=j]$ and $\mathbb{I}[i=k]$ implies that $i=j$ if the outer indicator is true so I can replace i inside the expression with j. The first term on the RHS looks a lot like ${}^{\partial s_{jt}}/_{\partial p_{kt}}$ but is not the same because of the β_r^2 term instead of β_r.

3.4.2.3 *Numerical Issues*

In addition to the typical numerical problems encountered in the BLP model (See 3.3.3.), it is also necessary to solve for the price equilibrium which is particularly challenging because the Bertrand-Nash FOCs are non-linear, as plotting slices of the FOCs about marginal cost shows. I use the Path 5.0.00 solver [Dirkse and Ferris, 1995, Ferris et al., 1999, Ferris and Munson, 2000], which is currently the best complementarity solver. Nevertheless, these non-linearities can easily confuse a Newtonian root-finder, such as that in Path. Even with an analytic Jacobian – which is essential for optimal numerical results – Path often did not converge quadratically, a sign of numerical instabilities. That market shares are often small, especially as J increases, makes the equations unstable and often leads to rank deficient systems.[23] Furthermore, small share values produce even smaller derivatives.

I discovered a novel transform – dividing the FOC for product j in market t by the market share s_{jt} – which removes most of the non-linearity and leads to consistent, quadratic convergence.[24] In addition, the solver is more likely to find the same optimum for different starts. To understand the intuition for this transformation, consider the FOCs for simple logit demand:

$$\begin{aligned}foc_{jt} &= s_{jt}+\left(p_{jt}-mc_{jt}\right)\frac{\partial s_{jt}}{\partial p_{jt}}\\ &= s_{jt}+\beta_{price}\left(p_{jt}-mc_{jt}\right)\left(1-s_{jt}\right)s_{jt}.\end{aligned}$$

[23] The solver sets values below a cutoff to 0 which can exacerbate the problem.

[24] In about 10% of cases, this rescaling makes the objective function too flat for the solver to make progress because the Jacobian has many vanishingly small elements. In this case, using the untransformed set of FOCs usually produces a reliable solution.

Clearly, if s_{jt} is small, then foc_{jt} will also be very small, leading to ill-conditioning. Dividing by market shares corrects this problem

$$\widetilde{foc_{jt}} = 1 + \beta_{price} \left(p_{jt} - mc_{jt} \right) \left(1 - s_{jt} \right)$$

so that the FOC is nearly linear in the region of the solution, making the system much more stable. In practice, I found that this transform worked very well for BLP demand and often makes the Jacobian diagonally dominant.

To ensure that I have found the true local optimum, I take multiple starts about a small markup over marginal cost. Because Path is a complementarity solver it sometimes converges to box constraints which prevent price from becoming negative. I select the optimum which produces the smallest residual for $\left\| \widetilde{foc_{jt}} \right\|_2$.

Because the markets are independent, I solve for equilibrium prices for each market individually. This promotes better convergence because different blocks require different scaling and stopping criteria: trying to solve the entire system will be less stable because the solver will use the wrong scaling or stopping condition for some blocks. Solving for individual markets also decreases memory usage and, in some cases, avoids the performance hit when paging is necessary in the virtual memory system of the computer. Nevertheless, solving for the equilibrium prices in each market is still computationally challenging because Newtonian root-finding scales as J^3 for dense systems, making systems with even $J = 100$ products per market quite slow to solve. Furthermore, increasing J increases non-linearities and decreases the share values, often making the Jacobian of the FOCs ill-conditioned. Trying to solve systems with many more products may require even higher precision arithmetic or advances in solver algorithms.

Lastly, generating 'random numbers' on a discrete computer is a treacherous business [Judd, 1998]. To avoid problems from spurious correlation due to poorly generated pseudo-random numbers, I use Mersenne Twister [Matsumoto and Nishimura, 1998], currently the best algorithm, to draw random numbers.[25]

[25] I use Richard Wagner's C++ implementation which can be downloaded from `http://www-personal.umich.edu/~wagnerr/MersenneTwister.html`. This implementation has several nice features, including generating normally distributed random variables correctly instead of using `erfinv()` which approximates the tails of the distribution poorly.

3.4.3 *Sampling Error*

Berry et al. [1995] and Berry et al. [2004b] devote considerable effort to minimizing bias from sampling error and show that it is negligible for their data. Often simulation studies – for example Dubé et al. [2011] – introduce artificial 'simulation error' by generating market shares using a pMC quadrature rule with very few draws. I compute simulated shares using a very accurate sparse grids integration rule in order to focus on model performance instead of distorting results with artificial simulation error. To study simulation error, one could simply replace the sparse grids rule with some suitable sampling process from the desired distribution of consumer types.

3.5 RESULTS

Using the infrastructure described above, I synthesize data and estimate the BLP model for a range of markets and products. The results show that for the configurations of markets and products often encountered in real-world datasets the BLP model suffers from considerable finite sample bias and that asymptotic inference may not yet be valid. Because GMM has finite sample bias, it is not surprising that the traditional BLP GMM-based estimation strategy is also biased. However, the magnitude of the bias and its persistence as the number of markets and/or products increase, even with cost-shifter IV, should concern practitioners.

Casual inspection of log files and point estimates to verify the estimation output confirms these results. For each replication, I estimated 50 unique starts, for which the solver usually found a valid optimum 95% of the time or better.[26] Some point estimates showed considerable variation in magnitude across starts despite valid exit codes, indicating that these parameters cannot be estimated precisely. In particular, the point estimates for the scale parameters and the constant term, θ_{11}, suffered from this problem more than the coefficients on the product characteristics and price. Using cost-shifter instruments instead of characteristics of rival products decreased this variation to some extent as well as producing smaller asymptotic standard errors.

[26] A valid optimum for SNOPT has the Inform code 1 (Success) and satisfies the market share and other constraints to high tolerances.

The following subsections describe the results in more detail for both instrumentation strategies and their effect on point estimates and elasticities. In addition, I discuss solver convergence and computational costs.

3.5.1 *BLP Characteristics Instruments*

I compute the bias, mean absolute deviation, median absolute deviation, and RMSE for the point estimates to measure the performance of the BLP model for a range of markets, $T \in \{1, 10, 25, 50\}$, and products, $J \in \{12, 24, 48, 100\}$. I compute these statistics for both point estimates and elasticities for each of the 100 replications of each combination of T and J. I also compare two different instrumentation strategies: characteristics of rival products and cost shifters. These statistics show that for these datasets, the BLP point estimates with product characteristics IV are consistently biased in one direction for all of these statistics and that inference based on the asymptotic distribution of the estimator is not yet valid.

I compute these statistics by comparing the point estimates with the true parameters for θ_1, the parameters for the mean utility, and θ_2, the scale of the random coefficients' taste shock. For example, bias $= 1/N \sum\limits_{n \in N} \hat{\theta}^{(n)} - \theta^0$, where $\hat{\theta}^{(n)}$ is the best estimate for the n-th replication and θ^0 is the truth. The mean and median absolute deviation are computed similarly for the absolute value of the deviation of the best starts from the truth. The Root Mean Squared Error (RMSE) is $\sqrt{1/N \sum\limits_{n \in N} \left(\hat{\theta}^{(n)} - \theta^0\right)^2}$.

In Table 14, Table 15, and Table 16, I summarize the results for mean utility parameters, θ_1. These tables show that the parameters for the constant term, θ_{11}, and for the product characteristic, θ_{12}, consistently have positive bias whereas the parameter for price, θ_{13} has negative bias. The computed values for both the absolute and relative bias are surprisingly large – recall that $\theta_1^0 = \begin{pmatrix} 3 & 2 & -5 \end{pmatrix}'$. More worryingly, increasing the number of markets and products has little effect on the bias and appears to actually increase it in some cases, though this is probably an artifact of only using 100 replications per experiment. That most of the variation in BLP models is usually with-in market [Berry et al., 2004b, Armstrong, 2011] means that increasing the number of markets will do little to decrease bias. Instead, identification must come from increasing the number of products. Furthermore, increasing the number of replications to $1,000$ will not improve results considerably because bias scales as $\sqrt{N}$. The other statistics – mean absolute deviation, median absolute deviation, and RMSE – tell the same story: for the numbers of markets and products considered there is persistent bias. The asymptotic theory for BLP developed in Berry et al. [2004b] argues that the BLP estimator is asymptotically normal as $J \to \infty$ (with one market) for certain (unspecified) price setting mechanisms. But, Armstrong [2011] proves that BLP is in fact only identified as $J \to \infty$ with cost shifter instruments for a Bertrand-Nash price equilibrium. I.e., using the characteristics of other firms' products will not identify the model when $J \to \infty$. For the values of J in these experiments, finite sample bias is present which is perhaps not surprising, given that the BLP estimator is a GMM estimator and GMM is know to have poor finite sample properties. Consequently, applied researchers should search for alternative identification strategies.[27]

Table 17, Table 18, and Table 19 present these statistics for standard deviation of the random coefficients. θ_{22}, the parameter on the product characteristics has little bias, but the other parameters have much larger bias. When checking the results for the point estimates, I often observed estimates for θ_{21} and θ_{23} which were either extremely large or small. This could be caused by the relative flatness of the BLP objective function along these dimensions: i.e., these parameters may be identified in theory, but in practice they cannot be estimated precisely. In fact, Armstrong [2011] fixes the scale parameters, θ_2, in his simulations to facilitate the solver's convergence to an optimum.

Because elasticities are crucial for characterizing the properties of demand systems in applied work, I also compute these bias statistics for own- and cross-price elasticities. Table 20 and Table 21 summarize these results for point estimates with rival product characteristics as instruments. Elasticities also suffer from large finite sample bias; in particular, RMSE can be the same order of magnitude as actual elasticities. When looking at these figures, note that the mean own-price elasticities range from -4 to -5, regardless of the number of products so the relative bias is quite large. Cross-price elasticities are also biased. Although the bias may appear to decrease as J increases, this decrease really occurs because of the increase in the number of

[27]With real data, misspecification further complicates identification because the normal mixing distribution for the random coefficients is usually misspecified [Keane and Wasi, 2009].

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	1.3	3.3	1.9	5.3
1	24	1.1	2.3	1.3	3.7
1	48	2.6	3.9	1.8	6.4
1	100	2.3	3.2	1.8	4.8
10	12	0.68	3.1	1.7	6.3
10	24	1	2.7	1.5	5.2
10	48	3.3	4.3	2.1	6.6
10	100	2.3	3.5	1.4	7.2
25	12	−0.27	2.8	2.1	4.2
25	24	1.5	3.2	1.8	5.1
25	48	3.4	4.9	1.9	11
25	100	2.8	4.1	1.8	7.8
50	12	−0.2	2.3	1.7	3.5
50	24	1.7	3.4	1.6	6.9
50	48	2.3	3.6	2.1	7.6
50	100	2.8	4.1	2	6.7

Table 14: Bias, Deviations, and RMSE for θ_{11} with only product characteristics instruments.

products. The magnitude of the average cross-price elasticity decreases from 0.2 to 0.03 as J increases from 12 to 100 products. Consequently, cross-price elasticities also have large relative bias. In addition, the bias of the elasticities is always in the same direction: own-price estimates are always too negative and cross-price are always too positive.

For the numbers of markets and products which are typical for many applied projects, characteristics of rival products produce biased estimates. Furthermore, asymptotics have not begun to produce consistent estimates so the asymptotic standard errors are probably unreliable. Although desirable, increasing the number of products in these experiments becomes computational difficult – perhaps even impossible – because the FOCs for the Bertrand-Nash price equilibrium are a dense Jacobian and the root-finding algorithm scales as J^3. Thus, $J = 100$ produces a matrix with $100^2 = 10,000$ entries which is difficult for most current root-finding algorithms.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	0.33	0.74	0.47	1.4
1	24	−0.033	0.49	0.37	0.67
1	48	−0.014	0.63	0.42	1
1	100	0.14	0.56	0.34	0.94
10	12	0.15	0.71	0.4	1.3
10	24	0.19	0.72	0.33	1.4
10	48	0.022	0.62	0.44	0.9
10	100	0.22	0.75	0.46	1.5
25	12	0.018	0.64	0.44	1.1
25	24	0.28	0.85	0.51	1.4
25	48	0.44	1.1	0.5	2.6
25	100	0.26	0.77	0.46	1.2
50	12	0.08	0.63	0.46	0.92
50	24	0.3	0.86	0.45	1.6
50	48	0.46	0.91	0.51	2.2
50	100	0.21	0.8	0.48	1.3

Table 15: Bias, Deviations, and RMSE for θ_{12} with only product characteristics instruments.

3.5.2 *Cost Shifter Instruments*

The results for point estimates and elasticities when the model is estimated with supply-side cost shifters weakly confirm Armstrong [2011]'s proof that the BLP model (and other discrete choice demand systems) is identified with cost-shifter IV but not product characteristics. Estimates for mean utility parameters (See Table 22, Table 23, and Table 24.) show that θ_{12} and θ_{13} are computed with a factor of 10 less bias than with BLP IV; other metrics also improve. Similarly, estimates for the scale parameters for the preference shocks (See Table 25, Table 26, and Table 27.) show that θ_{22} and θ_{23} have much less bias than before. Surprisingly, the constant parameter, θ_{11}, and its corresponding scale parameter, θ_{21}, are still estimated with considerable bias. Note that most parameters have consistent bias in one direction such as θ_{13}, the parameter on price, which almost always has negative bias.

As with characteristics IV, increasing the number of markets or products does not clearly reduce bias. Even for 50 markets and 100 products, asymptotics do not yet seem to apply: consequently, the

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	−2	3	1.3	5.7
1	24	−0.72	1.9	1.2	3.2
1	48	−0.52	1.9	1.2	3
1	100	−0.57	1.7	1.3	2.3
10	12	−1.7	2.6	1.1	6
10	24	−0.65	2	1.3	3.6
10	48	−0.64	1.9	1.3	3.2
10	100	−0.83	2	1	3.9
25	12	−0.62	1.9	1.2	3.1
25	24	−0.96	2.3	1.4	3.7
25	48	−1.3	2.8	1.2	7.6
25	100	−0.95	2.1	1.1	3.7
50	12	−0.39	1.6	1.1	2.7
50	24	−1.2	2.5	1.1	5.4
50	48	−1.2	2.2	1.3	5.2
50	100	−0.63	1.9	1.3	3

Table 16: Bias, Deviations, and RMSE for θ_{13} with only product characteristics instruments.

GMM formula for standard errors, which most practitioners use, is probably invalid and applied researchers should take care when performing inference. To help avoid finite sample bias, Conlon [2009] estimates a dynamic BLP model using Empirical Likelihood at the cost of extra computational complexity.[28] The small sample size of these experiments may obscure the effectiveness of increasing T or J to reduce bias. However, because bias falls as $\sqrt{N}$, even a 10x increase will not reduce bias significantly. Numerical errors in computing the true price equilibrium may also exacerbate these problems because Path often had difficulty finding an interior optimum when synthesizing the data, despite taking up to 30 starts and using a transform to facilitate convergence (See Section 3.4.2.3.).

Supply-side cost-shifters produce slightly better own- and cross-price elasticities than traditional BLP IV, as shown in Table 28 and Table 29. Bias and other metrics decrease but by less than an order

[28] Conlon [2009] solves the primal problem. Developing a dual problem formulation may make EL more accessible for BLP practitioners.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	3.1	3.9	1	7.3
1	24	4.8	5.3	1	10
1	48	5.7	6.5	1	23
1	100	2.1	2.7	1	5.2
10	12	3.5	4.1	1	8.1
10	24	2.9	3.3	0.94	7.1
10	48	4.7	5.1	1	9.9
10	100	1.7	2.2	0.6	6.7
25	12	3.6	4.1	1.7	7
25	24	3.3	3.6	1.1	7.2
25	48	2.9	3.3	0.98	7.4
25	100	2.2	2.7	0.76	6.7
50	12	2.5	3	1	5.6
50	24	4.1	4.5	1.1	11
50	48	1.5	2	0.98	3.6
50	100	2.7	3.1	0.85	7.4

Table 17: Bias, Deviations, and RMSE for θ_{21} with only product characteristics instruments.

of magnitude. Again, the number of markets and products has little effect on bias.

After examining several measures of bias for both point estimates and elasticities using two popular instrumentation strategies, the BLP model appears to be biased for the number of markets and products often encountered in applied research. Practitioners should be concerned that finite sample bias persists in all cases and that asymptotics appear to require many more products and/or markets to reduce bias. Better instruments and estimation strategies are needed to address these issues.

3.5.3 *Solver Convergence*

In addition to the results on bias, I find that BLP becomes increasingly difficult to solve as the number of markets and products increase or the quality of the instruments decreases. In Table 30, I tabulate solver exit codes for all replications and starts by both number of markets and products when the instruments are the characteristics

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	0.15	0.73	0.62	1.1
1	24	−0.12	0.47	0.38	0.58
1	48	−0.14	0.5	0.44	0.63
1	100	−0.029	0.4	0.24	0.62
10	12	−0.1	0.4	0.26	0.53
10	24	0.00034	0.3	0.16	0.47
10	48	−0.15	0.28	0.17	0.42
10	100	−0.049	0.27	0.16	0.48
25	12	−0.097	0.35	0.26	0.51
25	24	−0.0093	0.32	0.23	0.48
25	48	0.045	0.33	0.17	0.56
25	100	−0.0041	0.21	0.12	0.36
50	12	−0.01	0.33	0.23	0.45
50	24	0.011	0.28	0.16	0.48
50	48	0.11	0.27	0.12	0.7
50	100	−0.0052	0.25	0.13	0.43

Table 18: Bias, Deviations, and RMSE for θ_{22} with only product characteristics instruments.

of rival goods. In this table, each row summarizes the exit codes for some number of markets, T, and number of products, J. The columns, except for the last column, are labeled by the SNOPT exit code to make it easier to view the results. These exit codes [Gill et al., 2002] are explained in Table 31. The number of successful exit codes, Inform=1, decreases as both the number of markets and products increase. Furthermore, exit code 43 – which means that the initial constraints cannot be satisfied – is the most common source of failure and increases in frequency almost linearly with the number of markets. This problem could be caused by the non-convergence of Berry [1994]'s contraction mapping if it is less likely to converge for larger numbers of products (See Section 3.3.3.1.). These results show that the BLP model becomes harder to solve as its complexity grows, necessitating more starts to find the global optimum. Table 32 shows similar results, though smaller in magnitude, for cost-shifter instruments. Better instruments appear to facilitate solver convergence, as reflected in fewer unsuccessful exit codes.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	0.72	1.6	1	2.6
1	24	0.28	1.2	1	2.1
1	48	0.093	1.1	1	1.3
1	100	−0.047	0.88	1	1
10	12	0.58	1.3	1	2.3
10	24	0.026	0.79	0.89	0.99
10	48	0.16	0.79	0.67	1.1
10	100	0.12	0.66	0.47	0.99
25	12	0.035	0.9	1	1.1
25	24	0.1	0.77	0.67	1
25	48	0.22	0.73	0.41	1.8
25	100	0.21	0.61	0.37	1
50	12	−0.11	0.74	0.72	0.99
50	24	0.25	0.73	0.41	1.4
50	48	0.15	0.52	0.3	0.99
50	100	0.081	0.48	0.37	0.63

Table 19: Bias, Deviations, and RMSE for θ_{23} with only product characteristics instruments.

The last column in Table 30 is the total number of starting values observed for each combination of markets and products. It usually equals 5,000 (the number of replications, 100, times the number of starts, 50) except for cases where the solver failed to converge before the limit on the job's wall-clock time was met.[29] The distribution of wall-clock time required for different starts is quite skewed with a long, thin upper tail: i.e., most jobs complete very quickly whereas a few take an extremely long time. See Figure 5 for a typical example of the distribution of runtimes, in this case for 50 markets and 100 products with the traditional BLP instruments. The 'mass point' at 24 hours was caused by setting an upper limit of 24 hours in the PBS scripts which ran these estimation jobs.

