Understanding Multi-product Health Insurance Marketplaces: An Advancement in Aggregated Demand Estimation Using Bayesian Statistics

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Tsan-Yao Huang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Brynd Dowd, Adviser

November 2020

Acknowledgements

Thanks to my siblings, Elsa Li-Ya, Li-Ting and Kang-Hua for their help taking care of our parents while I spent time finishing my PhD. I also want to show my gratitude to my advisor Bryan Dowd for his invaluable support, encouragement, and suggestions. I am also grateful to Andres Musalem for sharing his codes and suggestions for the development of this thesis, to Caitlin Carroll, and Peter Huckfeldt for reviewing, commenting, and help me to polish this dissertation. And to my friends Coleman Drake, Jiani Yu, Xuyang Tang, Zoe Kao, Laura Smith, Fernando Alarid, Lucas Higuera and their family, my cats Feles and Clara for their support during my journey toward the finish line of this dissertation. It's not an end, just a beginning.

Dedication

To my parent, Chou-Yueh Chiao and Fu-Chuan Huang.

Abstract

Demand functions estimated by aggregated data used in economics and marketing often employ the approach of Berry, Levinsohn and Pakes (Berry et al., 1995). To apply the method, researchers are required to collect market shares for each product of interest along with product-level attributes. Yet in many applications observed market shares are aggregated by firms or brands which sell multiple products. My thesis addresses this empirical issue by advancing existing BLP estimation procedure from Musalem et al. (2009) by using aggregated market shares at the firm level (portfolio market shares) and product-level attributes. I provide a solution to recovering the distributions of preference weights and price elasticities when researchers are limited to data containing only market shares at firm-level but consumers make choices over product-level attributes. The applications are specifically applied to the Health Insurance Marketplaces in the US.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Affordable Care Act provided health insurance to 17 million previously uninsured Americans (Greene et al., 2015). About 23% of those newly insured are now covered by Health Insurance Marketplaces (Marketplaces). While the Marketplaces do provide new enrollees with better access to care and improved health outcomes (Goldman et al., 2018), recent trends of rising premiums, insurer exit and imperfect competitions in Marketplaces have led to policy discussions on reforming and replacing the Marketplaces.

An existing literature has studied consumer behaviors, market equilibria, and different subsidy designs in the Marketplaces. Yet the studies have by and large focused on State-based Marketplaces, e.g., California, Washington or Colorado (Tebaldi, 2017; Saltzman, 2019; Drake, 2019; Panhans, 2019). One reason for this is State-based Marketplaces provide ample opportunities for researchers to access individual level data. That is, researchers observe individual-level health plan choices along with enrollee demographics.

On the other hand, very little is known for consumer demand in the federal facilitated Marketplaces till very recently (Polyakova & Ryan, 2019; Aizawa & Kim, 2020). While enrollment data aggregated at market level are generally available through public data sources such as HealthCare.gov or the website of Center of Medicare &

Medicaid Services (CMS), the data structure is problematic have caused atypical difficulties for existing methods for demand estimation. Specifically, the data consist of aggregated market shares at level of multi-product firms while the researchers observe detailed attributes of products as independent variables. Existing methods for aggregated market shares such as BLP's (1995) require, however, that researchers observe both independent variables of choice attributes and dependent variables of market shares both at the product level.

When researchers have product-level characteristics but market shares aggregated by other dimensions of products, existing literature has proceeded estimating models by assuming a representative product exists. Specifically, Berry et al. (2004) use product-level attributes from modal vehicles to estimate demands on types of cars. In studies of Medicare Advantage plans, a similar issue arises as many plans are sold under a contract and market shares are reported at contract-county level. Hall (2007), and Nosal (2012) assume a contract can be represented by one of its plans so that they can match attributes of the representative plan to the contract-county level market share. Kim (2014), studying the same market, instead uses averaged characteristics over products as independent variables. However, the statistical properties of those models with representative products are not well understood. A particular concern is assuming "representative" product-level characteristics creates measurement error and will thus bias parameter estimates towards zero.

In this paper, I develop a new method for estimating demand functions in the presence of aggregated market shares extending an alternative Bayesian BLP approach from Musalem et al. (2009). This also corresponds to a specific research question in market science in that individuals are observed to make choices for multi-product firms, or brands, as outcomes, given detailed product-level attributes for each product as the independent variables. My approach focuses on observing aggregated choices, as neither the exact choice decisions for each individual selecting a specific product, nor the brand choice is unobserved.