[29] Wall-clock time is the elapsed time a job consumes while running, more precisely it is the sum of CPU time, I/O time, and the communication channel delay.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	−0.77	2.2	0.94	4.9
1	24	−0.095	1.5	0.77	3.3
1	48	−0.082	1.6	0.91	2.7
1	100	−0.39	1.5	0.98	2.5
10	12	−0.5	1.7	0.81	3.3
10	24	−0.57	1.7	0.83	3.3
10	48	−0.16	1.5	0.97	2.2
10	100	−0.53	1.7	0.93	3.3
25	12	−0.3	1.4	0.94	2.7
25	24	−0.72	1.8	1.1	3
25	48	−0.87	2.2	1.1	4.9
25	100	−0.61	1.7	0.97	2.7
50	12	−0.43	1.5	0.94	2.6
50	24	−0.77	1.9	0.91	3.8
50	48	−0.97	1.9	1.1	4
50	100	−0.56	1.8	1.1	2.9

Table 20: Bias, Deviations, and RMSE for own-price elasticities with only product characteristics instruments.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	0.094	0.15	0.035	0.57
1	24	0.017	0.051	0.013	0.2
1	48	0.0068	0.027	0.0072	0.08
1	100	0.0038	0.011	0.0029	0.034
10	12	0.075	0.12	0.051	0.27
10	24	0.032	0.059	0.022	0.15
10	48	0.0097	0.023	0.011	0.046
10	100	0.0056	0.013	0.0053	0.037
25	12	0.067	0.099	0.055	0.23
25	24	0.039	0.063	0.03	0.13
25	48	0.02	0.036	0.013	0.13
25	100	0.0065	0.013	0.0062	0.039
50	12	0.059	0.099	0.055	0.23
50	24	0.044	0.071	0.024	0.27
50	48	0.019	0.031	0.013	0.1
50	100	0.0054	0.013	0.0067	0.028

Table 21: Bias, Deviations, and RMSE for cross-price elasticities with only product characteristics instruments.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	−1.4	2.4	0.98	4.9
1	24	0.8	1.8	1	3.3
1	48	4.1	4.8	1.5	7.8
1	100	4.5	5.1	1.8	8.9
10	12	0.27	1.3	0.95	1.7
10	24	0.38	0.95	0.67	1.4
10	48	1.9	2.2	0.81	4.2
10	100	4.4	4.8	1.3	7.7
25	12	0.0036	1	0.72	1.5
25	24	0.65	1	0.74	1.4
25	48	2.2	2.4	0.91	3.9
25	100	4.6	4.9	1.4	7.9
50	12	−0.16	0.69	0.58	0.93
50	24	0.67	0.91	0.42	2
50	48	1.9	2.1	0.79	3.6
50	100	4.3	4.5	1.3	8.7

Table 22: Bias, Deviations, and RMSE for θ_{11} with cost-shifter instruments.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	0.043	0.37	0.22	0.54
1	24	−0.092	0.34	0.24	0.47
1	48	−0.098	0.28	0.24	0.36
1	100	−0.027	0.19	0.15	0.24
10	12	0.02	0.27	0.19	0.38
10	24	0.015	0.19	0.14	0.25
10	48	0.0086	0.15	0.11	0.19
10	100	−0.0065	0.071	0.053	0.089
25	12	0.061	0.22	0.17	0.29
25	24	0.049	0.15	0.11	0.22
25	48	−0.0054	0.1	0.068	0.15
25	100	−0.0059	0.052	0.041	0.071
50	12	0.065	0.16	0.1	0.25
50	24	0.1	0.16	0.078	0.45
50	48	0.02	0.068	0.051	0.1
50	100	−0.0028	0.037	0.025	0.057

Table 23: Bias, Deviations, and RMSE for θ_{12} with cost-shifter instruments.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	−0.38	1.1	0.83	1.5
1	24	−0.05	1	0.84	1.3
1	48	0.012	0.99	0.86	1.2
1	100	0.057	0.72	0.62	0.88
10	12	−0.62	1.3	0.92	2
10	24	−0.18	0.8	0.61	1.3
10	48	−0.15	0.62	0.52	0.86
10	100	−0.027	0.39	0.3	0.52
25	12	−0.38	1	0.71	1.6
25	24	−0.3	0.73	0.64	0.98
25	48	−0.11	0.45	0.34	0.63
25	100	−0.033	0.25	0.19	0.33
50	12	−0.081	0.79	0.67	1.1
50	24	−0.22	0.55	0.35	1
50	48	−0.026	0.28	0.18	0.4
50	100	0.003	0.19	0.14	0.26

Table 24: Bias, Deviations, and RMSE for θ_{13} with cost-shifter instruments.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	7.4	8.2	2.6	13
1	24	8.4	8.8	2.9	14
1	48	7.2	8.1	1.1	13
1	100	6.2	7.1	3.2	12
10	12	0.8	1.8	1	2.7
10	24	4	4.9	1	11
10	48	2.9	3.8	1	6.6
10	100	5.9	6.8	1.2	11
25	12	1.5	2.3	1	3.4
25	24	3.6	4.4	1.9	7.7
25	48	3.7	4.6	2.1	7
25	100	6.2	7	1.8	11
50	12	0.97	2	1	3.1
50	24	3.9	4.6	1	12
50	48	3.6	4.2	1.8	6.3
50	100	5.9	6.6	2.2	12

Table 25: Bias, Deviations, and RMSE for θ_{21} with cost-shifter instruments.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	0.15	0.52	0.4	0.66
1	24	−0.042	0.39	0.31	0.5
1	48	−0.058	0.33	0.26	0.43
1	100	0.032	0.22	0.18	0.3
10	12	−0.038	0.3	0.25	0.39
10	24	0.0067	0.2	0.17	0.27
10	48	−0.023	0.16	0.12	0.22
10	100	−0.0037	0.096	0.062	0.13
25	12	0.0026	0.22	0.17	0.3
25	24	0.00017	0.17	0.12	0.25
25	48	−0.0048	0.1	0.072	0.15
25	100	0.00058	0.06	0.044	0.077
50	12	0.035	0.18	0.13	0.25
50	24	0.065	0.13	0.082	0.25
50	48	0.028	0.063	0.044	0.083
50	100	0.01	0.043	0.033	0.057

Table 26: Bias, Deviations, and RMSE for θ_{22} with cost-shifter instruments.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	0.23	0.8	0.74	1
1	24	0.0036	0.76	0.83	0.98
1	48	−0.1	0.72	0.78	0.86
1	100	−0.18	0.58	0.54	0.69
10	12	0.17	0.86	0.87	1.2
10	24	−0.06	0.62	0.51	0.89
10	48	−0.022	0.46	0.39	0.6
10	100	−0.035	0.31	0.21	0.43
25	12	−0.035	0.7	0.65	0.88
25	24	0.034	0.45	0.34	0.61
25	48	0.028	0.27	0.17	0.42
25	100	0.0083	0.18	0.13	0.27
50	12	−0.18	0.56	0.43	0.72
50	24	−0.031	0.29	0.2	0.41
50	48	−0.026	0.16	0.099	0.24
50	100	−0.01	0.12	0.077	0.18

Table 27: Bias, Deviations, and RMSE for θ_{23} with cost-shifter instruments.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	0.059	0.86	0.52	1.4
1	24	0.17	0.83	0.55	1.3
1	48	0.11	0.85	0.6	1.3
1	100	−0.59	1.3	0.43	60
10	12	−0.098	0.69	0.48	1
10	24	−0.095	0.52	0.33	0.82
10	48	−0.15	0.48	0.28	4.2
10	100	−0.072	0.3	0.19	0.54
25	12	−0.23	0.56	0.38	0.83
25	24	−0.22	0.48	0.34	0.69
25	48	−0.062	0.3	0.19	0.45
25	100	−0.16	0.3	0.13	0.68
50	12	−0.27	0.54	0.32	0.92
50	24	−0.32	0.46	0.22	1
50	48	−0.1	0.2	0.12	0.33
50	100	−0.15	0.24	0.098	0.57

Table 28: Bias, Deviations, and RMSE for own-price elasticities with cost-shifter instruments.

T	J	Bias	Mean Abs Dev	Med Abs Dev	RMSE
1	12	0.047	0.082	0.031	0.16
1	24	0.016	0.034	0.012	0.075
1	48	0.003	0.015	0.0049	0.037
1	100	0.0017	0.0064	0.0016	0.24
10	12	0.03	0.062	0.036	0.1
10	24	0.012	0.026	0.015	0.044
10	48	0.0057	0.01	0.0053	0.079
10	100	0.0017	0.0032	0.0018	0.0057
25	12	0.041	0.066	0.04	0.1
25	24	0.021	0.03	0.019	0.045
25	48	0.0062	0.0087	0.005	0.014
25	100	0.0023	0.0032	0.0018	0.0062
50	12	0.028	0.057	0.033	0.093
50	24	0.022	0.028	0.014	0.057
50	48	0.0065	0.0085	0.0053	0.013
50	100	0.0022	0.0028	0.0017	0.0052

Table 29: Bias, Deviations, and RMSE for cross-price elasticities with cost-shifter instruments.

#Market.#Product	SNOPT Inform Codes								Total
	1	3	13	31	32	42	43	84	
T1.J12	4987	0	1	0	0	0	12	0	5000
T1.J24	4987	0	1	0	0	0	12	0	5000
T1.J48	4996	0	1	0	0	1	2	0	5000
T1.J100	4962	4	6	6	0	0	22	0	5000
T10.J12	4947	0	1	2	0	0	50	0	5000
T10.J24	4950	0	2	1	0	1	46	0	5000
T10.J48	4920	0	6	8	0	1	65	0	5000
T10.J100	4840	0	11	30	0	2	117	0	5000
T25.J12	4941	0	0	4	0	0	55	0	5000
T25.J24	4914	0	9	1	0	2	74	0	5000
T25.J48	4824	0	17	21	1	5	132	0	5000
T25.J100	4613	0	15	21	32	3	288	0	4972
T50.J12	4932	0	1	4	0	0	63	0	5000
T50.J24	4808	0	12	19	4	6	151	0	5000
T50.J48	4462	0	21	33	14	9	285	176	5000
T50.J100	4225	5	24	5	0	9	512	21	4801

Table 30: Count of Solver Exit Codes by Number of Markets and Products for BLP Characteristics Instruments

SNOPT exit code	description
1	optimality conditions satisfied
3	desired tolerance could not be achieved
13	nonlinear infeasibilities minimized
31	iteration limit reached
32	major iteration limit reached
42	singular basis
43	cannot satisfy the general constraints
84	not enough real storage

Table 31: SNOPT Exit Codes

#Market.#Product	SNOPT Inform Codes								Total
	1	3	13	31	32	42	43	84	
T1.J12	4983	0	6	0	0	0	11	0	5000
T1.J24	4987	0	0	0	0	0	12	0	5000
T1.J48	4993	0	4	0	0	0	3	0	5000
T1.J100	4969	5	12	1	0	0	13	0	5000
T10.J12	4976	0	0	1	0	0	23	0	5000
T10.J24	4964	0	1	3	0	0	32	0	5000
T10.J48	4984	0	0	2	0	1	13	0	5000
T10.J100	4970	0	1	3	0	0	26	0	5000
T25.J12	4973	0	1	1	0	0	25	0	5000
T25.J24	4921	0	3	3	0	1	72	0	5000
T25.J48	4911	0	7	10	0	3	69	0	5000
T25.J100	4852	0	9	8	5	3	107	0	4984
T50.J12	4936	0	1	3	0	0	60	0	5000
T50.J24	4881	0	10	7	6	3	93	0	5000
T50.J48	4776	0	8	14	8	2	110	82	5000
T50.J100	424	1	0	0	0	0	16	3	444

Table 32: Count of Solver Exit Codes by Number of Markets and Products for Cost-Shifter Instruments

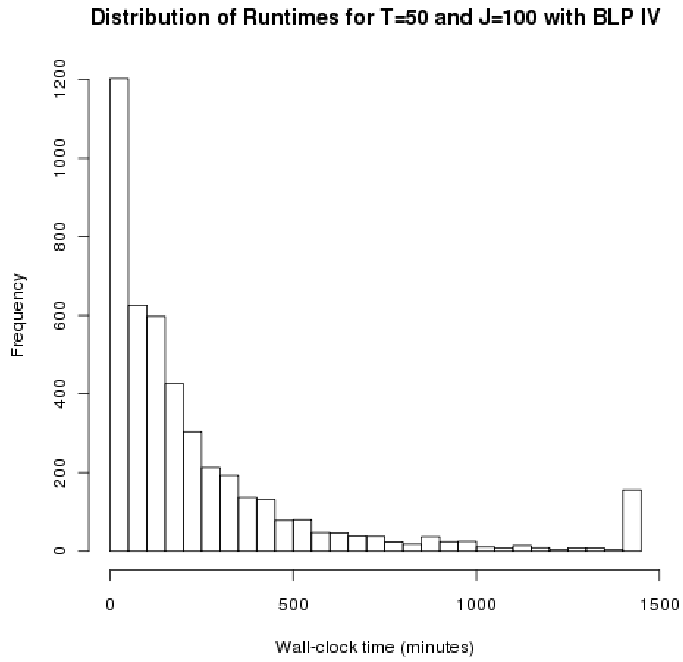

Figure 5: Distribution of runtimes for $T = 50$ markets and $J = 100$ products with characteristics of rival products as instruments.

Inspection of the various optima found by the solver for the different starting guesses confirms the difficulties in solving the BLP model. I found considerable variation in the point estimates – even those with the smallest residual for each replication. In particular, the standard error for the random coefficients seems to be difficult to estimate. The parameters for the constant term, θ_{11}, and its scale, θ_{21}, are usually the most difficult parameter to compute precisely.

3.5.4 *Bias from Quadrature Rules*

As discussed in Chapter 2, accurate numerical integration can have a profound effect on point estimates, elasticities, and other economic estimates. Most researchers adopt some form of a Monte Carlo rule – sometimes improved by using importance sampling or quasi-Monte Carlo methods. Yet, many practitioners, basing their intuition on the Law of Large Numbers, assume that as the number of draws approaches infinity the bias will disappear and estimates will be consistent asymptotically normal (CAN). However, the finite sample performance of these rules is unknown. Furthermore, if the optimizer finds a local optimum instead of the global optimum – which is particularly common with pMC rules – then the properties of the estimator are unknown. As a first step towards answering this question, I use the Monte Carlo infrastructure of this chapter to compute the bias for pMC and SGI quadrature rules and find that the SGI rule considerably outperforms the pMC rule for all metrics.

Table 33 summarizes the bias statistics for a simple synthetic dataset with 2 markets and 24 products. The data was generated using the DGP in Section 3.4 with an SGI rule. The estimation procedure was the same as that described in Section 3.3.3 except that I compute the pMC estimates with 165 nodes to insure that the computational cost is roughly the same as the SGI rule, which also uses 165 nodes.[30] The key difference between the rules is that the SGI rule is exact for all degree 11 monomials or less whereas the pMC rule is much less accurate. These results demonstrate the superiority of the SGI rule: for almost every measure of bias, SGI parameter estimates are closer to the truth. For the few cases where pMC is better, it is only slightly less biased. Consequently, polynomial-based rules appear to decrease the bias of estimates vis-a-vis pMC.

	Bias		Mean Abs Dev		Med Abs Dev		RMSE	
	SGI	pMC	SGI	pMC	SGI	pMC	SGI	pMC
θ_{11}	0.96	12.34	2.29	13.25	1.20	3.64	4.00	28.92
θ_{12}	0.02	−0.13	0.52	0.38	0.22	0.33	0.94	0.48
θ_{13}	−0.28	−0.38	1.47	1.21	0.62	0.99	3.01	1.51
θ_{21}	22.57	128.22	23.01	128.24	2.62	34.06	81.76	253.87
θ_{22}	0.02	−0.04	0.12	0.16	0.07	0.13	0.19	0.20
θ_{23}	0.08	0.64	0.36	0.75	0.16	0.79	0.75	0.90

Table 33: Comparison of bias in point estimates : SGI vs. pMC for T=2 markets and J=24 products with 165 nodes.

3.5.5 *Computational Cost*

These simulations and data analysis required 94,325 CPU-hours and 39,875 jobs to estimate the BLP model over 320,000 times.[31] These numbers may appear large compared to what is common for Eco-

[30] Currently, I only use one set of pMC draws, but this should not affect results significantly because I use 100 replications.

[31] 16 experiments × 100 replications × 50 starts × 2 types of instrumental variables × 2 for restarts on original optima = 320,000.

nomic research, but this usage level is negligible compared to what is now routine for the natural sciences and engineering. For example, some Physics and Chemistry models now require over 10,000,000 hours to complete and produce petabytes of output. Furthermore HTC computing resources are now available at most universities or through national organizations such as XSEDE (www.xsede.org) and the Open Science Grid (www.opensciencegrid.org) in the US or the National Grid Service (www.ngs.ac.uk) in the UK. In addition, 100,000 hours is the size of a typical exploratory grant at XSEDE, which is available with minimal paperwork.

3.6 BOOTSTRAPPING BLP

Traditionally, BLP practitioners have relied on the GMM formula to compute asymptotic standard errors and perform inference. To evaluate the performance of standard errors researchers typically use the bootstrap. Given the computational challenges of just estimating the BLP model reliably, no one has conducted a bootstrap study of BLP. However, using the infrastructure in this chapter, bootstrapping is now possible and an important subject for a future study to determine what market sizes and numbers of products are necessary to use the asymptotic standard errors.[32]

Because the unit of observation in BLP is effectively the market, bootstrapping is not just a straight-forward matter of resampling: correlation between products in a market and dual asymptotics in the number of markets and products invalidate simple resampling. Correlation across markets can add further complexity, although often researchers assume that different markets are independent. However, BLP datasets often have the same (observed) characteristics in each market.[33] Because market shares in a market are correlated and most variation is typically with-in market, asymptotics depend on $J \to \infty$. Consequently, each market functions a lot like one observation. Perhaps resampling should be done at the market level, which is similar to block resampling for time series processes. Using the parametric bootstrap would require computing ξ_{jt} which is highly inaccurate:

[32]Monte Carlo experiments where demand is a simple logit with an unobserved product market shock could also help establish bounds because it should be easier to estimate than BLP, making it possible to study datasets with much larger T and J.

[33]In the case where markets are different time periods as in the original BLP paper, dynamic considerations could matter.

even with MPEC I observe considerable variation in $\widehat{\xi_{jt}}$ at optima with essentially identical structural parameter estimates $\widehat{\theta}$. Further research is needed to determine the correct resampling procedure to bootstrap BLP.

3.7 CONCLUSION

Advances in parallel computing architecture, such as High Throughput Clusters and parameter sweeps, significantly ease the costs of performing large-scale Monte Carlo experiments. In fact, Monte Carlo experiments are quickly becoming a new pillar of the physical sciences on par with theory and experiment [U.S. National Science Foundation, 2007]. Using this technology, I developed the infrastructure to characterize the performance of the BLP estimator by running Monte Carlo experiments for different numbers of markets and products on a much larger scale than anyone has attempted before.

I find that for the configuration of markets and products found in many real datasets the BLP parameter estimates and elasticities are considerably biased by a variety of measures and there were not enough markets and/or products to reduce bias to the necessary levels. Using cost shifter instruments mitigates this to some extent, weakly confirming the asymptotic theory developed in Armstrong [2011] which showed that asymptotic identification requires supply-side cost-shifters because product characteristics drop out of the pricing equation as $J \to \infty$. However even with 50 markets and 100 products, results show finite sample bias. Consequently, asymptotic results are probably not valid and applied researchers need to develop better instruments, estimation strategies, and models. Because the BLP model is currently the dominant approach for estimating demand for differentiated products, practitioners must find new ways to reduce this bias – perhaps by using much larger datasets than we had previously thought were necessary. By running experiments for datasets generated from different parameters, the method developed in this chapter can be used to improve our understanding of the factors causing this bias and how it propagates through the BLP model.

Given that BLP has considerable finite sample bias and that pMC rules produce artificially tight standard errors (See Chapter 2.), applied researchers should avoid naive implementations of the BLP model, especially in regulatory settings: clearly, any welfare calculations based on such code will produce biased point estimates, misleading standard errors, and flawed welfare calculations.

This computational approach demonstrates that Economists can use Monte Carlo experiments on a much larger and more rigorous scale than before to study the properties of estimators. Using this methodology, we can now study estimators where no theory yet exists. Furthermore, these results can help guide theory in the right directions, just as theory has often pointed applied work in the correct direction in the past.

Part II

EVALUATION OF 2003 MORRISONS-SAFEWAY MERGER

The UK supermarket industry has been the focus of several anti-trust investigations by the UK Competition Commission for a variety of potential abuses of market power with respect to both consumers and suppliers. The final part of the thesis evaluates how the 2003 acquisition of Safeway by Morrisons affected firm profits, prices, and consumer welfare. I construct a structural model for the geographic distribution of demand, by combining a discrete/continuous model of consumer demand with census data. By modeling both store choice and expenditure in conjunction with the geographic locations of stores and consumers, I account for the impact of the spatial distribution of stores and households on both firms' behavior and consumer welfare. I compute the welfare consequences of the merger as well as a counter-factual merger between Safeway and Tesco. This Part improves on earlier studies, such as Smith [2004], by using a higher quality panel of data on consumer purchases (The TNS Worldpanel) and by combining a structural model of demand with disaggregate census data to study how store locations affect market power.

4

THE GEOGRAPHY OF GROCERY DEMAND IN THE UK: AN EVALUATION OF THE 2003 MORRISONS-SAFEWAY MERGER[1]

In 2003, the UK Competition Commission (CC) approved the acquisition of Safeway plc by Wm. Morrisons plc, respectively the fourth and sixth largest firms in the industry. Because Morrisons focused on the North and Safeway on the South, this merger had the potential to create a fourth national champion to rival Asda, Sainsbury's, and Tesco, hopefully improving competition, lowering prices, and improving quality for consumers. But, the merger could also have had an adverse affect on competition by creating pockets of local market power which the merged firm could exploit. To evaluate the CC's decision, I construct a geographic distribution of demand which models the local interactions between consumer demographics and store locations. My model has several parts. I estimate a Discrete/Continuous structural model of demand from a high quality panel of consumer micro-data (the TNS Worldpanel) to explain both store choice and conditional demand for groceries. After combining this demand system with disaggregate census data, I recover marginal costs and then predict store-level sales and profits as well as willingness-to-pay. I use these tools to evaluate the welfare implications of the merger and of a counter-factual merger between Safeway and Tesco. I find that the changes in prices, profits, and consumer welfare under either merger are quite small – although larger for a Tesco-Safeway merger. Although consumers are slightly worse off under these mergers, the results support the UK Competition Commission's approval of the merger.

[1] I gratefully acknowledge the financial support of the UK Economic and Social Research Council through a grant (RES-589-28-0001) to the ESRC Centre for Microdata Methods and Practice (Cemmap). The Computation Institute at the University of Chicago graciously hosted me during 2010-2011 academic year. Conversations with Walter Beckert, Andrew Chesher, Allan Collard-Wexler, Rachel Griffith, Jerry Hausman, Günter Hitsch, Kenneth L. Judd, Andrew Leicester, Lars Nesheim, C. Yeşim Orhun, and Adam Rosen have greatly improved this paper. Rachel Griffith generously provided access to the TNS and IGD data used in this paper.

On 9 January 2003, William Morrisons plc, a UK supermarket chain, made an unsolicited offer for its larger rival Safeway plc, prompting a flurry of takeover proposals from other firms seeking to purchase Safeway. Safeway's suitors included Asda, Sainsbury's, and Tesco as well as Kohlberg Kravis & Roberts (KKR), a private equity firm, and Trackdean Investments Ltd., a family investment vehicle. As the takeover battle escalated, the UK Secretary of State for Trade and Industry referred the acquisition proposals of the four supermarket firms to the UK Competition Commission under the 1973 Fair Trading Act: the Secretary believed that acquisition by Asda, Sainsbury's, or Tesco would cause a substantial lessening of competition and that a Morrisons-Safeway merger might also be anti-competitive. Ultimately the CC approved the merger with Morrisons subject to the divestment of 52 stores where the CC thought the merger would have an adverse impact on local competition. After the CC's decision, on 15 December 2003, Morrisons made a revised offer to acquire Safeway which received the approval of both firms' shareholders on 11 February 2004.

In theory, the merger appeared to make good business sense because Morrisons was concentrated in the North, Yorkshire, and Midlands whereas Safeway was strong in London, the South East, and Scotland. Furthermore, Safeway, although not a failed company, lacked a consistent pricing strategy and had struggled to differentiate itself vis-a-vis Asda, Sainsbury's, and Tesco whereas Morrisons had profitably differentiated itself and grown in a financially sound manner [Seth and Randall, 2001]. Consequently, a combined firm under Morrisons' management could become a strong national competitor through increased scale, greater purchasing power, and better coverage of regional markets. The concomitant increase in competition in the groceries industry should have improved consumer welfare through lower prices and/or higher quality. In practice, the firm struggled for the next several years to integrate Safeway's operations with its own, despite divesting 72 additional stores as well as 114 Safeway Compact convenience stores to further focus its positioning on larger format stores.

In addition to the 2003 merger investigation [UK Competition Commission, 2003], the UK supermarket industry has been the subject of several other investigations by the Competition Commission. In 2000, prior to the merger inquiry, the CC investigated the conduct of multiples towards suppliers [UK Competition Commission, 2000][2] as well as their pricing practices, leading to the adoption of a Supermarket Code of Practice (SCOP) between multiples and suppliers. In addition, most supermarkets voluntarily adopted a national pricing policy to allay concerns about 'price flexing', the exploitation of local market power to set higher prices. In addition, the report considered requiring planning approval for the dominant firms, Asda, Morrisons, Safeway, Sainsbury's, and Tesco, prior to opening large stores or performing large extensions of existing stores within fifteen minutes of another one of their stores. A subsequent investigation in 2006 resulted in a strengthened version of SCOP, called Groceries Supply Code of Practice (GSCOP), and a market test to determine whether the largest firms could open new stores or undertake renovations of existing stores [UK Competition Commission, 2006].[3]

Given this background, I evaluate the CC's approval of the Morrisons-Safeway merger by constructing a structural model of the geographic distribution of demand for groceries to explain how consumers' physical locations and preferences interact with stores' locations, pricing, and profitability. Understanding this interaction is central to developing good policy because the geographic distribution of consumers and their preferences determines the profitability of a store at a specific physical location. In addition, stores of the same fascia[4] or parent firm increase local market power because the owning firm will capture part of the demand lost from a price increase when consumers switch to the firm's other stores [Smith, 2004]. Furthermore, the locations of nearby stores as well as their prices and qualities, affect the number of consumers who choose a store and their expenditure. This strategic interaction between stores may also propagate pricing pressures over larger distances through chaining. Consequently, I combine a discrete/continuous model of demand with UK census data to calculate the expected demand each fascia's stores face at different locations. The discrete/continuous demand system [Dubin and McFadden, 1984, Hanemann, 1984] explains both store choice and

[2] 'Multiples' were defined as firms operating at least ten stores having sales areas greater than 600 square meters.

[3] Because the Competition Appeal Tribunal upheld Tesco's challenge to the lawfulness of the market test, the CC undertook an additional inquiry to address concerns about the economic costs of the test as well as its effectiveness [UK Competition Commission, 2009].

[4] *Fascia* refers to the brand on the front of a firm's stores such as *Tesco Express*, *Tesco Metro*, or *Tesco Extra* each of which is a different fascia, although all are owned by the same corporation, Tesco plc.

conditional demand. By aggregating over the distribution of consumer types and locations, as specified in the census data, I can compute the expected demand at each store, given prices. I then use this aggregate, geographic distribution of demand for several other calculations. After recovering marginal costs by inverting the first order equations for profit, I solve for Bertrand-Nash equilibrium prices, and use these prices to calculate firms' profits and bounds on consumers' compensating variation. Throughout these calculations, I assume that stores can set the optimal price for each household type's basket, which greatly simplifies estimation and computation of prices, profits, and consumer welfare. Finally, I use these tools to compare three different policy scenarios: the counter-factual state before the observed merger, the observed state after Morrisons acquired Safeway, and a counter-factual merger between Tesco and Safeway. Given Tesco's dominant position in the industry, the Tesco-Safeway counter-factual should put an upper bound on the adverse effects of the acquisition of Safeway by any of its suitors. Because the census data is disaggregate at the Output Area (OA) level,[5] I can compute welfare effects at this level as well as the change in profits at individual stores.

My approach to estimating demand is most similar to Smith [2004], who develops a discrete/continuous choice model to measure market power, both by computing elasticities and considering the welfare implications of a series of mergers and demergers. Like Smith, I use store characteristics and locations from the IGD data.[6] My estimates benefit from access to the Taylor Nelson Sofres (TNS) Worldpanel,[7] a high quality homescan panel of consumer microdata containing actual consumer choices, prices, and expenditure whereas Smith only observed price-cost margins and had to rely on survey data about consumers' stated store choice and expenditure. Consequently, I can estimate the demand parameters more precisely via Maximum Likelihood Estimation (MLE) instead of his two-step method. I also avoid the estimation and identification problems he encountered because of the limitations of his data. Lastly, I use aggregate units of groceries and price indexes for each fascia, region, and household type which

[5] An OA or *Output Area* is the smallest geographical area for which the UK census provides data. OAs roughly correspond to a ward. On average they contain 125 households and about 300 residents. See http://www.statistics.gov.uk/census2001/glossary.asp#oa.

[6] IGD was formerly known as the *Institute of Grocery Distribution* and was formed from the merger of the Institute of Certified Grocers and Institute of Food Distribution.

[7] Note: the TNS Worldpanel was recently rebranded as the Kantar Worldpanel.

Beckert et al. [2009] constructed from a combined TNS-IGD dataset in order to study the composition of consumers' shopping baskets. Although Beckert et al. [2009] use quadratic utility, I choose a specification with non-linear demand curves which is more appropriate for merger analysis.

There is a long literature on estimating consumer demand for groceries. Deaton and Muellbauer [1980b] provides a survey of 'classic' functional forms, demonstrating the advantages of their Almost Ideal Demand System (AIDS) [Deaton and Muellbauer, 1980a]. These studies were limited by the aggregate data and computational resources available at the time. There is a growing literature on the supermarket industry which builds on advances in modeling discrete choices. Beckert et al. [2009], which uses the same data as this chapter, examines how consumers choose baskets of goods and finds that often a large fraction of a given household type never purchases certain categories such as alcohol or pet food. Briesch et al. [2010] find that not only do about 20% of US consumers stop at multiple stores per shopping trip, but that they purchase different baskets from different stores and that destination categories determine store choice as well as complementarities between different fascia. Also, the pricing strategy of a fascia affects whether consumers prefer to purchase a large or a small basket: EDLP (Every Day Low Pricing) favors large baskets whereas a Hi-Lo strategy makes opportunistic purchases of small baskets of whatever is on promotion preferable [Bell and Lattin, 1998]. Pakes et al. [2006] develop an alternative approach to these estimation strategies using moment inequalities which may be more robust to identifying assumptions and specification.

This chapter continues as follows: first, I describe the model (Section 4.2) and discuss several datasets which I use for estimation and geographic aggregation (Section 4.3). Next, I explain the estimation methodology in Section 4.4 and the results in Section 4.5. Then, I use the geographic distribution of demand to recover firms' marginal costs and to compute price equilibria in Section 4.6. Finally in Section 4.7 I use these results to evaluate how the Morrisons-Safeway and counter-factual Tesco-Safeway mergers affect welfare. Section 4.8 concludes.

The geographic distribution of demand for groceries consists of a structural model of demand integrated over the empirical distribution of consumers to compute the expected demand for a store at a specific location. Several factors affect store choice and expenditure conditional on that choice: store and consumer characteristics, spatial locations of both stores and households, and consumer preferences. This section explains the model of consumer preferences at the core of the computations to evaluate welfare effects of the Morrisons-Safeway merger. These calculations – estimating demand, recovering marginal costs, and solving for pre- and post-merger Bertrand-Nash price equilibria – all depend on integrating predicted demand over the distribution of consumers.

Because firm behavior depends on the aggregate demand that each of store faces, I use a static, discrete/continuous model of demand based on Smith [2004] to estimate demand for each Government Official Region (GOR) and household composition – e.g., students, pensioners, couple with children, etc. (See 4.3 for more discussion of the data.). This model predicts both store choice and conditional expenditure for representative consumers. For the TNS data, described in 4.3.1, the empirical distribution of expenditure at chosen stores falls off exponentially for all levels above the lowest 10%.[8] Consequently, primary shopping constitutes the majority of expenditure so I model just one trip per period instead of a primary and 'top-up' trip as in Smith [2004]. By emphasizing aggregate purchases, I avoid unnecessary difficulties from estimating a more complex model of how consumers choose baskets of goods. In addition, the model is easy to estimate using maximum likelihood. If behavior is actually driven significantly by different shopping modes, the model will be misspecified. I also consider only one trip per household to avoid dealing with dynamic demand issues, such as correlations between trips, inventory keeping, search, and habit formation [Hendel and Nevo, 2004, Seiler, 2010].

[8]When I fit a line to $\log(\text{expenditure})$, adjusted R^2 is 0.97, indicating that an exponential fits the distribution well.

4.2.1 *Indirect Utility*

In the data I observe both consumer and store characteristics, some of which – such as the size of the car park[9] and the distance between the store and the consumer – affect only store choice and others which affect both store choice and grocery expenditure. I specify an indirect utility function, $U_{hj}\left(x_h, z_j, \xi_h, \epsilon_{hj}; \theta\right) = V_{hj}\left(x_h, z_j, \xi_h; \theta\right) + \epsilon_{hj}$, where ϵ_{hj} is an unobserved household-store shock. I split $V_{hj}\left(x_h, z_j, \xi_h; \theta\right)$ into V^{Loc} and V^{Exp} because some covariates affect only store choice but not conditional expenditure. Thus, V^{Loc} captures the utility derived solely from a the attributes of a store's location[10] whereas V^{Exp} is the utility from the location and expenditure at a specific store:

$$V_{hj}\left(x_h, z_j, \xi_h; \theta\right) = V_{hj}^{Loc}\left(x_h, z_j; \theta\right) + V_{hj}^{Exp}\left(x_h, z_j, \xi_h; \theta\right).$$

x_h is household h's attributes, z_j store j's characteristics, θ the parameters to be estimated, and ξ_h a household-specific random effect. The shocks ϵ_{hj} and ξ_h are observed by the household but neither the econometrician nor the firm: firms know the distribution of household types but not the type of a specific consumer.

Both V^{Loc} and V^{Exp} are based on the specification in Smith [2004],which is derived from Hanemann [1984]. Thus, V_{hj}^{Loc}, the utility consumer h receives from just store j's location, is

$$V_{hj}^{Loc} = \beta_{carpark} carpark_j + \beta_{dist} dist_{hj}$$

and $V_{hj}^{Exp}\left(\cdot\right)$, the utility household h receives from both location and shopping (i.e., their conditional expenditure), is

$$\begin{aligned} V_{hj}^{Exp} &= \mu\left[\delta + \alpha_{out}\log p_{out} + \alpha_{price}\log p_j + \alpha_{inc}\log y_h + \alpha_{area} area_j + \xi_h\right] \\ &\times \left(\frac{p_j}{p_{out}}\right)^{-\gamma_{price}}. \end{aligned}$$

[9] I.e., 'parking lot' in American English.

[10] Orhun [2005] shows in an extension of Seim [2006]'s model of entry under incomplete information, that location-specific unobservables play a crucial role in determining the spatial positioning of supermarkets in the US. I ignore this type of shock, though unobserved factors such as the accessibility of a store, unobserved store qualities, or local complementarities with other destinations could be important.

This functional form leads to expenditure shares which are log-linear in price and income. Here, y_h is household h's income, p_j the aggregate index of hedonic prices for a unit of groceries at store j, p_{out} the price of an outside good, $area_j$ the sales area, $carpark_j$ the size of the car park, and $dist_{hj}$ the Cartesian distance between the household and the store.[11] μ is the relative contribution of the utility from purchasing groceries versus the utility from just the store's location. The outside good consists of goods purchased on the trip other than groceries, such as stopping at the chemist, or for a take-away meal. The outside good is necessary to ensure that the utility function is homogeneous of degree zero. Because I do not observe purchases of the outside good, I normalize the price p_{out} to 1, so the term drops out of the utility specification.[12] For practical estimation purposes, I work with

$$V_{hj}^{Exp} = \mu\left[\delta + \alpha_{price}\log p_j + \alpha_{inc}\log y_h + \alpha_{area}area_j + \xi_h\right] p_j^{-\gamma_{price}}.$$

As in Dubin and McFadden [1984], heterogeneity enters the model in two ways, as the unobserved shock ϵ_{hj}, due to unobserved store characteristics, and the household fixed effect ξ_h, due to unobserved household characteristics. Note that ϵ_{hj} affects only store choice whereas ξ_h affects store choice and conditional expenditure. ϵ_{hj} captures unobserved utility which a household gains from a store's characteristics: for example, the store could stock some key good which the consumer values, be conveniently located on the consumer's regular commute, or be near other stores where the consumer shops. I assume that ϵ_{hj} follows a Type I Extreme Value distribution, yielding a multinomial logit specification which is tractable but often suffers from problems such as independence of irrelevant alternatives (IIA) and unrealistic substitution patterns [McFadden, 1981]. In particular, this assumption requires stores to be substitutes. Briesch et al. [2010] show that certain product categories affect fascia choice, that roughly 20% of trips involve stops at two different fascia, and that some fascia are complements with others. However, these facts have not been reproduced with UK data: i.e., UK shopping habits may be different because, among other reasons, US culture is much more car-centric and US houses typically have much more storage space, facilitating both more

[11] Phibbs and Luft [1995] show that Cartesian distance is a good approximation for travel time and that the longer the trip, the higher the correlation between travel time and Cartesian distance.

[12] Beckert et al. [2009] also assume that the price of the outside good is constant.

stops per shopping trip and larger purchases (or purchases of larger pack-sizes) because of reduced inventory costs. Many researchers now use random coefficients to rectify the problems from IIA. To some extent, I mitigate the problems of IIA by estimating demand separately for different types of consumers, as described below. I do not specify an outside option for store choice where $U_{h0} = 0 + \epsilon_{h0}$, as in most studies such as Smith [2004], because I actually observe all of a household's choices in the IGD data.

The random effect, ξ_h, captures unobserved household characteristics such as household size, variation in the outside option, and preferences for atypical items, all of which produce departures from the reference basket and may cause measurement error. Measurement error occurs from unobserved variations in quality, i.e. when the actual price and assortment in a store differ from the reference offer. For example, variations in price and other promotions cause the consumer to substitute from goods in the reference basket to similar products. In the data, I do not observe sales or promotions, so unobserved variations in price will distort the basket's composition from the reference unit of groceries. Furthermore, a store may not stock certain products in the reference basket either because of outages or local variations in the assortment of goods offered. For example, most fascia tend to stock more organic and luxury items in affluent neighborhoods. This variation is an important source of non-price competition and may be a significant, especially given the lack of price variation seen in UK data.[13] Consequently, the true price and quality of the basket actually purchased will differ from the price index for units of groceries and corresponding reference basket. ξ_h attempts to control for these unobserved variations in the quality and price. I assume that ξ_h is the same for all stores in a consumer's choice set to simplify estimation. This assumption depends on the correlation between non-price competition and geographic location. Finally, the actual home scan process used to collect the TNS data provides another source of measurement error (See 4.3.1 and Leicester and Oldfield [2009b]) because of fatigue, attrition, and reporting errors which vary by both product characteristics and household type. The differential reporting of different product types will also distort the observed basket from the baseline. However, these errors are typically about 5% of expenditure or less so I assume I can ignore them.

[13] I am indebted to Jerry Hausman for pointing this out to me.

Bias will occur when a significant component of the household random effect is measurement error and not exogenous variation in the composition of a household's basket. For example, richer households (higher y_h) will tend to purchase higher quality baskets so the price index will understate the true price paid. Similarly, a higher price for the reference basket may cause substitution to lower cost items of lower quality so that purchased quality is lower than the quality used for generic 'units of groceries'. Thus, I expect that in the presence of measurement error (or endogeneity), ξ_h will be correlated with both household expenditure, y_h, and the price index, p_j. In this case, the parameter estimates will be biased.

Following Smith [2004], I assume that the store and household shocks are independent to facilitate estimation via full information maximum likelihood. If ϵ_{hj} and ξ_h are correlated then the model is misspecified and will suffer from selection bias [Heckman, 1979]. These shocks could be correlated if characteristics which affect store choice also affect expenditure, such as advertising or product promotions. To correct this bias, I would need to use one of the estimation strategies in Dubin and McFadden [1984].

I do not specify time, fascia, or regional dummies because they are highly collinear with those included in the price index and the solver failed to converge when I specified them. Similarly, I was unable to estimate δ. Also following Griffith and Nesheim [2010] and Griffith et al. [2010], given the incredible richness of product characteristics in the data, I assume that I observe all product characteristics and that there is no need to control for a (potentially endogenous) unobserved product characteristic, unlike the BLP model [Berry et al., 1995] of differentiated products.

Following Beckert et al. [2009], I use household composition to categorize households by type and then estimate the model separately for each group. When the variation in each group is low, this method controls for some of the unobserved heterogeneity.[14] This assumption simplifies computation and is supported by the low estimated variance for ξ_h (See 4.5.).

To avoid infeasible expenditure shares, I assume that ξ_h is distributed as a truncated normal where the bounds are chosen to ensure that $w_{hj} \in [0, 1]$. This, unfortunately, causes the support of ξ_h to de-

[14]Random coefficients would probably improve the fit of the model.

pend on the data potentially making ML estimation inconsistent.[15] In the data I observe that the size of the tails which need to be truncated have extremely small probability mass. In addition, some households violate these expenditure share bounds, but they are only 59 out of 16897 observations (0.34%). These violations are spread evenly across household types. Consequently, in order to facilitate computation, I assume that $\xi \sim \mathrm{N}\left(0, \sigma^2\right)$.

Lastly, I set $\alpha_{inc} = 1$ in order to identify the scale parameter, μ. I do not impose the restriction that $\alpha_{inc} = \gamma_{price}$ as in Smith [2004], so γ_{price} is free. See 4.4.2 for further discussion.

After applying all of these considerations, the indirect utility from expenditure is

$$V_{hj}^{Exp} = \mu \left[\delta + \alpha_{price} \log p_j + \log y_h + \alpha_{area} area_j + \xi_h\right] p_j^{-\gamma_{price}}.$$

4.2.2 *Conditional Demand*

According to classical microeconomic theory, Roy's Identity can be used to derive the conditional demand. Because the indirect utility depends on an aggregate price index, not the prices of the individual goods, I assume that I can treat the price index as a 'price' and apply Roy's Identity. If this is not true, then the aggregate conditional demand for groceries must be derived by applying Roy's Identity to each individual good and then aggregating. For the case where utility depends only on the total price of the basket via an index restriction, the true derived aggregate conditional demand, $q_{hj}^{true} \propto q_{hj}$, and the constant of proportionality depends on the expenditure share weights, as shown in Appendix B.1.

Under this assumption, the conditional demand is

$$\begin{aligned} q_{hj} &= -\frac{\partial V_{hj} / \partial p_j}{\partial V_{hj} / \partial y_h} \\ q_{hj} &= \frac{y_h}{p_j} \left\{ \gamma_{price} \left[\delta + \alpha_{price} \log p_j + \log y_h + \alpha_{area} area_j + \xi_h\right] - \alpha_{price} \right\} \end{aligned}$$

and the expenditure share is

[15] ξ_h is a function of the parameters because the shock is obtained by solving the relationship for the observed and calculated conditional demand, which is computed below using Roy's Identity.

$$w_{hj} \quad = \quad \gamma_{price}\left[\delta + \alpha_{price}\log p_j + \log y_h + \alpha_{area} area_j + \xi_h\right] - \alpha_{price}.$$

Because $\alpha_{price} < 0$ for the price term, the expenditure share has a simple interpretation: the constant term, $-\alpha_{price}$, captures the subsistence expenditure required by a household and the bracketed term represents expenditure on goods which are not necessities.

4.3 DATA

Evaluating the Morrisons-Safeway merger requires predicting the demand a store faces and then computing the equilibrium prices under different policy scenarios and the changes in consumer welfare. I estimate the structural model of demand from data on household shopping trips, household attributes, and store characteristics. I use a dataset which Beckert et al. [2009] (BGN, hereafter) assembled by merging the TNS Worldpanel, which includes data on household consumption and characteristics, with the IGD database of store characteristics.[16] In addition, BGN construct a price index for units of a reference basket of groceries.

Recovering marginal costs, solving for Bertrand-Nash equilibrium prices, and calculating changes in welfare all utilize the same method: compute the expectation of the appropriate equation, such as the profit first order equations, and then solve for the appropriate factor, such as marginal costs or equilibrium prices. The expectation is formed by integrating predicted demand at each store over the empirical distribution of consumers as specified in Table KS020 of the UK 2001 Census, which contains OA-level data on household composition.

I now discuss the main features of the TNS and IGD datasets further.

4.3.1 *TNS Worldpanel*

I use the TNS data from November 2003 to November 2004 to estimate demand for units of groceries by household type and region.[17] This

[16] I only work with data for England and Wales because too many stores are missing from the data for Northern Ireland, Scotland, and the various Isles.

[17] Because I observe data only after the Morrisons-Safeway merger, I estimate the post-merger demand system and then use the structural model to compute the unobserved pre-merger state, which provides a basis for welfare comparisons. See 4.7.

data is a homescan panel of consumer purchases from the UK Worldpanel for Fast-Moving Consumer Goods (FMCG) sector. Households use a scanner to record information about their purchases at the SKU level, including the date, price, quantity, and location of purchase. In addition, they can enter items without a barcode, such as raw fruit and vegetables. The scanning system transmits results electronically to TNS who periodically verify the information using the consumer's receipts. In addition, TNS provides further data on household characteristics from which I determined each household's composition type. The only store characteristics in the data are the address and fascia.