Using the proposed new method, I find that researchers can recover demand primitives including both preference weights, their variances and covariances, and the mean utilities of choice characteristics independent of specific individuals. The insight of the Bayesian approach is to account for the data aggregation explicitly by simulating a set of individual consumer choices that are consistent with observed aggregated shares.

Second, I also contribute to the field of health economics by applying my Bayesian BLP model to consumer demand in the Marketplaces. First, I test my model using data from Covered California, one of the largest State-based Marketplaces. The advantage of focusing on Covered California is that I can compare the results of my Bayesian BLP model with existing work that uses granular, individual data (Tebaldi, 2017). I find the my estimates using Bayesian BLP with aggregated shares at the firm-level are comparable to those estimates from granular data at individual level.

I then estimate my Bayesian BLP model using data from Pennsylvania (PA), a Federally Facilitated Marketplace where only aggregated data is available. I find that consumers in PA have similar preferences as shown in other states. A young, unsubsized, and single enrollee in PA has a preference weight of -1.15 over premium, while a subsidied enrollee is more price sensitive, and the senior or family enrollees tend to be less price elastic. I use these results to conduct a policy change simulation: When a state changes its model of managing its Marketplace from federal-based platform to State-based platform, how much is the change in consumers' welfare? To address this question, I conduct a hypothetical cut of user fees of issuers of health plans, and thus the decrease in premiums, when the state of Pennsylvania switches its Marketplace to State-based model. I find an individual in the Marketplace of PA can have a gain in consumer welfare for about \$70 to \$150 due to the saving of user fees.

The thesis proceeds in the following arrangements. In Chapter 2 I review the existing literature on empirical studies of the Marketplaces from the perspective of health economics. I also review a particular branch of literature of industrial organization for multi-product firms and marketing, as well as the empirical approaches for estimating demand functions using aggregated data. In Chapter 3 I discuss the institutional background of individual Marketplaces in general. Chapter 4 covers the advancement of empirical approach of estimating demand functions using aggregated data, especially in the context of multi-product firms, and Bayesian statistics. The empirical application using the Marketplace in California, CoveredCA, is conducted with the proposed new method in this thesis and compared to another empirical study in CoveredCA in Chapter 5. In Chapter 6 I evaluate the welfare gain of consumers in the federally facilitated Marketplace in Pennsylvania. Chapter seven concludes.

Chapter 2

Literature Review

2.1 Theories and Empirical Studies in Multi-product Firms

The empirical difficulties of this thesis originate from the problem of multi-product firms. In theory, a specific branch of literature in industrial organization has discussed oligoplistic firms selling multiple products. For example, Shaked and Sutton (1990) assume a two stage game to study the conducts of multi-product firms. In the first stage firms decide how many products to provide in a market. In the second stage firms compete by setting prices of products. To track the implications of firm conducts, they assume firms sell up to three products and show there exist multiple pricing equilibria when firms compete in a typical Betrand pricing game. The finding of more recent development in theory, as Noche and Schultz (2018b) indicate, is there may exist multiple static equilibria for pricing conducts in general. To prove there could exist a unique pricing equilibrium, Noche and Schultz (ibid.) assume a consumer's relative ranking of any two goods is invariant to the presence of a third alternative (i.e., independence of irrelevant alternatives).

Empirically, many studies have recognized that firms could offer multiple products and account for the strategic pricing explicitly in policy simulations. For example, Nevo (2000a) conducted a merger analyses of firms by simulating the change of product ownership from different firms to one company. The empirical approach of Nevo (ibid), however, is based on the BLP approach. Berry et al. (1995) study the automobile demand using aggregated shares of vehicles and vehicle attributes in a discrete choice framework. Additionally, Smith (2004) studies retail stores at multiple locations – the market shares are thus aggregated by many locations of the same brand of retailers to the firm-level. To proceed with policy analysis measuring the market power of the chain of retail stores in England, Smith (ibid.) explicitly considers the data generating process in which 54 pre-specified types of consumers with one dimensional unobserved attribute visit one of many retail stores. Then predicted shares at the store-level are aggregated by choice probabilities of all types of consumers and by the portfolio of store locations. Smith then estimates the model using maximum likelihood.