The data which I use was prepared by BGN. They divide the data by GOR and household type. The household types are similar to those in the UK 2001 Census's Table KS020. Because marital status should not affect consumption behavior, BGN aggregate household types which differ only by marital status: e.g., a married couple with dependent children is treated the same as an unmarried couple with dependent children. In addition, BGN construct a price index for a unit of groceries from an aggregate of hedonic price indexes for each fascia by region and household type. I use this aggregate index as the price when estimating demand. This index varies by year, month, GOR, and fascia which complicates estimation by decreasing the amount of variation in price. In addition, as explained in Section 4.2.1, the price index assumes that each household type purchases the same reference basket and could introduce measurement error to the extent that a household's basket differs from the reference basket. See Section B.3.3 in the Appendix for further discussion of the price index.

Although, the TNS data includes all shopping trips which a household reports during a period, BGN draw a single trip at random for each household. This facilitates estimation by avoiding dynamic issues such as correlation, inventories, and habit formation, but ignores much of the information in the data. For merger evaluation, this may not matter because the analysis depends on the expected demand a store faces, which is formed by aggregating over all representative consumers who shop at a store. Because of regional pricing, aggregate demand at all of a fascia's stores determines each fascia's regional pricing behavior; consequently, dynamics are less important for accurate assessment of the welfare consequences than predicting the consumption of individual consumers. In addition, BGN assume each household's choice set consists of the 30 stores closest to their home. For most households, this is a reasonable assumption because of consumers' aversion to traveling more than 10 to 15 minutes to shop [Smith, 2004]. However, BGN find that a small fraction of households shop at more distant stores so they exclude these consumers from the dataset. These distant purchases may be determined by the shoppers' commute paths, as Houde [2011] found for gasoline purchases.

Section B.3 provides more information on the issues associated with using TNS data.

4.3.2 *IGD Data*

Because the TNS data lacks store characteristics, BGN supplement it with the IGD data, which contain the store characteristics for all of the stores of the major supermarket firms as well as many smaller regional and local companies. The data include the date of opening, closing, and the last renovation as well as the store's address, post code, sales area, gross area, and car park size. The IGD data span 1900-2004 which enables me to determine which stores operated under the Safeway fascia prior to their acquisition by Morrisons. After data cleaning and allowing for store acquisitions, conversions, and closures, my dataset has $10,883$ stores for England and Wales.

Following BGN, I group the data into the following fascia: Aldi, Asda, Budgens, Coop, Iceland, Kwik Save, Lidl, Marks & Spencer, Morrisons, Netto, Other, Safeway, Sainsbury's, Somerfield, Tesco, and Waitrose as well as SainS and TescS, for Sainsbury's and Tesco's convenience stores. Coop consists of the different regional Cooperative movement stores whereas Other contains the fringe of regional and smaller grocery retailers. This simplification facilitates estimation but overstates these firms' market power because they lack the centralization, logistics, and focused strategies of the other fascia.

4.4 DEMAND ESTIMATION

I estimate the demand system via maximum likelihood. Because of its efficiency, MLE is an good choice when endogeneity and unobserved product characteristics are not important. However, the estimator will be inconsistent if the assumptions in Section 4.2 fail.

The likelihood for a household, $L_h(x;\theta)$, is composed of two pieces: the likelihood of choosing a location (i.e. a specific store) and the conditional likelihood of expenditure:

$$\begin{aligned} L_h(x;\theta) &= \Pr\left[h \text{ chooses } j, \text{buys } q_{hj}\right] \\ L_h(x;\theta) &= \Pr\left[h \,\text{chooses}\, j\right] \times \Pr\left[q_{hj}\,|h \,\text{chooses}\, j\right] \end{aligned}$$

where q_{hj} is the units of groceries household h purchases at store j. Consequently,

$$\log L_h(x;\theta) \quad = \quad \ell^{choice} + \ell^{q}$$

where ℓ^{choice} is the log-likelihood computed from the probability of household h choosing store j and ℓ^{q} is the log-likelihood of purchasing q_{hj} units of groceries conditional on shopping at j. These two components of the log-likelihood are easily calculated using the shocks, ϵ_{hj} and ξ_h, because they follow Type I Extreme value and normal distributions, respectively.

4.4.1 *Computation of the Log-likelihood*

Give the distributional assumptions, I estimate the full model using maximum likelihood in a single step as follows:

1. Compute a set of initial guesses by drawing points about the OLS point estimates of conditional expenditure on the covariates. I use multiple, randomly-drawn, starting points in a region about the OLS estimates.

2. Maximize the full log-likelihood:

    a) Compute the household-specific shock, $\xi_h(\theta)$, by inverting the equation for conditional demand

    b) Compute the log-likelihood for store choice, $\ell_h^{choice}\left(x_h, z_j, \xi_h(\theta);\theta\right)$

    c) Compute the log-likelihood for conditional expenditure, $\ell_h^{q}\left(x_h, z_j, \xi_h(\theta);\theta\right)$

Because $\epsilon_{hj} \sim$ Type I Extreme Value, the probability of choosing a store (the *discrete* part of the discrete/continuous choice model) is the convenient multinomial logit form [McFadden, 1981], which facilitates computation of the log-likelihood for the consumer's store choice:

$$\Pr\left[h\, chooses\, j\right] = \frac{\exp\left[V_{hj}\left(x_h, z_j, \xi_h; \theta\right)\right]}{\sum\limits_{k\in J_h}\exp\left[V_{hk}\left(x_h, z_k, \xi_h; \theta\right)\right]}$$

where, J_h is household h's choice set and j the chosen store. Consequently, the log-likelihood, ℓ_h^{Store}, is

$$\ell_h^{choice}\left(x_h, z, \xi_h; \theta\right) = -\log\sum_{k\in J_h}\exp\left[V_{hk} - V_{hj}\right]$$

where I have rewritten the fraction in the choice probability to be more stable numerically.

ℓ^{choice} is a function of the indirect utility, V_{hk}, which in turn depends on the shock, ξ_h. I obtain ξ_h by inverting the equation for expenditure share

$$\xi_h = \frac{1}{\gamma_{price}}w_{hj} + \frac{\alpha_{price}}{\gamma_{price}} - \left[\delta + \alpha_{price}\log p_j + \log y_h + \alpha_{area} area_j\right].$$

I also use the shock ξ_h to compute ℓ^q, the log-likelihood from consumer expenditure conditional on store choice. Because ξ_h is a truncated normal, the likelihood is

$$\begin{aligned}\ell_h^q\left(x_h, z_j, \xi_h; \theta\right) &= -\frac{1}{2}\log\sigma^2 - \frac{1}{2}\log 2\pi - \frac{1}{2\sigma^2}\xi_h\left(w_{hj}, x_h, z_j; \theta\right)^2 \\ &- \log\left[\Phi\left(b_h\right) - \Phi\left(a_h\right)\right].\end{aligned}$$

Φ is the normal cumulative distribution function, a_h and b_h are the truncation bounds which ensure that $w_{hj} \in [0, 1]$, and σ^2 is the variance of the shock. I have written ξ_h as a function of covariates and parameters to emphasize that the support of ξ_h depends on the data and parameter estimates. Under these conditions, MLE is still consistent but may converge faster than $\sqrt{N}$ [Donald and Paarsch, 1996]. I dropped the term $\log\left[\Phi\left(b_h\right) - \Phi\left(a_h\right)\right]$ from the estimation procedure because in practice the bounds are so far apart that this normalization term is nearly constant.

4.4.2 *Identification*

Variation in household and store characteristics affects the indirect utility and, thus, likelihood in two ways. Some characteristics – such as the size of the car park and household-store distance – only affect store choice whereas others – such as price, income, and sales area – affect both store choice and conditional expenditure. Consequently, variation in the data leads to the identification through one or both of these channels. In addition, identification depends on the restrictions imposed on the model as well as the distributional assumptions for the errors. These restrictions include the standard logit normalization of the variance of ϵ_{hj} to $\pi^2/6$ [Train, 2009] and the normalization $\alpha_{inc} = 1$ in order to identify the scale parameter, μ.[18] Identification of α_{price}, α_{area}, γ_{price}, and σ^2 follows from the identification of μ, the log-linear functional form of the conditional expenditure equation, and the variation of the covariates income, y_h, price, p_j, and sales area, $area_j$. $\beta_{carpark}$ and β_{dist} are identified through variation in the size of car park and household-store distance when otherwise similar households choose different stores.

In theory, δ should be identified from the conditional expenditure and variation of the covariates, but the solver diverges when I include δ in the model. Lack of variation in the price index could contribute to this problem because the standard deviation of the price index for each GOR, household type, and month is usually less than 20% and often closer to 10% of the mean price. Firms' national pricing policies further reduce price variation. Ellis [2009] finds that although UK grocery prices fluctuate by up to 40% per week, monthly observations understate the amount of variation. In addition, UK consumers are extremely price sensitive so weekly fluctuations in price and promotions influence expenditures and lead to departures from the reference basket of groceries, i.e. measurement error. Both effects decrease the correlation between expenditure and the monthly price index. Diagnostic OLS regressions for the expenditure share equation have R^2 less than 0.2. Consequently, a model with a higher-frequency price index[19] or which focuses on aggregate expenditure may produce more reliable results. I control for the measurement error in the basket of

[18]This normalization is reminiscent of the restriction that $\alpha_{inc} = 1$ in the AIDS model, as required by economic theory.

[19]A household-specific price index is another option, especially because some consumers never purchase certain goods such as alcohol.

groceries by estimating the model by household composition using price indexes which are computed for each region, household type, fascia, and month. ξ_h captures the remaining unobserved variation in consumer preferences. Adding random coefficients to the model might reduce this error further.

In regions of the parameter space where γ_{price} is close to zero, the utility reduces to an expression which is nearly linear in covariates, making identification of μ impossible because the interaction between household and store characteristics disappears. [Train, 2009]. This problem complicates estimation because the price index is normalized so that $p_j \approx 1$ and point estimates for $\hat{\gamma}_{price}$ range from 0.26 to 0.4 (See Table Table 34.), making $p_j^{-\gamma_{price}}$ quite flat and close to 1.

The model could also be misspecified because of the failure of Roy's Identity, endogeneity, or incorrect treatment of shocks. Assuming Roy's Identity holds for the price index, the parameter estimates should be consistent and unbiased after controlling for selection via the model for store choice. But, when Roy's Identity is not valid, the parameters will not be identified. For example, if prices enter the indirect utility via an index restriction, then conditional expenditure is really

$$\begin{aligned} w_{hj} &= \Psi \left\{ \gamma_{price} \left[\delta + \alpha_{price} \log p_j + \log y_h + \alpha_{area} area_j + \xi_h \right] \right. \\ &- \left. \alpha_{price} \right\}, \end{aligned}$$

for some constant of proportionality Ψ, which is a function of the expenditure shares (See Section B.1 for the derivation.). Then δ and γ_{price} are not identified from conditional expenditure alone because only $\Psi \gamma_{price}$ can be identified from variation in y_h (α_{price} and α_{area} are still identified). Thus, δ and γ_{price} must be identified from variation in store choice. Assuming Roy's Identity holds for the price index is equivalent to the restriction that $\Psi = 1$.

Endogeneity potentially affects the estimation results through simultaneity and measurement error. Price endogeneity is a problem, particularly if firms observe ξ_h and incorporate it into their price setting mechanism. If households which spend more also prefer more expensive, higher quality goods, then the coefficient on price will be biased. In addition, income could be correlated with ξ_h if more affluent households spend more because they purchase higher priced, higher quality goods, such as organic produce, which deviate from the reference basket. But, the correlation between ξ_h and income is likely small because, ceteris paribus, the different fascia position themselves to appeal to certain consumer groups. By controlling for fascia in the price indexes, I attempt to control for these sources of endogeneity.

The discrete/continuous choice specification should control for selection bias by explicitly modeling store choice as long as the sample of households in the dataset is representative. However, if $\mathbb{E}\left[\epsilon_{hj}\xi_h\right] \neq 0$ then selection bias is possible: for example, households which prefer certain fascia spend more at those fascia. Or, an advertising campaign could affect both store choice and expenditure. Then the parameter estimates will be biased.

The unobserved quality of store locations – such as ease of access, proximity to other stores, or exceptional staff – may matter, as Orhun [2005] found for US supermarkets. Such a shock is similar to unobserved product-market characteristic, ξ_{jt}, in models like Berry et al. [1995] and often correlated with price. But fascia pursue a national pricing strategy so local characteristics only affect price through the resulting equilibrium price for a region. Consequently, local fixed effects should not be significant.

4.4.2.1 *Estimation and Numerical Issues*

I discuss computational and numerical issues in Section B.2 of the Appendix.

4.5 DEMAND ESTIMATION RESULTS

Table 34 displays the point estimates for the model of household demand for groceries by household type.[20] The structural parameters consist of the β coefficients, which only affect the utility a consumer obtains from a store's location, and the other coefficients, which affect the utility from a store's location and from consumption. The coefficient σ^2_ξ is the variance of the household-specific shock, ξ_h. All point estimates cannot be rejected at the 5% level (or better) and have the expected signs: i.e., the results show that consumers prefer lower prices, shorter travel times, easier parking, and more variety (i.e. more

[20] When comparing the scale of the point estimates for location and conditional expenditure, it is important to multiply coefficients of V_{hj}^{Exp} by the scale parameter, μ. E.g., to compare β_{dist} to $\mu\alpha_{price}$, instead of α_{price}.

sales area). In addition, estimates are similar in sign and magnitude across household types.[21]

Although the estimates for the variance of the household shock, ξ_h, appear small, they have the same order of magnitude as the variation in price. Small $\widehat{\sigma_{\xi}^2}$ may be caused by estimating by household type to control for heterogeneity in household composition. This result supports the assumption that households purchase a type-specific reference basket of units of groceries and that departures from this basket, in terms of quality, assortment, and price are small: i.e., the price index captures consumer behavior well and there is little measurement error. The bounds on the expenditure share, w_{jh}, further limit the variation in the household-specific shock.

The largest parameter in magnitude is $\hat{\beta}_{dist}$ for all household types, indicating that consumers dislike traveling, which agrees with other studies [Smith, 2004]. Single households – young, pensioners, and single parents – have by far the greatest distaste for distance, reflecting either higher values of time or lack of a partner to share the work of running a household. The coefficient for the size of $\beta_{carpark}$ is also large and significant. Both of these point estimates reflect the importance of convenience to consumers.

Because the conditional own-price elasticity of a household for a store is $\eta_{hj} = -1 + \gamma_{price}\alpha_{price}w_{hj}^{-1}$, the product of γ_{price} and α_{price} affects a household's elasticity: price sensitive household types have more negative values for $\gamma_{price}\alpha_{price}$. The estimates show, then, that single pensioners and single parents with children are the most price sensitive whereas single non-pensioners are the least price sensitive, possibly because single pensioners and single parents tend to have lower incomes. When I examined the mean and medium expenditure vs. household type, this story only holds for single pensioners. Perhaps, single parents effectively have lower incomes vis-a-vis other household compositions because they lack the scale economies of couples and, consequently, have less time for 'home production' activities such as shopping and housekeeping. Also, total expenditure on groceries in the TNS data may not be a reliable measure of income for households receiving welfare benefits.

[21] I performed likelihood ratio tests to check if I could aggregate similar household types. In all cases, the test rejected the hypothesis that any of these household types could be combined at the 1% level. Consequently, I do not consider aggregation of household types further.

To recast these results in more economically meaningful terms, I compute elasticities and profits and compare them with the perceived ranking of firms in the marketing literature and popular press (See Section 4.6.). These metrics confirm that Asda, Morrisons, Sainsbury's, and Tesco dominate the UK supermarket industry and exert significant pressures on their rivals while being constrained primarily by each other.

4.5.1 *Elasticities*

To understand the extent of firms' market power, I compute own- and cross-price elasticities as well as elasticities for income, area, and distance. First, I calculate the elasticities observed for each store in an OA's choice set. Using these store-level elasticities, I can evaluate the local market power at a highly disaggregate level. Next, I average elasticities over all fascia and OA in a GOR, weighting by the population in OAs whose choice sets include the relevant fascia. Consequently, the reported elasticities represent the market power a typical consumer encounters when a fascia is in their choice set.[22] The averaged elasticities measure the regional market power of the different fascia. This approach handles the differentiation of stores, both by location and quality. To obtain the national average elasticities, I average the elasticities over GOR, weighted by the population in each GOR.

More formally, I compute the elasticities for each OA's choice set from the expected demand, q_{ja}, each store j faces in OA a:

$$q_{ja} \quad = \quad \sum_{h \in H_a} \int \int n_{ha} s_{jha} q_{jha} dF\left(y, \xi\right),$$

where H_a is the set of household types in OA a, n_{ha} is the number of households of type h in a, and s_{jha} is the probability of someone in a choosing store j from the stores in a's choice set. In addition, I integrate over the distribution for income and ξ, assumed to be log-normal and normal, respectively. Then, I compute the price-elasticities for store j with respect to price for fascia k for each OA's choice set:

[22] To some extent, this is an artifact of only defining a store's choice set as the 30 closest stores. However, most consumers visit more distant stores infrequently because of their strong distaste for travel and propensity to shop near their homes.

$$
\begin{aligned}
\eta_{j,k} &= \frac{p_k}{q_{ja}}\frac{\partial q_{ja}}{\partial p_k} \\
\eta_{j,k} &= \frac{p_k}{q_{ja}}\sum_{h\in H_a}\int\int n_{ha}\left(\frac{\partial s_{jha}}{\partial p_k}q_{jha}+s_{jha}\frac{\partial q_{jha}}{\partial p_k}\right)dF\left(y,\xi\right) \\
\eta_{j,k} &= \frac{p_k}{q_{ja}}\sum_{h\in H_a}n_{ha}\int\int s_{jha}\left(q_{jha}\left(\frac{\partial V_{hja}}{\partial p_j}\mathbb{I}\left[j\cap k\right]-\sum_{i\in C_a}s_{iha}\frac{\partial V_{iha}}{\partial p_k}\right)\right. \\
&- \left.\mathbb{I}\left[j\cap k\right]\left(\frac{q_{hja}}{p_j}+\gamma_{price}\alpha_{price}\frac{y_{ha}}{p_j^2}\right)\right)dF\left(y,\xi\right).
\end{aligned}
$$

Here, C_a is the set of stores in the choice set for OA a and $\mathbb{I}\left[j\cap k\right]$ is 1 iff stores j and k have the same fascia and 0 otherwise. This formula shows that a price increase affects demand at a store at both the extensive margin through the choice probability, s_{jha}, and the intensive margin through the conditional demand, q_{jha}. Competition from rival stores, on the other hand, only operates through the choice probability via the term $-\frac{p_k}{q_{ja}}\sum_{h\in H_a}\int\int s_{jha}q_{jha}\sum_{i\in C_a}s_{iha}\frac{\partial V_{iha}}{\partial p_k}$, which explains why some fascia devote considerable resources to non-price competition to capture consumers at the extensive margin. When several stores with the same fascia are in an OA's choice set, their local own-price elasticities increase because the fascia's other stores capture some of the customers who would substitute away on a price increase. This multi-store effect increases local market power for major firms by making demand less elastic. If firms were free to pursue price flexing to exploit their local market power, then most elasticities would increase in magnitude because the term $\sum_{i\in C_a}s_{iha}\partial V_{iha}/\partial p_k$ reduces to $s_{kha}\partial V_{kha}/\partial p_k$. Consequently, regional (or national) pricing may be an effective strategy to make demand less elastic, increasing both market power and profits. Regional pricing may also facilitate tacit collusion because it is easier to monitor and coordinate on price.

Table 35 presents population-weighted own- and cross-price elasticities. Each row represents a fascia's set of elasticities with respect to the prices of the fascia listed in the column. All own-price elasticities are negative, ranging from -4.22 to -5.54 which correspond to a price-cost margin of about 18-24%, using the Lerner formula. This markup is higher than the low margins typically reported for the supermarket industry because I can only recover marginal costs and do not observe the fixed costs involved in distribution, logistics, and other infrastructure.

Cross-price elasticities are significantly smaller and show considerable variation in the market power of the different fascia, ranging from 0.00 to 0.52. These surprisingly low values show that firms wield considerable local power, driven by consumers' distaste for travel and multi-store effects. Most second tier firms have relatively larger cross-price elasticities with respect to the Asda, Morrisons, Sainsbury's, and Tesco (AMST) fascia than the AMST fascia have with them, except for a few fascia such as Iceland. These results show that the AMST fascia exert more competitive pressure on smaller firms than vice-versa. Cross-price elasticities are also larger between fascia which pursue similar formats and strategies, such as Aldi vs. Kwik Save or Netto vs. Kwik Save which compete for similar market segments. Tesco's dominance is clear: almost every fascia's largest cross-price elasticity is with respect to Tesco. These results are similar to those found in the CC's 2006 supermarket investigation [UK Competition Commission, 2006].

The elasticities for the Limited Assortment Discounters (LAD) (Aldi, Lidl, and Netto) are surprising because these stores pursue similar strategies so their cross-price elasticities should be symmetric. However, the elasticities with respect to changes in Lidl's price are extremely small because the price index for Lidl is 35% larger than Aldi's and Netto's prices on average, making Lidl a particularly unappealing substitute for price-sensitive consumers. In addition, cross-price elasticities for Aldi and Lidl with respect to Netto are about half of Netto's because Aldi and Lidl have twice as many stores as Netto and the geographic averaging involved in the calculations.

Elasticities also show some limitations to my model. The model overstates the market power of the Coop and Other fascias. Coop stores are part of a larger buyer co-operative and lack the central coordination and economies of scale and scope of AMST. The Other fascia contains many independent, fringe firms which operated a limited number of stores on a regional or smaller basis. Classifying them as a single fascia engaged in regional price setting considerably overstates their market power and influence on the computed equilibrium or aggregated elasticities. Also, fascia which consistently operate smaller formats have smaller elasticities. Finally, the logit functional form assumption requires that rival fascias are substitutes.

Another possible source of low cross-price elasticities is missing data. If many stores are missing, then the choice sets constructed from the IGD data will understate the density of stores, increasing the distance between stores and, hence, decrease cross-price elasticities. Because most of the missing stores are independent operators and small convenience stores, this should not have a significant impact on results.[23]

As a check, I compared my results to Smith [2004], who computes elasticities using a similar model. He finds own-price elasticities which are roughly a factor of two larger in magnitude than mine and cross-price elasticities which are a factor of 10 to 100 larger. These differences arise for several reasons: he does not construct elasticities at the choice set level; he estimates demand using data on fascias' price-cost margins and from consumer surveys about shopping habits; and, he performs sensitivity analysis to determine α_{price} and α_{inc} (β_1 and β_2 in his notation) because he cannot identify them separately.

4.6 GEOGRAPHY OF COMPETITION

By combining the parameter estimates with census and postcode data, I constructed a geographic distribution of demand to understand how store locations and the distribution of consumers affect competition and consumer welfare. These spatial locations help determine how much local market power a fascia can exploit and how pricing pressures propagate via chaining (where a store's price affects more distant stores through its impact on intervening stores). I use this method of aggregating demand over the empirical distribution of consumers in order to calculate marginal costs, price equilibria, fascia profits, and consumer welfare. This section of the chapter focuses on how the method works and the recovery of marginal costs. In Section 4.7, I use this technique to evaluate the welfare consequences of the Morrisons-Safeway merger and a counter-factual merger between Tesco and Safeway.

Supermarkets make a complex offer to consumers by varying price, quality, range, and service (PQRS), all of which firms choose to maximize their profits. UK Competition Commission [2006] found that supermarkets, with the exception of Coop and Somerfield, set national prices, based on Tesco's argument that prices propagate through the

[23] I am grateful to Patrick Mitchell-Fox, an analyst at IGD, for clarifying this issue.

chain of substitution, producing a national price. I assume that firms sets prices for their fascia regionally to capture the strategic effect of regional variation in the assortment (range) of products: in practice, assortment varies on an even smaller scale.[24] Unfortunately, this variation is difficult to observe. This model of competition is incorrect if either variation in assortment is large and local or it is small. In the former case, firms essentially choose a store-specific price for a unit of groceries and pursue price flexing; in the latter case, they choose a national price. The regional pricing assumption overstates (understates) firms' market power if pricing is at the national (store) level. However, if Tesco's argument is valid and there is sufficient density of stores and consumers, then the chain of substitution produces national prices.[25] It is not clear how these different models of competition would affect the welfare change under a merger.

For all of these calculations, I assume that a static model captures the relevant economics for this policy analysis. The IGD data show that during the 1980s and 1990s, firms' built primarily mid-sized (280 m^2 to 1,400 m^2) and large-sized (> 1,400 m^2) stores. Since 2000, however, competition has focused on convenience stores (< 280 m^2), because of legal restrictions on building larger formats.[26] The market test proposed by the CC in 2006 strengthened these planning constraints [UK Competition Commission, 2006].

4.6.1 *Geographic Aggregation*

I recover each fascia's marginal cost for each region and household type by inverting the first order conditions (FOCs) for expected profit, assuming Bertrand-Nash competition in prices. I assume that each firm functions as a product in a multi-product, oligopoly and sets prices for its fascia regionally.

To calculate marginal costs, I calculate the geographic expectation of the FOCs as follows:

1. For each GOR, compute pair-wise distances between all stores and OAs using the UK Office of National Statistics (ONS) 2006 All Fields Postcode Directory (AFPD), which maps post codes and

[24] Controlling for fascia should capture most of the variation in quality and service because firms try to provide a uniform shopping experience across stores.

[25] Using the machinery of this chapter, it should be possible to test Tesco's national pricing argument.

[26] Firms may also have been constrained by a lack of suitable sites for larger stores.

OAs into 'northings' and 'eastings', a standard X-Y coordinate system for UK geographical data. Given the high precision of UK postcodes this introduces little error compared to a household's true location.

2. Form choice sets for each OA from the 30 closest stores.

3. Calculate the expected FOCs for profit at each store by integrating demand over household types, income, and OAs using the KS020 Table of the UK 2001 Census. The demand is the product of the probability that a household chooses a specific store and the conditional demand. Then, aggregate across stores by fascia.

4. Invert the first order equations and recover marginal costs for each fascia, household type, and region, allowing for strategic interactions for the firms (Sainsbury's and Tesco) which control multiple fascia.

After recovering marginal costs, I use a similar method to compute expected firm profits and price equilibria.

I assume that marginal cost is the same for all of a fascia's stores and that scale economies and network effects do not affect marginal costs; i.e., that a linear specification is sufficient. To the extent that these factors matter, my model will be misspecified. Both Jia [2008] and Holmes [2011] show that both of these forces affect competition among US discounters, but the smaller scale of the UK may decrease the importance of these factors.

Before discussing the results in Section 4.6.2, I first explain the equilibrium assumption, profit function, and derivation of the first order equations.

4.6.1.1 *Equilibrium Assumption*

Marginal cost recovery is based on computing profits and FOCs at equilibrium prices. But, solving for equilibrium prices when there are multiple consumer types and hundreds of products is technically challenging, even if competition is limited to just several hundred key value items (KVI) [UK Competition Commission, 2006]. Consequently, I assume that the fascia compete via Bertrand-Nash competition in each region and they choose an optimal price for the reference basket of each household type. This assumption simplifies the problem of solving for the price equilibrium and marginal costs.

Because there are many more goods than household types and households purchase a type-specific reference basket, this assumption is equivalent to each firm choosing the vector of prices of individual goods, p, such that they satisfy

$$\begin{pmatrix} \tilde{p}_1 \\ \dots \\ \tilde{p}_H \end{pmatrix} = \begin{bmatrix} \omega_1^T \\ \dots \\ \omega_H^T \end{bmatrix} p$$

for some optimal price indexes, $\tilde{p}_h$, and vectors of expenditure shares, ω_h. Thus, firms can adjust prices in a way that generates the profit-maximizing price for each type's basket. Clearly, this equation mapping prices to basket prices only has a solution if the matrix of expenditure shares is full rank.[27] Appendix 7 in UK Competition Commission [2000] provides further support: Sainsbury's, Somerfield/Kwik Step, Tesco, and Waitrose all report that they focus pricing strategy on the basket price consumers face as part of their offer. Furthermore, Tesco buying managers set prices to target certain subgroups, consistent with central Tesco pricing policy.

4.6.1.2 *Profits Under Multi-Product Oligopoly*

Given the equilibrium optimization assumption, the geographic profit function aggregates each household's expected conditional demand over household types and OAs. Thus, π_{jha}, the profit at store j from household type h in OA a, is

$$\pi_{jha} = n_{ha} \left(p_j - c_j \right) q_{jha} s_{jha}.$$

Aggregating over all OAs in a region, conditional on household type, yields a store's conditional profits:

$$\pi_{jh} = \sum_{a \in \mathcal{OA}} n_{ha} \left(p_j - c_j \right) q_{jha} s_{jha}.$$

Then, total profit per store is just the sum over consumer types:

[27]When the matrix is full rank there will also be multiple solutions because of the non-trivial null space of the expenditure share matrix.

$$\pi_j = \sum_{h\in\mathcal{H}}\sum_{a\in\mathcal{OA}} n_{ha}\left(p_j - c_j\right) q_{jha}s_{jha} + F_j,$$

where F_j is the fixed cost of opening and operating store j. To obtain a fascia's profits, I aggregate over the set of all stores the fascia operates, $\mathcal{F}_f$:

$$\pi_f = \left(p_f - c_f\right)\sum_{j\in\mathcal{F}_f}\sum_{h\in\mathcal{H}}\sum_{a\in\mathcal{OA}} n_{ha}q_{jha}s_{jha} + \left|\mathcal{F}_f\right| F_f.$$

Because I observe neither a household's income nor shock, ξ_h, I integrate over their distributions, assumed to be normal and log normal, respectively. The moments for income are computed from the TNS data.[28]

Sainsbury's and Tesco operate multiple fascias which increases their market power, according to the theory of multi-product firms. Consequently, the FOCs must include these effects. Let firm f operate T different fascia. A firm's individual stores are enumerated in sets $\mathcal{F}_f^1, \ldots, \mathcal{F}_f^T$ for each fascia type t. Then, the firm f's set of stores $\mathcal{F}_f = \bigcup_{t\in T}\mathcal{F}_f^t$ is the union of the set of stores for each fascia type. Now, each firm's profit is the sum of the profits of each its fascia:

$$\begin{aligned}\pi_f &= \sum_{t\in T}\sum_{j\in\mathcal{F}_f^t}\sum_{h\in\mathcal{H}}\sum_{a\in\mathcal{OA}}\int\int n_{ha}\left(p_f^t - c_f^t\right) q_{jha}s_{jha}dF\left(y\right)dF\left(\xi_h\right)\\ &+ \left|\mathcal{F}_f^t\right| F_f^t.\end{aligned}$$

In most firms, this reduces to the profits from a single fascia.

4.6.1.3 *FOCs Under Multi-Product Oligopoly*

Recovering marginal costs or solving for a new price equilibrium under different market structures requires computing the multi-product profit FOCs. Consequently, cross-price effects are important when firms set prices if the markets served by the different fascia overlap sufficiently. E.g., if the markets for Tesco and Tesco Metro have have large

[28] I perform the integration via a Gaussian-Hermite product rule with three nodes in each dimension because using more nodes had negligible impact on the results. I tried using sieve estimation [Chen, 2007] to approximate the income distribution but there was not enough data to use a non-parametric method.

cross-price elasticities – i.e. are good substitutes – then assuming that competition is single-product oligopoly will understate Tesco's true market power.

Let the multi-product firm f choose prices $p_f^1, \ldots, p_f^T$. Because firms choose the optimal price for each household type's basket, this complex optimization problem reduces to a set of separate optimization problems for each household type. Thus, firms solve a non-linear system of Bertrand-Nash FOCs where each fascia's FOC conditional on household type is:

$$\begin{aligned}
0 &= \frac{\partial \pi_f}{\partial p_f^v} \\
0 &= \sum_{j \in \mathcal{F}_f^v} \sum_{h \in \mathcal{H}} \sum_{a \in \mathcal{OA}} n_{ha} q_{jha} s_{jha} \\
&+ \sum_{t \in T \setminus \{v\}} \left(p_f^t - c_f^t \right) \sum_{j \in \mathcal{F}_f^t} \sum_{h \in \mathcal{H}} \sum_{a \in \mathcal{OA}} n_{ha} \left\{ q_{jha} \frac{\partial s_{jha}}{\partial p_f^v} \right\} \\
&+ \left(p_f^v - c_f^v \right) \sum_{j \in \mathcal{F}_f^v} \sum_{h \in \mathcal{H}} \sum_{a \in \mathcal{OA}} n_{ha} \left\{ q_{jha} \frac{\partial s_{jha}}{\partial p_f^v} + s_{jha} \frac{\partial q_{jha}}{\partial p_f^v} \right\}.
\end{aligned}$$

This equation shows that prices affect profits through the price charged, the relative desirability of other stores, and the change in demand at each store. To recover marginal costs, rewrite the equations in matrix notation:[29]

$$\begin{aligned}
0 &= D_f + G_f \left(p_f - c_f \right) \\
D_{f,(v)} &= \sum_{j \in \mathcal{F}_f^v} \sum_{h \in \mathcal{H}} \sum_{a \in \mathcal{OA}} n_{ha} q_{jha} s_{jha} \\
G_{f,(v,t)} &= \sum_{j \in \mathcal{F}_f^t} \sum_{h \in \mathcal{H}} \sum_{a \in \mathcal{OA}} n_{ha} \left\{ q_{jha} \frac{\partial s_{jha}}{\partial p_f^v} + s_{jha} \frac{\partial q_{jha}}{\partial p_f^v} \mathbb{I} \left[t = v \right] \right\}.
\end{aligned}$$

The subscripts denote the element in the vector or matrix. The vector D_f is the total demand at each fascia. G_f is a matrix of changes in demand at the margin. The cross-partial for market share in these terms is:

[29] $\mathbb{I}[b]$ is the indicator function which is 0 if b is false and 1 if b is true.

$$\frac{\partial s_{jha}}{\partial p_f^v} = \begin{cases} s_{jha}\left(\frac{\partial V_{jha}}{\partial p_f^v} - \sum_{m\in\mathcal{F}_f^v} s_{mha}\frac{\partial V_{mha}}{\partial p_f^v}\right) & j \in \mathcal{F}_f^v \\ -s_{jha}\sum_{m\in\mathcal{F}_f^v} s_{mha}\frac{\partial V_{mha}}{\partial p_f^v} & j \in \mathcal{F}_f^t, t \neq v \\ 0 & \text{otherwise} \end{cases}$$

After inverting the FOCs, the marginal costs are

$$c_f = p_f + G_f^{-1} D_f.$$

Conversely, given a set of marginal costs, FOCs can be solved for a new equilibrium price vector under different market structures.

4.6.2 *Results*

I now discuss the results on marginal costs, price-cost margins, and profits. I focus on how results vary by either GOR or household composition, depending on which is more important economically.[30]

4.6.2.1 *Marginal Costs*

According to Smith [2004], the UK supermarket industry has a simple cost structure which consists of purchasing goods, distribution, labor, and store operations. He argues that the first three are marginal costs, based on research by the UK Competition Commission [2000]. Like Smith, I do not observe advertising, headquarters overhead, and other firm-level costs so I assume they are fixed.

Because the price index varies by month, region, and fascia, I recover marginal costs for each month, region, and fascia from the FOCs, as explained in Section 4.6.1.3. The discussion of results focuses on May 2004 because the same patterns persist across periods, although results for other months show some small variation in costs. Marginal costs can only be recovered in GOR where a fascia operates: for example, I do not observe any Budgens or Waitrose stores in the North East

[30]Most results in this paper can be viewed by fascia, household composition, or GOR. To facilitate comprehension, I aggregated or averaged along one of these dimensions. A detailed appendix with results along all three of these dimensions is available upon request.

and, consequently, cannot recover their marginal costs in this region. In these cases, the marginal cost is coded as Not a Number (NaN) or 'Missing' to ensure that these cases are handled correctly in subsequent computations.

Cost recovery depends on the assumption that firms play a Bertrand-Nash pricing game for multi-product firms, where each fascia functions as a product because of the regional pricing assumption.[31] I assume that Sainsbury's and Tesco are the only firms which operate multiple fascia (brands), which is a reasonable approximation of actual behavior because most firms focus on a specific market segment.[32]

Table 36 reports population-weighted marginal costs with multi-fascia effects averaged across household type. Allowing for the multi-fascia effect lowered marginal costs by 0.7 to 7 percent for Sainsbury's and Tesco's fascias, depending on household type and region. Rival fascias have the same marginal costs in either case. Operating multiple fascias increases Sainsbury's and Tesco's market power because some of the demand lost from a price increase at one fascia is captured at their other fascia. Ignoring this strategic effect overstates Sainsbury's and Tesco's true costs.

Smith [2004] also found that failure to consider multi-fascia effects leads to incorrect marginal costs. My recovered marginal costs also agree with the marketing literature [Seth and Randall, 2001]: Aldi and Netto pursue a deep discount strategy and are the low cost leaders; Iceland and Kwik Save also offer low costs and target working and lower-middle class households; Asda, Tesco, Morrissons-Safeway, and Sainsbury's – listed in order of increasing marginal costs – occupy the middle ground; Tesco's convenience store fascia enjoys a cost advantage over Sainsbury's; and, lastly, Marks & Spencer and Waitrose have significantly higher costs than all other fascia, reflecting their emphasis on higher quality. Surprisingly, Lidl's marginal costs are much higher than Aldi's and Netto's, driven by Lidl's much higher prices. In addition, fascia with higher marginal costs charge higher prices. Each fascia also has the same ordering for the costs of serving different household types. These trends persist when viewed across

[31] In the most general sense, each individual store is a product. Because firms pursue a regional pricing strategy, they set one price per fascia-household type in each GOR. When price flexing occurs, each store functions as a product.

[32] For example, Tesco's fascias include Tesco Express, Tesco Metro, Tesco, and Tesco Extra. The assignment of other firms' stores to one fascia is based on the fact that only Sainsbury's and Tesco focus on both the supermarket and convenience store formats.

either household types or region.[33] Marginal costs vary by household type because the baskets for different types vary in size and composition. Geographical differences, such as distance to distribution centers, density, wages, or congestion, also cause regional variations in marginal costs. For example, average marginal costs are consistently higher for the East Midlands, London, the South East, and Wales for all fascia.

4.6.2.2 *Price-Cost Margin*

From the marginal costs and price indexes, I calculate the price-cost margin, $PCM = (p - c) / p$, for each fascia by household composition (See Table 37.). Price-cost margins vary primarily by household type and not by region. These margins also agree with the marketing literature. Of the big four, Asda, Morrisons-Safeway, and Tesco all have margins ranging from 22% to 23% , which are slightly lower than Sainsbury's (PCM from 23% to 24%). Sainsbury's margin for its convenience store format is also higher than Tesco's, though both firms have higher markups for this format than for the standard format. Similarly, Aldi and Netto – the deep discounters – and Iceland and Kwik Save – which target lower income households – have lower PCM. Lastly, Marks & Spencer also has comparably low margins, driven by their extremely high costs.

Margins vary by household composition in a similar way for each fascia: single pensioners, single parents, and pensioner couples are more profitable whereas couples with children, other without children, and other with children are less profitable. Although total marginal costs are lower for serving the former, per capita marginal costs usually are not. Consequently, more profitable household types purchase items with higher margins, such as prepared food.

The margins I calculated agree with those in Smith [2004]'s Table 3, column (i), which computes the revenue minus the cost of purchasing goods divided by revenue. His value for the margin is based on all marginal costs (purchase of goods, distribution and labour) and is smaller. The CC estimated a multinomial choice model and also

[33] I examined the standard deviation of both marginal cost and the price cost margin for each fascia-household type pair across GOR. Both quantiles and histograms indicate that $\widehat{\sigma_{MC}} / MC$ and $\widehat{\sigma_{PCM}} / PCM$ are less than 10% at the 75th (90th) quantile: the distribution is clumped near 0 with a sharply declining right tail. Fascia-household pairs with higher variation are those which are poorly represented in the TNS data, such as the young and pensioners, and fascia with few observations in a GOR, such as Sainsbury's convenience store format.

finds that margins are 20% in concentrated markets and 15% in non-concentrated markets [UK Competition Commission, 2006]. These margins are much higher than the market literature's value of roughly 5%. As discussed at the beginning of this section, there are several fixed costs, such as advertising, store-level overhead, and headquarters costs, which I do not observe and assume are fixed costs. If I could observe these costs and recompute the margins, I would expect to find margins that were closer to those in the marketing literature.

4.6.2.3 *Geographic Distribution of Profits*

Using the geographic distribution of demand, I predict both fascia and store-level profits by region and household composition. I summarize the different fascias' share of profits by GOR in Table 38. Asda, Morrisons, Sainsbury's, and Tesco (AMST) have the largest shares other than Coop and Other, whose apparently large market shares result from aggregation, as discussed in Section 4.3.2. These fascia lack the coordination and integration of AMST. AMST also have particularly large market shares among the most profitable household types (couples, couples with children, others, and others with children) which tend to be larger, both in numbers of adults and children. In addition, Tesco occupies a dominant position followed by Sainsbury's and Asda, who have comparable market shares. Tesco's lead is even stronger when combined with its convenience store fascia: Tesco's small format stores have much more market share, lower costs, and higher margins than Sainsbury's. The combined Morrison-Safeway fascia is a more effective national competitor than these fascia were individually when compared with the counter-factual pre-merger state in Section 4.7: Safeway and Morrisons are often strong where the other is weak. These results also show that Iceland is a strong niche competitor.

Table 38 also shows which fascia have the largest share of profits in each region. Of the big four, Asda is clearly strongest in the North and weaker further South; Morrison-Safeway favors the North East and Yorkshire; Sainsbury's has significant market share across the country, but is concentrated in London and the South; and, Tesco is strong everywhere, but quite dominant in the East, South East, and Wales with market shares there of 19-26%. Other patterns are visible, such as the deep discounters' focus on the North and Midlands, and Waitrose's concentration on the more affluent areas in London and the South East.

There is considerable variation in the profitability of individual stores, as shown in Table 39 which reports summary statistics for total profit by store and by sales area.[34] These statistics show that fascia which operate larger format stores, as expected, have larger total sales per store but that convenience stores and smaller format stores, such as Budgens and Iceland, are much more profitable per sales area. The deep discounters, Aldi and Netto are also quite profitable per sales area: Lidl, however, is not as profitable because of its much higher marginal cost. The profitability of AMST's stores has considerably more variance than most other fascia, probably because AMST operate more extremely large format stores (hypermarkets). For all fascias, there is a long upper tail: some stores are extremely profitable, either because they are well located or have huge sales areas. Also, firms may operate less profitable stores to foreclose entry to rivals, preventing a situation which would be even less profitable.

4.7 MERGER EVALUATION AND POLICY EXPERIMENT

In 2003 the UK Competition Commission approved the acquisition of Safeway by Morrisons instead of Tesco or a handful of other suitors. To evaluate the merger, I compare the pre-merger state with both the actual merger and a counter-factual Tesco-Safeway merger. This comparison uses firm profits, prices, market shares, and compensating variation to quantify changes in consumer welfare and firm profits. The counter-factual Tesco-Safeway merger provides an upper bound on the adverse welfare consequences of the acquisition of Safeway because Tesco has the most market power and is strong in every region.

My method is similar to the method used in Section 4.6.1 to recover marginal costs and compute firm profits. In contrast to other merger evaluations, I only observe consumer shopping decisions after the Morrisons-Safway merger has occurred. Consequently, I must reconstruct firm behavior prior to the merger. Fortunately, the IGD data starts considerably earlier than 2003 and includes observations on which stores operated under the Safeway fascia. I use this data to construct two counter-factual datasets, one for the state before the observed merger and another for after a hypothetical Tesco-Safeway

[34] A store whose sales area was miscoded in the TNS data causes the anomalously large maximum profit for Marks & Spencer in Table 39.

merger. For both datasets, I solve for the Bertrand-Nash equilibrium prices and use them to compare firm and consumer welfare under each scenario.

In each scenario, I assume that marginal costs remain constant for each fascia regardless of the industry structure. The only exception is Safeway whose pre-merger marginal costs are assumed to be the same as Morrisons's and whose post-merger costs are the same as the firm which acquired Safeway (i.e. Morrisons or Tesco).[35] Because Safeway's strategy lacked focus [Seth and Randall, 2001], any suitor would expect to improve the performance of Safeway stores by replacing Safeway's purchasing, distribution, and operations with their own business processes. Consequently, Safeway's pre-merger costs are probably higher than Morrisons's, and this assumption understates the welfare gains from the merger.