A relevant school of thoughtis modeling brand/firm choices in marketing science, while using granular data of consumer choices of products and product attributes. For example, Guadagni et al. (1983) have granular sales data on consumers' purchases of coffee by Universal Product Codes (UPCs). The research question in marketing, however, is not modeling choices of products but modeling brands of coffee makers that sell multiple types of UPCs. The empirical strategy of Guadagni et al. (ibid.) is to reduce the dimension of UPCs of coffee powder into pre-classified types of brand-ounce of coffee powder and argue other product-or UPC-level attributes bring consumers no utility. For example, the consumers don't perceive granularity of coffee powder as an important factor upon purchase. Notable attributes of the choices on brand-sizes include prices, promotions of brand-size, and consumers' loyalties to brand and UPCs of the same brands.

Although researchers can have access to granular data like UPCs, another empirical need naturally arises as some researchers sometimes model products by different types or representative choices. For instance, Berry et al. (2004) use product-level attributes from modal vehicles to estimate demand for types of cars. While this apporach isn't empirically attractive– it potentially reduces the dimension of products in the consideration set of consumers– a possible concern is that using an averaged attribute induces measurement error. If consumers, for example, have made purchases of the good with higher prices than the average price measured by the researcher, the difference of the actual price and the averaged price becomes an unobserved product attribute.The scholar would then underestimate the true preference weights of the price.

In this thesis I propose an alternative approach to estimate a demand model with product attributes for brand and/or firm choices, regardless the dependent variable is aggregated. I focus on the implication of the proposed method in the context of using aggregated brand choices (i.e., BLP-type demand model). I estimate demand functions for multi-product firms by explicitly aggregating product choices to aggregated market shares at the brand level by firms' product portfolio. I apply this approach in the Marketplaces.

2.2 Literature on Health Insurance Marketplaces

The literature has used individual-level data on plan selection to estimate demand for insurance in the State-based Marketplaces. Saltzman (2019) studies the demand of insurance plans in State-based Marketplaces in California and Washington, and Tebaldi (2017) studies the equilibrium of insurance demand and supply and proposes optimal subsidy designs in Marketplaces using data from California. Drake (2019) collects networks of health plans in CA and evaluates the willness to pay for broader networks. Drake et al. 2020 further account for consumers inertia of insurance demands using the individual-level data of enrollments in CA. Panhans (2019) uses individual-level enrollments and medical claim data from Colorado's Marketplace and shows adverse selection exists in the market. That is, premiums of health plans are positively related with incurred medical costs from claims data.

Yet the studies on the Marketplaces in states using federal facilitated model are limited. Aizawa and Kim (2020) evaluate the effect of cutting advertisements on the enrollments in both federal facilitated and State-based Marketplaces. They study enrollment at the county-level, controlling for average county-level age, income health status and advertisements; thus, they implicitly assume that there is no unobserved heterogeneity across consumers and use the approach of Berry (1994) to estimate decreased demand resulting from government advertisements. Their estimates imply that, if federal government cuts an extra dollar on advertising, the Marketplace enrollment would descrease.

Another relevant study is by Polykaova and Ryan (2019) who study welfare losses due to mean-tested subsidy design in any market of imperfect competition. In particular, they argue, when firms have market power but cannot price-discriminate, targeted subsidies could lead to a shift of aggregated demand curve of firms but the net effects of optimal pricing is ambiguous. In theory, they demonstrate there exists "demographic externalities" in any imperfectly competitive market with multiple types of consumers with universal prices of goods while the amounts of subsidies are linked by the universal prices. Intuitively, if the aggregated demand shifts up by less price-sensitive consumers, the universal prices of goods could rise and reduce the welfare of price-sensitive consumers of target subsidies. Alternatively, it is also possible that a shift of aggregated demand results from the increase in the type of price sensitive consumers creates pressure to drive down the universal prices of goods.