4.7.1 *Computation of Welfare*

Unfortunately, the standard, analytic expression for the change in compensating variation does not apply to my model of utility because the marginal utility of income is non-linear in income [Train, 2009, Anderson et al., 1992]. Consequently, I cannot easily compute the expected maximum utility and solve for the compensating variation. However, it is possible to compute bounds on the compensating variation [McFadden, 1999]:

$$\sum_{j \in C_a} s^0_{hja} C_{jj} \leq \mathbb{E}\left[CV_{ha}\right] \leq \sum_{j \in C_a} s^1_{hja} C_{jj},$$

where C_a is the choice set for OA a, C_{jj} is the compensating variation if the consumer chooses store j in both the pre- and post-merger scenarios, $i = 0$ ($i = 1$) indicates the pre-merger (post-merger) scenario, and s^i_{hja} is the probability household h chooses store j under scenario i. The intuition for this formula is that the consumer would need more (less) compensation if he were unable to change his choice of store before (after) the policy change. Let p^0_j and p^1_j be the prices at store j before and after the merger. Then C_{jj} can be calculated by solving for C_{jj} from the definition of compensating variation

[35] In theory, I should be able to recover Safeway's actual marginal costs if I had access to TNS data prior to the merger.

$$V_{hja}\left(y_h, p_j^0\right) = V_{hja}\left(y_h - C_{jj}, p_j^1\right)$$

which yields

$$\begin{aligned} C_{jj} &= y_h - \exp\left\{\left(\frac{p_j^1}{p_j^0}\right)^{\gamma}\left(\log y_h - \alpha_p \log p_j^0 + \alpha_{area} area_j + \xi_h\right)\right. \\ &+ \left.\alpha_p \log p_j^1 - \alpha_{area} area_j - \xi_h\right\}. \end{aligned}$$

Note that $C_{jj} = 0$ if the price does not change. Furthermore, only the utility of consumption affects the willingness-to-pay because utility from location does not change. $C_{jj} < 0$ means that the consumer needs compensation for his lower post-merger utility. Finally, I compute the total lower and upper bounds on compensating variation for each GOR and household type by integrating over the distribution of consumers, income, and the shock ξ_h. For this problem, this method yields tight bounds and eliminates the need for a more complex procedure to calculate the change in consumer welfare.

4.7.2 *Results*

The observed Morrisons-Safeway merger and counter-factual Tesco-Safeway merger cause little change in consumer welfare and firm profits vis-a-vis the pre-merger state. Both mergers caused small changes in prices, profits, market shares, and compensating variation, although these effects are larger for the Tesco-Safeway merger which is also worse for consumers.

Table 40 compares the changes in market shares, Δs, profits, $\Delta \pi$, and prices, $\Delta price$, where the subscripts refer to whether Morrisons or Tesco acquired Safeway. The first column, s_{pre}, provides pre-merger market shares for comparison. The changes in profits and prices are computed via population-weighted averages across GOR and household composition. I compare the total pre-merger market share and profits of the acquiring firm and Safeway with the post-merger values for the combined firm. The change in market shares for both mergers was small – less than 0.02% for Morrisons and 0.35% for Tesco. The lack of geographic overlap between Morrisons and Safeway explains the small magnitude of the change, especially given the assumption that Safeway's per-merger marginal costs were the same as Morrisons. If Safeway's true costs were larger, then the change in market share should be larger because the true pre-merger price would be higher. The merger should also produce a larger increase in consumer welfare from the lower prices associated with the combined firm's cost savings.

When Safeway merges with either Morrisons or Tesco the change in profits is also small: 0.07% and 0.67% of pre-merger profits, respectively. Almost all firms increase their profits after either merger, although by less than 1%. Most firms can increase their profits and prices much more under the Tesco-Safeway merger because Tesco's market power enables it to raise prices more than Morrisons and prices are strategic complements under Bertrand-Nash competition. The main exception is Waitrose, whose profits decrease by about 1% because Waitrose decreases its price by 0.25% or more in response to these mergers, probably because of increased competition in the South and London where the majority of Waitrose's customers live. In addition, the multi-fascia strategic effect propagates the changes in Sainsbury's and Tesco's prices from their supermarket fascia to their convenience store fascia, causing price increases of about 1%.

The bounds on consumer-willingness-to-pay are summarized by region and household type in Table 41 and Table 42. These tables show the upper and lower bounds on compensating variation, CV_{low} and CV_{upper}, for three scenarios. The labels 'Pre', 'Morr', and 'Tesco' in the column headings indicate which scenarios are being compared: pre-merger, after the Morrisons-Safeway merger, and after the counterfactual Tesco-Safeway merger.[36] These bounds are usually quite tight – i.e. less than 10% – and, consequently, provide a good estimate of consumer welfare.

When aggregated to the national level or by household composition, these mergers appear to decrease consumer welfare. The largest negative welfare changes appear to be for couples, couples with children, and others with children. However, these categories tend to be relatively more numerous. There is considerable regional variation and both mergers improve welfare in some regions. For example, the Morrisons-Safeway merger reduces consumer welfare by at most -0.12% of pre-merger profits, but improves welfare in the East Midlands,

[36]The change in compensating variation for Morr vs. Tesco is not just the difference between Pre vs. Tesco and Pre vs. Morr because the marginal utility is non-linear in income.

London, and the South West by providing stronger competition with Sainsbury's and Tesco. Similarly, the Tesco-Safeway merger causes at most a -0.54% decrease in consumer welfare, but produces positive compensating variation in Yorkshire and the East Midlands. The small magnitude of these changes could be caused either by an extremely competitive market where there is sufficient competition to prevent higher prices or by a situation where each store is effectively a local monopolist. Furthermore, regional pricing may be an effective strategy for tacit collusion to sustain higher prices. Any of these causes could prevent a merger in the supermarket industry from significantly changing the nature of competition.

I also computed the compensating variation for the welfare change between Morrisons and Tesco acquiring Safeway. The results show that acquisition by Tesco would be worse for consumer types when aggregated at the national level, though consumers in the Northwest, Yorkshire, and the West Midlands would be slightly better off. These regions correspond to areas where Morrisons has traditionally been stronger and Tesco relatively weaker.

Overall, my results support the UK Competition Commission's decision to approve Morrisons's acquisition of Safeway because it has little effect on firm profits and consumer welfare – certainly significantly less than a counter-factual merger between Tesco and Safeway. In the long run, Morrisons-Safeway could have developed into another strong national competitor to Asda, Sainsbury's, and Tesco if Morrisons had not had such difficulty integrating Safeway's operations. Moreover, the CC's requirement that Morrisons divest itself of some stores where local competitiveness could be affected further minimized potential adverse consequences of this merger.

4.8 CONCLUSION

This chapter evaluates the impact of Morrisons's acquisition of Safeway on consumer welfare and firm profits. By combining estimates for the demand system with census data, I compute changes in prices, expected profits, and consumer-willingness-to-pay for both the observed merger and a counter-factual Tesco-Safeway merger. Because the data is highly disaggregate these calculations provide information about the nature of demand and competition at even the OA and store-level. The results show that the Morrisons-Safeway merger, which the CC approved in 2003, had little impact on consumer welfare. A counterfactual merger with Tesco would have been an order of magnitude worse for consumers, though still have had a small impact.

These results depend on several assumptions, but especially that locations are fixed exogenously. Adding a stage to the model where fascia choose locations could affect results. In addition, the unobserved quality of locations [Orhun, 2005], scale economies, and network effects might also influence the results [Holmes, 2011, Jia, 2008].

Unfortunately, the current model cannot distinguish between a competitive equilibrium and tacit collusion, both of which could cause small changes in welfare. Using the method developed in this paper, I can, in theory, compute the price equilibrium if firms pursue price flexing instead of regional pricing and how these pricing strategies affect welfare. From a policy point of view, it is also worth investigating whether regional pricing facilitates collusion.

Future work could use this geographical distribution of demand to investigate market definition, whether chaining effects propagate pricing pressure across catchment areas, the local market power of different stores, and whether regional pricing facilitates collusion.

	1 Young	1 Pen	1 Parent	Couple	Pen Couple	Coup child	Other no	Other with
β_{dist}	-23.186	-23.6066	-23.5437	-20.7668	-21.2056	-21.7874	-22.1953	-22.2463
s.e.	(0.5452)	(0.5335)	(0.7623)	(0.4082)	(0.4433)	(0.3423)	(0.365)	(0.477)
$\beta_{carpark}$	8.1156	5.6679	10.4491	8.8168	8.4139	8.3756	7.4687	5.8252
s.e.	(0.7079)	(0.6385)	(0.9534)	(0.7159)	(0.5463)	(0.447)	(0.46)	(0.6318)
α_{price}	-1.2631	-1.2524	-1.2716	-1.3631	-1.1699	-1.5193	-1.4123	-1.4281
s.e.	(0.0508)	(0.0547)	(0.065)	(0.0487)	(0.0509)	(0.0467)	(0.0469)	(0.0582)
α_{area}	3.9843	4.2313	4.0631	3.2774	2.8035	4.621	4.4084	5.509
s.e.	(0.4316)	(0.4749)	(0.6907)	(0.2593)	(0.3406)	(0.3017)	(0.3117)	(0.4575)
γ_{price}	0.2624	0.2711	0.2899	0.3343	0.2902	0.3982	0.3586	0.3814
s.e.	(0.0108)	(0.0122)	(0.0157)	(0.0123)	(0.0131)	(0.0132)	(0.0126)	(0.0171)
σ^2_{ξ}	0.4132	0.3421	0.3622	0.3021	0.2455	0.3364	0.348	0.3685
s.e.	(0.0144)	(0.0118)	(0.0181)	(0.0091)	(0.0081)	(0.0081)	(0.0087)	(0.012)
μ	5.3991	4.0861	3.4768	6.2805	4.1435	5.0852	5.5838	5.0072
s.e.	(0.6514)	(0.5439)	(0.6675)	(0.6096)	(0.5463)	(0.3787)	(0.4402)	(0.4737)

Table 34: Point Estimates

	Aldi	Asda	Budg	Coop	Icel	Kwik	Lidl	MandS	Netto	Other	Morr	Sain	SainS	Some	Tesc	TescS	Wait
Aldi	-5.25	0.38	0.02	0.15	0.24	0.35	0.04	0.02	0.12	0.24	0.23	0.19	0.00	0.09	0.36	0.06	0.03
Asda	0.20	-4.73	0.03	0.15	0.23	0.26	0.04	0.02	0.09	0.22	0.20	0.22	0.01	0.09	0.37	0.06	0.04
Budg	0.07	0.19	-4.91	0.13	0.26	0.06	0.02	0.03	0.03	0.31	0.14	0.30	0.02	0.11	0.52	0.09	0.13
Coop	0.14	0.22	0.03	-4.33	0.19	0.21	0.03	0.03	0.07	0.15	0.18	0.19	0.00	0.10	0.32	0.05	0.03
Icel	0.17	0.30	0.04	0.14	-5.00	0.23	0.05	0.03	0.09	0.27	0.19	0.26	0.02	0.13	0.40	0.07	0.06
Kwik	0.27	0.34	0.02	0.19	0.24	-4.81	0.04	0.01	0.12	0.23	0.23	0.16	0.00	0.09	0.30	0.05	0.01
Lidl	0.14	0.26	0.02	0.13	0.24	0.19	-4.65	0.03	0.06	0.23	0.18	0.19	0.01	0.15	0.33	0.06	0.07
MandS	0.13	0.25	0.04	0.13	0.24	0.15	0.04	-5.54	0.06	0.26	0.19	0.24	0.02	0.10	0.36	0.07	0.08
Netto	0.23	0.29	0.01	0.16	0.20	0.30	0.03	0.01	-5.04	0.23	0.25	0.18	0.01	0.09	0.31	0.05	0.01
Other	0.12	0.21	0.04	0.09	0.21	0.17	0.04	0.03	0.07	-4.22	0.18	0.20	0.02	0.11	0.34	0.06	0.03
Morr	0.18	0.28	0.03	0.15	0.23	0.21	0.04	0.03	0.11	0.26	-4.68	0.21	0.01	0.12	0.36	0.06	0.04
Sain	0.12	0.27	0.05	0.12	0.25	0.14	0.04	0.04	0.06	0.28	0.17	-4.62	0.02	0.13	0.43	0.08	0.07
SainS	0.04	0.15	0.05	0.07	0.34	0.05	0.03	0.06	0.08	0.45	0.19	0.32	-4.79	0.14	0.32	0.12	0.05
Some	0.13	0.24	0.04	0.15	0.27	0.16	0.05	0.03	0.07	0.29	0.20	0.25	0.02	-4.76	0.41	0.07	0.04
Tesc	0.14	0.27	0.05	0.16	0.24	0.17	0.04	0.03	0.07	0.26	0.19	0.25	0.01	0.13	-4.50	0.05	0.05
TescS	0.12	0.25	0.05	0.13	0.28	0.14	0.05	0.05	0.07	0.31	0.18	0.29	0.03	0.15	0.40	-4.90	0.06
Wait	0.06	0.20	0.06	0.11	0.24	0.04	0.04	0.07	0.03	0.29	0.17	0.29	0.02	0.12	0.44	0.08	-4.84

Table 35: Population-Weighted Average Price Elasticities

	NE	NW	York	E Mid	W Mid	E Eng	Lon	SE	SW	Wales
Aldi	0.59	0.59	0.59	0.64	0.59	0.60	0.63	0.62	0.60	0.63
Asda	0.71	0.72	0.71	0.76	0.70	0.70	0.75	0.76	0.70	0.76
Budg	−−	−−	0.70	0.71	0.69	0.69	0.72	0.71	0.69	−−
Coop	0.79	0.82	0.79	0.85	0.78	0.75	0.82	0.84	0.73	0.83
Icel	0.64	0.64	0.64	0.65	0.62	0.63	0.66	0.65	0.62	0.66
Kwik	0.63	0.63	0.63	0.68	0.63	0.64	0.68	0.66	0.63	0.68
Lidl	0.79	0.80	0.79	0.80	0.77	0.77	0.79	0.79	0.76	0.86
MandS	1.07	1.08	1.10	1.09	1.02	1.01	1.06	1.05	1.12	1.14
Netto	0.58	0.60	0.60	0.62	0.57	0.57	0.60	0.64	−−	0.63
Other	0.67	0.74	0.72	0.79	0.69	0.69	0.74	0.77	0.69	0.75
Morr	0.75	0.78	0.76	0.82	0.75	0.74	0.80	0.83	0.74	0.82
Sain	0.79	0.80	0.79	0.85	0.78	0.77	0.82	0.86	0.77	0.83
SainS	−−	0.79	0.77	−−	0.74	0.70	0.80	0.83	0.71	0.73
Some	0.74	0.74	0.73	0.79	0.73	0.72	0.77	0.76	0.72	0.78
Tesc	0.74	0.75	0.75	0.80	0.73	0.73	0.78	0.80	0.73	0.78
TescS	0.70	0.71	0.70	0.76	0.70	0.69	0.73	0.72	0.70	0.75
Wait	−−	0.95	0.94	1.01	0.91	0.90	0.96	1.01	0.90	1.03

Table 36: Average Marginal Costs with Multi-Fascia Effects by GOR

	1 Young	1 Pen	1 Parent	Couple	Pen Couple	Coup child	Other no	Other with
Aldi	0.23	0.26	0.27	0.18	0.26	0.17	0.17	0.17
Asda	0.24	0.27	0.28	0.20	0.27	0.18	0.19	0.19
Budg	0.24	0.27	0.28	0.20	0.28	0.19	0.19	0.20
Coop	0.27	0.30	0.29	0.22	0.30	0.21	0.22	0.22
Icel	0.23	0.27	0.28	0.19	0.27	0.18	0.19	0.19
Kwik	0.23	0.27	0.28	0.19	0.27	0.18	0.18	0.19
Lidl	0.25	0.28	0.28	0.21	0.28	0.20	0.21	0.20
MandS	0.22	0.25	0.25	0.18	0.26	0.14	0.17	0.16
Netto	0.22	0.26	0.27	0.18	0.26	0.17	0.17	0.17
Other	0.27	0.30	0.30	0.23	0.30	0.21	0.22	0.22
Morr	0.25	0.28	0.28	0.21	0.28	0.20	0.20	0.20
Sain	0.25	0.28	0.29	0.21	0.28	0.20	0.21	0.20
SainS	0.28	0.31	0.31	0.24	0.31	0.23	0.24	0.23
Some	0.25	0.28	0.29	0.21	0.28	0.20	0.20	0.20
Tesc	0.25	0.28	0.29	0.20	0.28	0.19	0.20	0.20
TescS	0.26	0.29	0.30	0.22	0.29	0.21	0.21	0.22
Wait	0.24	0.27	0.28	0.21	0.27	0.19	0.20	0.20

Table 37: Average Price-Cost Margins with Multi-Fascia Effects by Household Composition

	NE	NW	York	E Mid	W Mid	E Eng	Lon	SE	SW	Wales	Total
Aldi	9.51	9.95	4.81	5.49	6.64	2.96	1.02	1.33	2.96	6.50	4.41
Asda	14.49	14.68	11.31	8.38	9.48	9.11	4.80	7.51	9.86	8.43	9.33
Budg	0.00	0.00	0.20	1.31	0.18	3.34	1.57	2.20	0.46	0.00	1.15
Coop	9.39	8.43	10.48	15.51	8.05	7.64	5.98	9.05	8.99	8.67	8.85
Icel	7.30	7.11	4.15	7.09	6.27	6.71	10.89	8.40	7.28	9.37	7.64
Kwik	13.34	14.75	10.50	9.90	12.32	0.24	1.22	0.52	2.14	16.90	6.56
Lidl	1.85	1.15	1.30	2.54	1.00	1.49	1.88	1.65	3.31	3.65	1.87
MandS	0.55	0.53	0.54	0.42	0.59	0.77	2.15	0.99	6.02	0.55	1.45
Morr	15.84	6.95	15.17	7.81	9.51	4.39	6.25	6.12	7.17	5.50	7.79
Netto	5.83	2.71	8.60	3.30	1.66	1.37	1.60	0.60	0.00	0.19	2.23
Other	3.96	12.10	14.09	12.95	19.22	15.11	20.19	14.53	16.31	10.90	14.96
Sain	6.13	6.57	5.21	5.97	7.98	11.50	14.94	13.13	8.61	2.62	9.38
SainS	0.00	0.12	0.09	0.00	0.25	0.05	2.40	0.55	0.21	0.08	0.52
Some	2.03	2.03	2.42	4.32	4.74	5.07	4.21	5.47	10.05	6.02	4.77
Tesc	8.84	10.59	9.77	12.76	9.89	25.64	12.22	19.58	12.80	18.75	14.62
TescS	0.94	2.27	1.27	1.92	1.96	2.94	3.94	3.43	2.65	1.58	2.58
Wait	0.00	0.06	0.10	0.32	0.25	1.68	4.73	4.93	1.19	0.30	1.89
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 38: Share of profits by GOR (Percentage): Morrisons-Safeway Merger

	Profit per Store				Profit per Sales Area				
	Mean	Median	Std.Dev	Max	Mean	Median	Std.Dev	Max	N
Aldi	8.78	8.22	3.34	42.22	0.0011	0.0011	0.0005	0.0070	303
Asda	23.24	19.06	16.46	142.62	0.0005	0.0004	0.0002	0.0015	242
Budg	5.01	4.89	2.15	12.26	0.0013	0.0008	0.0013	0.0074	138
Coop	2.82	2.56	1.73	11.64	0.0005	0.0003	0.0005	0.0036	1891
Icel	6.41	5.99	2.35	16.70	0.0014	0.0012	0.0006	0.0049	719
Kwik	6.49	6.11	2.45	24.07	0.0009	0.0009	0.0005	0.0061	610
Lidl	3.24	3.06	1.10	6.18	0.0003	0.0003	0.0001	0.0009	349
MandS	2.69	1.20	19.62	354.27	0.0001	0.0001	0.0001	0.0005	325
Morr	8.52	7.59	5.11	43.37	0.0004	0.0003	0.0004	0.0037	551
Netto	8.10	7.42	3.03	18.61	0.0013	0.0012	0.0005	0.0036	166
Other	3.60	3.26	2.11	16.09	0.0011	0.0009	0.0008	0.0060	2502
Sain	13.31	10.11	13.70	124.73	0.0003	0.0003	0.0002	0.0035	425
SainS	4.37	3.98	2.54	15.04	0.0012	0.0013	0.0007	0.0033	72
Some	5.05	4.50	2.52	24.47	0.0006	0.0005	0.0005	0.0039	570
Tesc	17.85	12.96	16.82	155.75	0.0004	0.0004	0.0002	0.0016	494
TescS	5.83	4.76	5.68	63.82	0.0014	0.0012	0.0010	0.0045	267
Wait	7.08	2.98	26.40	276.00	0.0004	0.0002	0.0013	0.0133	161

Table 39: Store profits: England + Wales

	s_{pre}	Δs_{Morr}	Δs_{Tesco}	$\Delta\pi_{Morr}$	$\Delta\pi_{Tesco}$	$\Delta price_{Morr}$	$\Delta price_{Tesco}$
Aldi	4.41	0.00	−0.02	0.17	0.26	0.02	0.10
Asda	9.32	0.01	−0.03	0.13	0.37	0.04	0.14
Budg	1.15	−0.00	−0.00	0.06	0.42	0.02	0.14
Coop	8.85	0.00	−0.04	0.09	0.23	0.02	0.10
Icel	7.64	0.00	−0.03	0.07	0.30	0.03	0.14
Kwik	6.55	0.01	−0.03	0.20	0.18	0.02	0.18
Lidl	1.87	−0.00	−0.01	0.06	0.28	0.01	0.10
MandS	1.45	−0.00	−0.01	0.00	0.19	0.03	0.13
Morr	2.49	0.01	−0.01	0.13	0.10	0.26	0.10
Netto	2.23	0.00	−0.01	0.14	0.13	−0.16	0.08
Other	14.96	−0.00	−0.06	0.04	0.25	−0.00	0.11
Safe	5.29	−−	−−	−−	−−	−−	−−
Sain	9.38	−0.00	−0.03	0.04	0.39	−0.01	0.14
SainS	0.52	−0.00	−0.00	−0.06	0.09	0.30	0.85
Some	4.78	−0.00	−0.02	0.02	0.35	−0.01	0.14
Tesc	14.62	0.00	0.35	0.08	2.44	0.09	0.59
TescS	2.58	−0.00	−0.03	0.06	−0.36	−0.02	1.20
Wait	1.91	−0.02	−0.02	−0.96	−0.64	−0.37	−0.25

Table 40: Percent change in market shares, profits, and price post-merger.

	Pre vs. Morr		Pre vs. Tesco		Morr vs. Tesco	
	CV_{low}	CV_{upper}	CV_{low}	CV_{upper}	CV_{low}	CV_{upper}
NE	−22.21	−21.90	−24.68	−23.37	−2.66	−1.45
NW	−8.20	−8.08	−2.13	−0.16	6.04	7.93
York	−21.75	−21.36	8.29	10.47	29.94	31.85
E Mid	14.17	14.41	9.20	11.58	−5.34	−2.66
W Mid	−40.53	−39.88	−30.57	−28.33	9.07	11.89
E Eng	−29.45	−28.93	−79.60	−77.08	−50.11	−48.18
Lon	43.46	44.83	−10.16	−4.50	−53.36	−49.53
SE	−11.33	−11.24	−91.19	−85.74	−79.92	−74.45
SW	15.34	15.99	−74.71	−72.53	−90.83	−88.28
Wales	−8.96	−8.26	−28.14	−26.89	−19.85	−18.38
Total	−69.47	−64.42	−323.70	−296.55	−257.00	−231.25

Table 41: Bounds on expected compensating variation by GOR

	Pre vs. Morr		Pre vs. Tesco		Morr vs. Tesco	
	CV_{low}	CV_{upper}	CV_{low}	CV_{upper}	CV_{low}	CV_{upper}
1 Young	−9.04	−8.63	−11.49	−8.44	−2.72	0.29
1 Pen	−6.99	−6.72	−18.06	−16.42	−11.21	−9.69
1 Parent	−0.79	−0.53	−20.15	−18.80	−19.49	−18.24
Couple	−12.26	−11.29	−44.69	−37.69	−33.00	−26.19
Pen Couple	−4.68	−4.36	−14.27	−11.92	−9.76	−7.53
Coup child	−22.33	−20.76	−141.97	−136.37	−120.41	−115.38
Other no	−10.23	−9.21	−57.03	−51.61	−47.39	−42.19
Other with	−3.14	−2.93	−16.05	−15.29	−13.02	−12.32
Total	−69.47	−64.42	−323.70	−296.55	−257.00	−231.25

Table 42: Bounds on expected compensating variation by household composition

Part III

APPENDIX

A

COMPUTATIONAL AND NUMERICAL ISSUES IN ESTIMATING THE BLP MODEL

This appendix provides additional numerical and computational details about how to estimate the BLP model and perform large-scale Monte Carlo simulations. First, I explain how to implement BLP as an MPEC and then discuss how to run large-scale simulations using the PBS job manager.

A.1 SNOPT FORMULATION OF BLP MPEC

The most reliable way to estimate the BLP model is to specify the problem as an MPEC and use a modern solver such as SNOPT [Su and Judd, 2010, Dubé et al., 2011]. The SNOPT interface formulates the optimization problem as

$$\begin{array}{rl} \min\limits_{x} & F_0(x) \\ \text{subject to} & l_x \le x \le u_x \\ & l_F \le F(x) \le u_F \end{array}$$

where

$$\begin{aligned} F_0(x) &= F_{objective}(x) \\ F_i(x) &= constraint_i(x). \end{aligned}$$

$F(\cdot)$ is a vector function whose first element is the value of the objective function and whose other elements are the constraints. l_x and u_x specify the upper and lower box constraints on x. Box constraints speed up optimization by preventing the solver from considering clearly infeasible regions of the parameter space. l_F and u_F are the upper and lower bounds on the objective function and non-linear constraints. For the BLP MPEC problem, the vector

$$x = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \delta_{jt} \\ \eta \end{pmatrix}.$$

The gradient of the objective function and the Jacobian of the constraints are specified in a matrix $G(x)$ where

$$G_i(x) = \nabla_x F_i(x),$$

i.e.

$$G_{ij}(x) = \frac{\partial F_i(x)}{\partial x_j}.$$

Thus, for the BLP MPEC,

$$F(x) = \begin{pmatrix} \eta' W \eta \\ S_{jt}^{observed} = s_{jt}\left(\delta\left(x, \xi; \theta_1\right); \theta_2\right) \\ \eta - Z'\left(\delta - X\theta_1\right) \end{pmatrix}$$

and

$$\begin{aligned} l_x &= \begin{pmatrix} -100 & -50 & -10^8 & -10^8 \end{pmatrix}' \\ u_x &= \begin{pmatrix} 100 & 50 & 10^8 & 10^8 \end{pmatrix}' \\ l_F &= \begin{pmatrix} -10 & \log\left(S_{jt}^{observed}\right) & 0 \end{pmatrix}' \\ u_F &= \begin{pmatrix} 10^8 & \log\left(S_{jt}^{observed}\right) & 0 \end{pmatrix}'. \end{aligned}$$

These box constraints rule out extreme parameter values to improve solver convergence. Although the objective function is positive semi-definite, the lower bound on F_0 is set to -10.0 to force the solver to do its best. $l_F = u_F$ for both the market share constraints and moment conditions because these constraints are equality constraints.

Similarly, the Jacobian G is

$$G(x) = \begin{pmatrix} 0 & 0 & 0 & 2W\eta \\ 0 & \dfrac{\partial s(\delta;\theta_2)}{\partial\theta_2} & \dfrac{\partial s(\delta;\theta_2)}{\partial\delta} & 0 \\ Z'X & 0 & -Z' & I \end{pmatrix}.$$

The first row is actually the gradient of the objective function for the SNOPT interface. This interface makes it easier to specify the sparseness structure of the problem more economically when implementing the estimation software.

A.2 Example PBS script for estimating BLP

Listing 2 shows a PBS script which executes 10 BLP estimations per job using a parameter sweep. Chunking faster jobs together is an important technique for minimizing scheduler overhead and increasing throughput. The script uses the bash shell and special comments which start with `#PBS` to specify options to the job manager, such as limits on wall-clock time for the job (12 hours in this case) or the indexes for instances of this job (specified with the `-t` option). Each instance of this job is passed a unique index between 0 and 499, inclusive, via the environmental variable `PBS_ARRAYID`. I use this variable to determine which replication and starts to process. The script abstracts as much state information as possible – the number of markets, T, the number of products, J, the type of instrumental variables, etc. – in other environmental variables to make it easier to modify this script for other experiments as well as to compare different runs.

Listing 2: Example PBS script for estimating 10 BLP starts per job.

```
#!/bin/bash
#PBS -l nodes=1:ppn=1,walltime=12:00:00
#PBS -q long
#PBS -t 0-499
#PBS -m n

#------------------------------------------------------------
# Purpose: run one BLP start at a time by converting the
# parameter sweep index to a start and replication.
#------------------------------------------------------------
```

```
# Global definitions to describe this run
export N_STARTS_PER_JOB=10
export N_STARTS=50
export IVType=Char
export T=25
export J=100

export PROJ_DIR=/gpfs/pads/projects/CI-SES000069
export BASE_DIR=${PROJ_DIR}/data/blp/arm.T${T}.J${J}

# Ensure that GFORTRAN output occurs as written instead of
# all C/C++ followed by FORTRAN
export GFORTRAN_UNBUFFERED_ALL='y'


#-------------------------------------------------------------
# Regular logic : run ten estimations per job
#-------------------------------------------------------------
ix=${PBS_ARRAYID}
ixBegin=$(( ix * N_STARTS_PER_JOB ))

for(( ix = ixBegin ; ix < ixBegin + N_STARTS_PER_JOB ; ix++ ))
do

  ixRep=$(( ix / N_STARTS ))
  ixStart=$(( ix % N_STARTS ))

  echo
  echo '>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>'
  echo '====>>>> Commencing ONE BLP Estimation'
  echo '         Replication : ' ${ixRep}
  echo '         Start       : ' ${ixStart}
  echo "         Using ${IVType} instruments"
  echo '<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<'
  echo

  inFile=${BASE_DIR}/rep${ixRep}/config/DFS.${IVType}.config.${
      ixStart}.prop

  ${PROJ_DIR}/sbox/blpcpp/src/blpDriver -f ${inFile}

done
```

B

SUPERMARKET MERGER EVALUATION

This appendix contains supplementary information about using Roy's identity with a price index, numerical issues involved in estimating the model, and the data and how it was cleaned.

B.1 ROY'S IDENTITY AND PRICE INDEXES

Because Roy's identity holds for true prices but not necessarily a price index, treating the price index as the true price can produce an incorrect result when deriving demand. Let the indirect utility be $V_{hj} = V_{hj}\left(p_j, y_h\right)$, where p_j is the price index for a unit of groceries at store j and y_h is household h's expenditure. The true specification of the indirect utility is $V_{hj}^{true} = V_{hj}\left(\breve{p}_j, y_h\right)$ where $\breve{p}_j$ is the vector of all prices of goods at the store. The price index is a function of this vector of prices, $p_j = f\left(\breve{p}_j\right)$. Assume that prices enter the indirect utility via an index restriction. I.e., the price index $p_j = \omega \cdot \breve{p}_j$, where ω is a vector of expenditure shares. Then, $V_{hj}^{true} = V_{hj}\left(\omega \cdot \breve{p}_j, y_y\right)$. If $\breve{p}_{jk}$ is the k-th component of the vector $\breve{p}_j$ and $\breve{q}_{jk}$ k-th component of demand, then Roy's identity implies that

$$\begin{aligned}
\breve{q}_{jk} &= -\frac{\partial V_{hj}^{true}}{\partial \breve{p}_{jk}} / \frac{\partial V_{hj}^{true}}{\partial y_h} \\
&= -\left[\frac{\partial V_{hj}}{\partial p_j}\frac{\partial p_j}{\partial \breve{p}_{jk}}\right] / \frac{\partial V_{hj}}{\partial y_h} \\
&= -\omega_k \frac{\partial V_{hj}}{\partial p_j} / \frac{\partial V_{hj}}{\partial y_h}.
\end{aligned}$$

Aggregating individual expenditures to compute the total units of groceries demanded by a household shopping at store j yields

$$\begin{aligned}
q_j &= \omega \cdot \breve{q}_j \\
&= -\left(\sum_k \omega_k^2\right) \frac{\partial V_{hj}}{\partial p_j} / \frac{\partial V_{hj}}{\partial y_h},
\end{aligned}$$

assuming that the aggregation of demand uses the same weights as the price index. Consequently, the result from Roy's Identity must be scaled by the sum of the squares of the weights on the individual goods (or categories) if the price index is used instead of individual price. If the weights $\omega_k \approx 1/n$ then $\sum_{k=1}^{n} \omega_k^2 \approx 1/n$ which is small because the index is composed of thousands of products. If this factor is not accounted for, the reported parameter estimates will be too small. The scaling factor, μ, in the indirect utility helps account for this scaling difference between the discrete and continuous choice parts of the model so the model provides reasonable predictions of both store choice and expenditure.

B.2 ESTIMATION AND NUMERICAL ISSUES

To calculate reliable point estimates, I use SNOPT 7.0 [Gill et al., 2002], one of the best solvers currently available to maximize the log-likelihood. SNOPT solves sparse, large-scale, constrained optimization problems using a sequential quadratic programming (SQP) algorithm. The SQP algorithm determines the search direction by minimizing a sequence of quadratic sub-problems which approximate the Lagrangian locally, subject to linearized constraints. Although other algorithms, such as the interior point method used in KNITRO [Byrd et al., 2006], are sometimes faster, interior point methods only work well on convex problems. For this problem SNOPT is much more reliable, especially for starting values in some regions of parameter space where the estimated variance of ξ_h is extremely small. SNOPT, KNITRO, and IPOPT [Wachter and Biegler, 2006], a free interior point solver which is similar to KNITRO, produce similar solutions and reliably outperform MATLAB's fmincon.

In addition, I used multiple starting values and computed accurate gradients using Complex Step Differentiation (CSD) [Al-Mohy and Higham, 2009] (See Section B.2.1.). In theory, CSD is more accurate than finite difference methods and often as accurate as analytic derivatives if you take a very small step size. In practice, I suspect that the solver will perform better with analytic gradients.

B.2.1 *Complex Step Differentiation*

CSD uses imaginary numbers to compute derivatives numerically which are more accurate than finite difference methods and often as good as automatic differentiation [Martins et al., 2001], a compiler-like tool which analyzes source code and computes efficient analytic derivatives. CSD calculates gradients numerically by taking a very small step[1] in the complex direction and approximating the gradient as $\frac{\partial f(x)}{\partial x_i} \approx \texttt{Imag}\left[\frac{f(x + ih \cdot e_i)}{h}\right]$ where e_i is a unit vector in the $\underline{i}$-th direction. From the Taylor series expansion of $f(x)$, it is easy to show that the error is $O\left(h^2\right)$ but with a much smaller h than for (two-sided) finite difference [Al-Mohy and Higham, 2009].

B.2.2 *Verification of Optimum*

Because the log-likelihood of a discrete choice model may have multiple local maxima, it is important to ensure that the solver has converged to a local maximum and that it is the global maximum [McFadden, 1984].

To verify that the solver converged to a local optimum, I try multiple starting values, use several solvers (SNOPT, KNITRO, Ipopt), and test my code on a Monte Carlo data set. For all households, the solver's exit codes indicate that it found locally optimal solutions, although not always within the desired tolerance. Numerical difficulties occur when the variance of the household-specific shock, σ_{ξ}^2, is small, which causes numerical truncation in the calculation of ℓ_h^q.[2] I also compared the results from SNOPT 7.0 to those from Ipopt 3.5.5, a modern Interior Point solver [Wachter and Biegler, 2006]. For all households, the two solvers found the same optima to at least four decimal places. In addition, I confirmed that the gradients and Lagrange multipliers were close to 0, that the condition number was small, and that all eigenvalues of the Hessian were positive, indicating that the solver's solutions were local optima.[3]

---

[1] CSD's superior accuracy depends on taking a smaller step size than finite difference methods.

[2] I am grateful to Todd Munson in the MCS division of Argonne National Laboratory for explaining this point to me.

[3] I computed the Hessian using CSD because BFGS approximations to the Hessian often differ considerably from the true value, particularly when the solver converges quickly.

To examine the properties of the solution on a larger scale, I compare the solver's optimum to the value of the log-likelihood at 500 quasi-Monte Carlo Niederreiter points in a hypercube of $\pm\hat{\sigma}$ about the solver's solution.[4] For every household type, the solver's solution was larger than at any of the Niederreiter points, indicating that the solver had probably found the global maximum.

Proving that the local maximum is the global maximum is more difficult. McFadden [1984] discusses functional form restrictions that ensure that a local maximum of a logit model is the global solution. One practical option is to compare the values of the log-likelihood and a quadratic approximation – analogous to that used by most solvers – about the solver's solution on a set of quasi-Monte Carlo points to examine if the quadratic approximation is accurate, if there is consistent bias in one direction, and if the objective function has multiple peaks.[5]

B.3 TNS WORLDPANEL

The TNS Worldpanel is a complex, high quality home scan panel of consumer shopping behavior which was designed for marketing purposes and contains rich detail about both household demographics and product characteristics. Participants scan every grocery item which they purchase and bring into the home using a scanner provided by TNS.[6] The data is sent to TNS electronically and periodically verified using till receipts. If an item lacks an SKU – such as raw fruit and vegetables – then the household enters the price and other details manually. Households typically participate in the panel for about two years and in return receive modest compensation which is designed to avoid influencing consumer spending on groceries [Leicester and Oldfield, 2009a].

Despite the precision of the measurement, there are several potential sources of bias: sampling bias, learning how to scan, attrition, and fatigue. Sampling bias occurs because a disproportionate number of participating households are those in their prime consuming years (couples with and without children). The young (students) and the old

[4] Niederreiter points are superior to standard Monte Carlo draws because they do a better job of covering the hypercube than pseudo-Monte Carlo draws [Judd, 1998].

[5] Ken Judd suggested this technique to me.

[6] E.g., if they purchase lunch at Marks & Spencer and eat it at the office the transaction is not recorded in the TNS data.

(pensioners) are under-represented and often missing completely in some regions. TNS reweights the data to correct for this bias. To avoid learning effects, TNS discards the first two weeks of data from new participants. Also, scanning behavior changes over time because of fatigue or scanning causing households to modify their consumption habits [Leicester and Oldfield, 2009b]. In addition, household types attrit at different rates. Nevertheless, TNS selects households and locations to maintain a representative distribution of consumer types. Leicester and Oldfield [2009b] discusses these issues in more detail as part of a comparison with the benchmark Expenditure and Food Survey (EFS). They conclude that scanner data matches the general trends of the EFS although with about 25% lower recorded expenditures.

Following Beckert et al. [2009], I simplify the complexity of the data by grouping similar fascia together, constructing household types which are based on the census categories for household composition, and constructing a price index for units of a reference basket of groceries. The following subsections examine these decisions in more detail.

B.3.1 *Household Type*

I assign each household in the TNS data a type based on the KS020 Household Composition Table in the UK 2001 Census (See Table 44.).[7] Then, I combine types which differ by only marital status, which should not affect consumption. I assume that each household purchases units of groceries of its type-specific reference basket of groceries at the price specified by the price index as discussed further in Section B.3.3.

B.3.2 *Aggregation of Fascia*

The TNS and IGD data provide much more detailed fascia than can be feasibly used for estimation. Consequently, similar 'sub-fascia' are grouped into 18 broader categories with one fascia for each of the major firms except Sainsbury's and Tesco, which each have an additional fascia type for their convenience store format. In addition, I coalesce the various regional Cooperative movement stores into one

[7] Household type 11 (KS0200015 Other, all pensioner) is not included because there are no observations of this type in the TNS data.

TNS HOUSEHOLD	TNS HH TYPE(S)	KS020 HOUSEHOLD
1 Single Young	2	KS0200003 Single non-pensioner
2 Single Pensioner	1	KS0200002 Single Pensioner
3 Single Parent	7	KS0200011 Single parent, dependent children
4 Childless couple	4	KS0200005 + KS0200008 Couple, no children
5 Pensioner couple	3	KS0200004 Pensioner couple
6 Couple with children	5	KS0200006 + KS0200009 Couple dependent children
7 Others – no children	6, 8,10, 11, 12	KS0200007 + KS0200010 Couple, with non-dependent children ; KS0200012 Single parent non-dependent children ; KS0200013 Other, dependent children ; KS0200014 Other, all student ; KS0200016 Other, other
8 Others – with children	9	KS0200013 Other, with dependent children

Table 44: Conversion of Census KS020 to TNS Household Type

Coop fascia and assign the fringe of regional and independent grocery retailers to the Other fascia.

The firms which are represented by Other typically have only a local presence. In all GOR, 75% or more of Other stores are the size of convenience stores. In some regions, several firms under the Other fascia have have a comparable number of stores to the national players. However, given their consistently small sales areas and regional presence, these stores probably have little market power and lack the cost advantages of the fascia which have more national presence, more efficient supply chains, and economies of scale and scope. Consequently, treating all these fringe stores as one fascia overstates their market power.

Similarly, using a single Coop fascia overstates the market power of the Cooperative movement stores because they lack the focus and centralized coordination of the major fascia.

B.3.3 *Price Index for Aggregate Consumption*

I use the aggregate hedonic price indexes which Beckert et al., 2009 construct by performing hedonic regressions within product categories and then aggregating with category expenditure weights. The hedonic regressions control for regional, fascia, and monthly fixed effects. The price index aggregates bar-code level hedonic price indexes to produce an aggregate index for the price of a unit of groceries. This unit of groceries represents the price of the reference basket of goods which each household type faces in a specific fascia-region-month. The hedonic weights from the index can also be used to impute prices for stores in the consumer's choice set where I do not observe prices because the consumer shops elsewhere.

This aggregation assumes that, conditional on household type, households behavior roughly satisfies the Gorman Polar form – i.e., their preferences are homothetic across different goods, perhaps with some satiation level of spending. The assumption that household expenditure is split across goods in constant fractions, regardless of income, has important implications for the price equilibrium and cost recovery, as discussed in Section 4.6.

The price data consists of prices, goods, product characteristics, quantities, and pack sizes for over 16 million purchases. The variation occurs both in the choice of basket – which is across household type and region – and prices – which is across fascia, regions, and months. Using this data, the price index is constructed as follows:

1. Group similar barcode-level purchase data into specific categories (shampoo, milk, mineral water, etc.) and then compute a hedonic regression

   $$\log p_{bst} = \alpha_f^r + t_t + \alpha^r t_{ft} + r_r + z_b^{'} \beta^r + \epsilon_{bst}$$

   for each category, using the observations at store s, fascia f, barcode (product) b, geographic region g, time t and product category group r. I suppress the subscripts for individual i of household-type h who actually purchased the item on some trip.

2. From the category regression, compute predicted prices for counter factual purchases at all fascia, dates, and regions:

$$\hat{p}_{bst} = \exp\left[\hat{\alpha}_f^r + \hat{t}_t + \hat{\alpha}^r t_{ft} + \hat{r}_r + z_b^{'}\hat{\beta}^r + \hat{\epsilon}_{bst}\right].$$

$\hat{\epsilon}_{bst}$ is the residual for the price which was observed for the item b at a specific fascia-region-date. I assume that this shock is an unobserved product characteristic which would be the same for any household purchasing the same item at a different fascia and date, regardless of household type.

3. Compute the predicted price for fascia-product category-month-region-household type by taking the mean over these factors:

$$\hat{p}_{rfhgt} = \frac{1}{N_{rfhgt}} \sum_{(\tilde{b},\tilde{s},\tilde{t})\in A_{rfhgt}} \hat{p}_{\tilde{b}\tilde{s}\tilde{t}}$$

where N_{rfhgt} is the number of predicted prices for a fascia-product category-household type in a given month and region and A_{rfhgt} is the set of the indices for a given fascia-product category-household type, (b, s, t).

4. Compute the product category expenditure weights for each individual household type by region:

$$w_{rhg} = \sum_{i\in h\cap g}\sum_{b\in r} e_{ib} / \sum_{i\in h\cap g}\sum_{b} e_{ib}$$

where e_{ib} is household i's annual expenditure on product b, $h \cap g$ is the set of households of type h in region g. Thus, w_{rhg} is the observed expenditure weight for all households of type h in region g on product category r. Annual expenditure is either the total expenditure observed per year or the observed expenditure scaled by $12/$ (Months of Data) if there is less than a year of data for a household i. Note: these expenditure weights are assumed to be constant over time.

5. Finally, calculate the expenditure-weighted aggregate hedonic price index (hereafter, the 'price index') for each household type-fascia-region-month

$$P_{fhgt} = \sum_r w_{rhg} \hat{p}_{rfhgt}.$$

Constructing a universal 'unit of groceries' remains a challenge because quality is not the same across fascia or even within fascia. In addition, (unobserved) price tiers for different locations and formats, promotions, stock outages, and seasonal effects further complicate these measurements. Also, non-price competition is important in the UK supermarket industry, especially the variation of the assortment of goods based on local market demographics, such as stocking more organic products in more affluent neighborhoods. As shown below, households do not purchase the same baskets, e.g. some households never purchase alcohol or pet food. Also, the fraction of the basket spent on different kinds of goods changes with income. The price index averages over these different factors. To minimize the impact of measurement error, I include a household shock, ξ_h.

The use of this price index and units of a reference basket of groceries to estimate demand depends on the validity of assumption that consumers, conditional on household type, purchase the same basket of goods at each fascia. The mean, median, and standard deviation of expenditure by household type and fascia on fresh food, prepared food, alcohol and non-food at supermarkets show considerable variation: for example, the standard deviation is often the same order of magnitude as the mean (See, for example, the mean and standard deviation for alcohol in Table 45 and Table 46.).[8] Furthermore, median alcohol expenditure is zero for all household type-fascia pairs and median non-food expenditure is almost always zero. In addition, research for US consumers indicates that they purchase different baskets from different stores, that different fascia are even complementary, and consumers often make more than one stop on a trip [Briesch et al., 2010]. Because consumers often deviate from the reference basket both in quality and composition, consumer expenditure is observed with measurement error. If the shock ξ_h fails to control for this heterogeneity, the model will suffer from endogeneity.

In addition, expenditure and basket composition are correlated with fascia because different fascia target different demographic segments. This correlation could also be caused by TNS's sampling mechanism if household type is correlated with other demographics which predispose a household to choose a specific fascia or basket as well as

[8] Tables for the other expenditure categories are available upon request.

the location where they live. This endogeneity would be even more important to handle in a dynamic model where firms (and households) choose locations.

B.3.4 *Computation of Choice Sets and Distance*

To simplify estimation and welfare calculations, I assume that the choice set of each household or OA consists of the 30 nearest stores. This assumption seems reasonable, based on the statistics in Table 47 which show mean, median, and 75-th percentile distances for the 30 stores closest to each OA by region. Because consumers have high disutility from travel [Smith, 2004], most households are unlike to consider stores outside this choice set unless a particular store is conveniently located or stocks a good which the consumer values highly.

A small fraction of households actually shops at more distant stores; Beckert et al. [2009] drop these observations from the data which could bias the results. The choice set may also overstate demand for convenience stores if consumers are less willing to travel to smaller stores than larger formats. Constructing choices sets of a uniform size does, however, facilitate implementation of the estimation code, and captures the majority of the stores that each household considers.

	1 Young	1 Pen	1 Parent	Couple	Pen Couple	Coup child	Other no	Other with
Aldi	0.0512	0.1296	0.0716	0.0855	0.0936	0.0452	0.1045	0.0207
Asda	0.0658	0.0429	0.0505	0.0801	0.0891	0.0496	0.0660	0.0532
Budg	0.0717	0.0000	0.0000	0.0432	0.0699	0.0511	0.0000	0.2391
Coop	0.0335	0.0437	0.0603	0.0714	0.0574	0.0781	0.0806	0.0519
Icel	0.0258	0.0103	0.0442	0.0588	0.0018	0.0279	0.0153	0.0068
Kwik	0.0993	0.0792	0.0577	0.1074	0.0270	0.0678	0.0804	0.0695
Lidl	0.0564	0.0473	0.0742	0.0422	0.0933	0.0503	0.0604	0.0616
MandS	0.0251	0.0340	0.0155	0.0228	0.0198	0.0243	0.0264	0.0169
Morr	0.0495	0.0592	0.0437	0.0959	0.0756	0.0445	0.0639	0.0461
Netto	0.0524	0.1312	0.0209	0.1131	0.0424	0.0315	0.0721	0.1486
Other	0.0428	0.0000	0.0000	0.0043	0.0182	0.0096	0.0173	0.0008
Sain	0.0630	0.0665	0.0340	0.0921	0.0629	0.0500	0.0708	0.0604
SainS	0.1782	0.0000	0.0000	0.0000	0.0000	0.0413	0.0520	0.0000
Some	0.0446	0.0379	0.0638	0.0662	0.0827	0.0441	0.0445	0.0339
Tesc	0.0617	0.0731	0.0375	0.0921	0.0729	0.0539	0.0610	0.0505
TescS	0.0562	0.0506	0.0257	0.0037	0.0189	0.0346	0.0682	0.0145
Wait	0.0540	0.0642	0.0262	0.0852	0.0336	0.0269	0.0572	0.0719

Table 45: Mean Expenditure on Alcohol by Household Composition

	1 Young	1 Pen	1 Parent	Couple	Pen Couple	Coup child	Other no	Other with
Aldi	0.1169	0.2566	0.1187	0.1690	0.2050	0.1067	0.1768	0.0617
Asda	0.1496	0.1517	0.1181	0.1734	0.1877	0.1300	0.1442	0.1370
Budg	0.1353	0.0000	0.0000	0.1365	0.2614	0.1694	0.0000	0.3112
Coop	0.1105	0.1712	0.1878	0.2049	0.1779	0.2230	0.2101	0.1662
Icel	0.1229	0.0570	0.1140	0.1562	0.0126	0.1140	0.0766	0.0313
Kwik	0.2541	0.2227	0.2010	0.1948	0.1297	0.1902	0.2098	0.1635
Lidl	0.1918	0.1635	0.1669	0.1140	0.2188	0.1374	0.1489	0.1776
MandS	0.1088	0.1669	0.0536	0.0840	0.0773	0.0874	0.1285	0.0533
Morr	0.1424	0.1757	0.1439	0.2116	0.1724	0.1280	0.1509	0.1268
Netto	0.1217	0.2827	0.0437	0.2278	0.0942	0.0815	0.1449	0.2618
Other	0.1940	0.0000	0.0000	0.0473	0.1304	0.0918	0.1276	0.0061
Sain	0.1604	0.1655	0.0799	0.1970	0.1655	0.1246	0.1558	0.1611
SainS	0.3648	0.0000	0.0000	0.0000	0.0000	0.0819	0.1274	
Some	0.1566	0.1434	0.1438	0.1734	0.2289	0.1337	0.1553	0.1102
Tesc	0.1454	0.1762	0.0985	0.1954	0.1693	0.1304	0.1552	0.1264
TescS	0.1606	0.1446	0.0852	0.0200	0.0700	0.0826	0.1980	0.0768
Wait	0.1369	0.1736	0.0787	0.1966	0.1042	0.0888	0.1370	0.1210

Table 46: Std. Dev. of Expenditure on Alcohol by Household Composition

	Mean (km)	Median (km)	Q75 (km)	Std Dev	% > 20 km	N
NE	6.38	5.00	7.63	5.79	3.16	8599
NW	5.06	3.85	5.72	4.04	1.67	22712
York	6.51	4.67	7.32	5.89	3.48	16793
E Mid	7.31	5.16	10.22	5.42	2.82	14107
W Mid	5.94	3.92	7.84	4.92	2.89	17458
E Eng	8.44	7.72	10.49	4.46	2.47	18199
Lon	2.40	2.33	2.87	1.23	0.00	24149
SE	7.17	6.30	9.85	4.12	0.40	26642
SW	9.57	8.17	13.67	6.57	6.95	17013
Wales	10.92	7.58	13.06	9.52	13.98	9766

Table 47: Statistics for Store-OA Distance by GOR

BIBLIOGRAPHY

A.H. Al-Mohy and N.J. Higham. The complex step approximation to the fréchet derivative of a matrix function. 2009.

J.G. Altonji and L.M. Segal. Small-sample bias in gmm estimation of covariance structures. Journal of Business & Economic Statistics, 14 (3):353–366, 1996. ISSN 0735-0015.

S.P. Anderson, A. De Palma, and J.F. Thisse. Discrete choice theory of product differentiation. MIT press, 1992.

T. Armstrong. Large market asymptotics for differentiated product demand estimators with economic models of supply. 2011.

Walter Beckert, Rachel Griffith, and Lars Nesheim. A micro-econometric approach to geographic market definition in local retail markets: Demand side considerations. Draft, 2009.

D.R. Bell and J.M. Lattin. Grocery shopping behavior and consumer response to retailer price format: Why 'large basket' shoppers prefer edlp. Marketing Science, 17(1):66–88, 1998.

Alexandre Belloni and Victor Chernozhukov. High dimensional sparse econometric models: An introduction. In Pierre Alquier, Eric Gautier, and Gilles Stoltz, editors, Inverse Problems and High-Dimensional Estimation, volume 203 of Lecture Notes in Statistics, pages 121–156. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-19989-9.

S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. Econometrica, 63(4):841–890, 1995.

S. Berry, J. Levinsohn, and A. Pakes. Differentiated products demand systems from a combination of micro and macro data: the new car market. Journal of Political Economy, 112(1):68–105, 2004a.

S.T. Berry. Estimating discrete-choice models of product differentiation. The RAND Journal of Economics, pages 242–262, 1994.

S.T. Berry and P.A. Haile. Identification in differentiated products markets using market level data. National Bureau of Economic Research Cambridge, Mass., USA, 2010.

Steve Berry, Oliver B. Linton, and Ariel Pakes. Limit theorems for estimating the parameters of differentiated product demand systems. Review of Economic Studies, 71(3):613–654, 2004b.

Steven Berry and Ariel Pakes. The pure characteristics demand model. International Economic Review, 48(4):1193–1225, 2007.

T.F. Bresnahan, S. Stern, and M. Trajtenberg. Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. The RAND Journal of Economics, 28:17–44, 1997. ISSN 0741-6261.

Richard A. Briesch, Edward Fox, Joon Park, and William R. Dillon. Destination categories and store choice: The role of assortment. 2010.

Robert Bringhurst. The Elements of Typographic Style. Version 2.5. Hartley & Marks, Publishers, Point Roberts, WA, USA, 2002.

R. Byrd, J. Nocedal, and R. Waltz. Knitro: An integrated package for nonlinear optimization. Large-scale nonlinear optimization, pages 35–59, 2006.

X. Chen. Large sample sieve estimation of semi-nonparametric models. Handbook of Econometrics, 6:5549–5632, 2007. ISSN 1573-4412.

Christopher T. Conlon. An mpec approach to empirical likelihood demand estimation. 2009.

Christopher T. Conlon. A dynamic model of costs and margins in the lcd tv industry. 2010.

R. Cools. An encyclopaedia of cubature formulas. Journal of Complexity, 19(3):445, 2003.

Ronald Cools. Constructing cubature formulae: the science behind the art. Acta Numerica 1997, pages 1–54, 1997.

Ronald Cools. Advances in multidimensional integration. Journal of Computational and Applied Mathematics, 149(1):1–12, 2002.

Peter Davis and Eliana Garcés. Quantitative techniques for competition and antitrust analysis. Princeton University Press, 2009.

Angus Deaton and John Muellbauer. An almost ideal demand system. The American Economic Review, 70(3):312–326, 1980a.

Angus Deaton and John Muellbauer. Economics and Consumer Behavior. Cambridge University Press, 1980b.

S. P. Dirkse and M. C. Ferris. The path solver: A non-monotone stabilization scheme for mixed complementarity problems. Optimization Methods and Software, 5:123–156, 1995.

S.G. Donald and H.J. Paarsch. Identification, estimation, and testing in parametric empirical models of auctions within the independent private values paradigm. Econometric Theory, pages 517–567, 1996.

Jean-Pierre H. Dubé, Jeremy T. Fox, and Che-Lin Su. Improving the numerical performance of blp static and dynamic discrete choice random coefficients demand estimation. SSRN eLibrary, 2011.

J.A. Dubin and D.L. McFadden. An econometric analysis of residential electric appliance holdings and consumption. Econometrica, 52(2): 345–362, 1984.

Colin Ellis. Do supermarket prices change from week to week? SSRN eLibrary, 2009.

M. C. Ferris and T. Munson. Complementarity problems in GAMS and the PATH solver. Journal of Economic Dynamics and Control, 24:165–188, 2000. URL `ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-12.ps`.

M. C. Ferris, C. Kanzow, and T. Munson. Feasible descent algorithms for mixed complementarity problems. Mathematical Programming, 86:475–497, 1999. URL `ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-04.ps`.

D.G. Fiebig, M.P. Keane, J. Louviere, and N. Wasi. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. Marketing Science, 29(3):393–421, 2010. ISSN 1526-548X.

A. Gandhi. Inverting demand in product differentiated markets. Draft, 2010.

P.E. Gill, W. Murray, and M.A. Saunders. Snopt: An sqp algorithm for large-scale constrained optimization. SIAM Journal on Optimization, 12(4):979–1006, 2002.

D. Goldberg. What every computer scientist should know about floating-point arithmetic. ACM Computing Surveys, 23(1):5–48, 1991. ISSN 0360-0300.

R. Griffith and L. Nesheim. Estimating households' willingness to pay. CeMMAP working papers, 2010.

R. Griffith, L. Nesheim, and M. O'Connell. Sin taxes in differentiated product oligopoly: an application to the butter and margarine market. CeMMAP working papers, 2010.

W.M. Hanemann. Discrete/continuous models of consumer demand. Econometrica: Journal of the Econometric Society, pages 541–561, 1984.

J.A. Hausman. Valuation of new goods under perfect and imperfect competition. In Timothy F. Bresnahan and Robert J. Gordon, editors, The Economics of New Goods, chapter 5, pages 207–248. University Of Chicago Press, 1997.

James J. Heckman. Sample selection bias as a specification error. Econometrica, 47(1):pp. 153–161, 1979. ISSN 00129682. URL http://www.jstor.org/stable/1912352.

Florian Heiss and Viktor Winschel. Likelihood approximation by numerical integration on sparse grids. Journal of Econometrics, 144 (1):62–80, 2008.

I. Hendel and A. Nevo. Intertemporal substitution and storable products. Journal of the European Economic Association, 2(2-3):536–547, 2004.

T.J. Holmes. The diffusion of wal-mart and economies of density. Econometrica, 79(1):253–302, 2011. ISSN 1468-0262.

J.F. Houde. Spatial differentiation and vertical contracts in retail markets for gasoline. 2011.

Panle Jia. What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. Econometrica, 76(6): 1263–1316, 2008. ISSN 1468-0262.

K.L. Judd. Numerical Methods in Economics. The MIT Press, 1998.

M. Keane and N. Wasi. Comparing alternative models of heterogeneity in consumer choice behaviour. 2009.

Yuichi Kitamura. Empirical likelihood methods in econometrics: Theory and practice. SSRN eLibrary, 2006.

A. Leicester and Z. Oldfield. An analysis of consumer panel data. 2009a.

Andrew Leicester and Zoë Oldfield. Using scanner technology to collect expenditure data. Fiscal Studies, 30(3-4):309–337, 2009b.

J. Martins, P. Sturdza, and J.J. Alonso. The connection between the complex-step derivative approximation and algorithmic differentiation. AIAA paper, 921, 2001.

M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation (TOMACS), 8(1):3–30, 1998. ISSN 1049-3301.

M.J. Mazzeo. Product choice and oligopoly market structure. The RAND Journal of Economics, 33(2):221–242, 2002.

D. McFadden. Econometric analysis of qualitative response models. Handbook of Econometrics, 2:1395–1457, 1984.

D. McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica: Journal of the Econometric Society, pages 995–1026, 1989.

D. McFadden. Computing willingness-to-pay in random utility models. In J. Melvin, J. Moore, and R. Riezman, editors, Trade, Theory and Econometrics: Essays in Honour of John S. Chipman, pages 253–274. Routledge London, 1999.

D. McFadden and K. Train. Mixed mnl models for discrete response. Journal of Applied Econometrics, 15(5):447–470, 2000.

Daniel McFadden. Econometric models of probabilistic choice. In Charles F. Manski and Daniel McFadden, editors, Structural Analysis of Discrete Data with Econometric Applications, chapter 5, pages 198–272. MIT Press, Cambridge, MA, 1981.

N. Nethercote and J. Seward. Valgrind: a framework for heavyweight dynamic binary instrumentation. ACM SIGPLAN Notices, 42(6): 89–100, 2007a.

N. Nethercote and J. Seward. How to shadow every byte of memory used by a program. In Proceedings of the 3rd international conference on Virtual execution environments, pages 65–74. ACM, 2007b.

A. Nevo. Mergers with differentiated products: The case of the ready-to-eat cereal industry. The RAND Journal of Economics, 31(3):395–421, 2000a.

A. Nevo. A practitioner's guide to estimation of random-coefficients logit models of demand. Journal of Economics & Management Strategy, 9(4):513–548, 2000b.

A. Nevo. Measuring market power in the ready-to-eat cereal industry. Econometrica, pages 307–342, 2001.

J. Nocedal and S.J. Wright. Numerical Optimization. Springer, 2000.

Chris Nosko. Competition and quality choice in the cpu market. Manuscript, Harvard University, 2010.

Y. Orhun. Spatial differentiation in the supermarket industry. 2005.

A. Pakes and D. Pollard. Simulation and the asymptotics of optimization estimators. Econometrica: Journal of the Econometric Society, pages 1027–1057, 1989.

A. Pakes, J. Porter, K. Ho, and J. Ishii. Moment inequalities and their application. Unpublished Manuscript, 2006.

A. Petrin. Quantifying the benefits of new products: The case of the minivan. Journal of Political Economy, 110(4):705–729, 2002.

Ciaran S. Phibbs and Harold S. Luft. Correlation of travel time on roads versus straight line distance. Medical Care Research and Review, 52(4):532, 1995.

P.C.B. Phillips and H.R. Moon. Linear regression limit theory for nonstationary panel data. Econometrica, pages 1057–1111, 1999. ISSN 0012-9682.

P.C.B. Phillips and H.R. Moon. Nonstationary panel data analysis: An overview of some recent developments. Econometric Reviews, 19 (3):263–286, 2000. ISSN 0747-4938.

W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. Numerical recipes: the art of scientific computing. Cambridge Univ Pr, 2007.

J. Reynaerts, R. Varadhan, and J.C. Nash. The convergence properties of the blp (1995) contraction mapping and alternative nonlinear algorithms. 2010.

John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. Econometrica, 55(5):999–1033, 1987. ISSN 00129682. URL http://www.jstor.org/stable/1911259.

S. Seiler. The impact of search costs on consumer behavior: a dynamic approach. 2010.

Katja Seim. An empirical model of firm entry with endogenous product-type choices. The RAND Journal of Economics, 37–3:619–640, 2006.

A. Seth and G. Randall. The grocers: the rise and rise of the supermarket chains. Kogan Page Ltd., 2001. ISBN 0749435496.

A. Shaked and J. Sutton. Relaxing price competition through product differentiation. The Review of Economic Studies, 49(1):3–13, 1982.

Benjamin S. Skrainka and Kenneth L. Judd. High performance quadrature rules: How numerical integration affects a popular model of product differentiation. Working Paper, 2011.

Howard Smith. Supermarket choice and supermarket competition in market equilibrium. Review of Economic Studies, 71(1):235–263, 2004.

AH Stroud. Approximate Calculation of Multiple Integrals. Prentice Hall, 1971.

Che-Lin Su and Kenneth L. Judd. Constrained optimization approaches to estimation of structural models. SSRN eLibrary, 2010.

K. Train. Halton sequences for mixed logit. Manuscript, Department of Economics, University of Berkeley, 1999.

Kenneth E. Train. Discrete Choice Methods with Simulation. Cambridge University Press, second edition, 2009.

UK Competition Commission. Supermarkets: a report on the supply of groceries from multiple stores in the united kingdom. http://www.competition-commission.org.uk/rep_pub/reports/2000/446super.htm, 2000.

UK Competition Commission. Safeway plc and asda group limited (owned by wal-mart stores inc); wm morrison supermarkets plc; j sainsbury plc; and tesco plc. http://www.competition-commission.org.uk/inquiries/completed/2003/safeway/index.htm, 2003.

UK Competition Commission. The supply of groceries in the uk market investigation. http://www.competition-commission.org.uk/inquiries/ref2006/grocery/index.htm, 2006.

UK Competition Commission. Groceries market investigation: Remittal of the competition test by the competition appeal tribunal. http://www.competition-commission.org.uk/inquiries/ref2009/groceries_remittal/index.htm, 2009.

U.S. National Science Foundation. Cyberinfrastructure vision for 21st century discovery. National Science Foundation, Cyberinfrastructure Council, 2007.

A. Wachter and L.T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming, 106(1):25–58, 2006.

colophon

This thesis was typeset with LaTeX 2ε using André Miede's elegant `classicthesis` LaTeX document class as the foundation for the design. `classicthesis` implements a typographic style based on *The Elements of Typographic Style* [Bringhurst, 2002] and is available for LaTeX via CTAN as "classicthesis". Nicholas Mariette's and Ivo Pletikosić's LyX port of `classicthesis` makes such beauty easily accessible to the masses. Both the LaTeX class and LyX template are available under the GNU General Public License. Those who require further details about fonts, line spacing, and other minutia should consult the documentation provided on CTAN.

Final Version as of January 13, 2012 at 22:31.