Polykaova and Ryan (ibid.) test this theory by estimating the equilibrium of demand and supply of health insurance plans using data from the federal facilitated Marketplaces. They indicate a policy of universal subsidies to eligible consumers can improve social welfare by about one-thirds compared to the existing mean-tested subsidy design of Affordable Care Act. However, Polykaova and Ryan (ibid.) also face the same issue of using market shares aggregated by multiple-product firms. Their unit of analysis is enrollment by demographic groups of age and income groups at the county level, rather than at the product level. This is the typical reason that the standard BLP approach for estimating demand using aggregated data isn't applicable. Instead, they adopt a two stage non-linear least squares approach and restrict their parameters of interests on the demand function of plans and premiums, while controlling for actuarial fair value and fixed effects of health insurance plans. The estimation procedure of non-linear least square has two steps. First, the algorithm first fixes a particular set of initial values to the parameters of interest and predicts the aggregated enrollments at the county level by age and income groups. In the second step, the algorithm searches for candidates of all possible parameters that minimize the difference between the predicted enrollments and observed enrollments by those groups. In the model specifications of Polykaova and Ryan (ibid), similar to Aizawa and Kim (2020), consumer heterogeneity is only explained by observed demographics of ages and income levels.

In my thesis, I contribute to the Marketplace literature by advancing the approach of BLP in Bayesian framework. The approach is conceptually similar to Smith (2004) and Polykaova and Ryan (2019) in that the predicted probabilities or market shares at multi-product firms are explicitly accounted for. However, my Bayesian approach more closely follows the standard discrete choice modeling approach that allows for heterogeneity in consumer preferences through the use of random coefficients. This approach not only allows for observed consumer heterogeneity, but also unobserved heterogeneity not accounted for by individual-specific attributes. The latter is particularly important. Ericson and Starc (2012) indicate unobserved heterogeneity exists in Massachusetts' pre-ACA Marketplace. The flexibility of my model is also adaptable to other questions in aggregated demand not just by firms, but potentially by "coarser" aggregations like the shares of groups of plans by actuarial value metal levels, or the shares of the population insured by health plans by county-year or state-year, etc. The approach is based on Musalem et al. (2009) and is explained in the next subsection.

2.3 Comparison of Empirical Strategies

The Bayesian BLP approach of Musalem et al. (ibid) is an algorithm for estimating BLP using product-level market shares. The algorithm intuitively behaves similar to a discrete choice model at the individual level. While the exact choices of individuals are unknown, the choices can be simulated from the aggregated product-level market share as a sequence of categorical indexes of products. The sequence is consistent with observed product-level market shares in that the relative frequencies of each product category from the simulated consumers in the same market is the same as the observed market share of that product[1].

This approach acknowledges that aggregated data at the product level are indeed generated by unobserved consumers with unobserved preference heterogeneity. Researchers only observe the aggregated choices of those consumers. Yet the missing linkage from individual-level data to aggregated data is that consumers become de-identified. Researchers can build a discrete choice model by simulate individuals' choices through the data-augmentation approach in Bayesian statistics (Tanner & Wong, 1987). The Bayesian estimation procedure is to search for the parameters of interest once the missing data of choices are augmented from observed aggregated market shares. Researchers can proceed with estimation procedure with any standard discrete choices at the individual level.

[1]Depending on how the sequence is drawn, the relative frequencies are subjected to rounding errors.

The contribution of this thesis is that I generalize the algorithm of Musalem et al. (2009) and the empirical estimation of BLP with aggregated data in the following ways. First and most importantly, I accommodate the case in which aggregated market shares are not only at the product level, but can be aggregated through the product portfolio of multi-product firms. My approach starts by simulating individual choices for brands or firms from observed portfolio market shares of multi-product firms[2], incorporating additional information of portfolios of products of each brand. Then the parameter estimates are reported as the algorithm searches for sets of parameters that results in data consistent individual choice behaviors.

Second, I extend the Bayesian BLP of Musalem et al. (ibid) with random coefficients of unobserved heterogeneity and allow for observed heterogeneity of particular preference weights by interacting product-level attributes with individual observed demographic characteristics. This addition results in a modeling difficulty as augmenting choices with observed demographics would imply different sets of parameters of demand primitives that are consistent with observed market shares. I address this issue by incorporating the sampling algorithm of choices from Musalem et al. (ibid) for my model of repeated cross-sectional choices. Intuitively, the algorithm conducts repeated choices for the same set of individuals in a market with known demographics. Then each individual has a positive probability of selecting one of the choice option that in aggregation is consistent with observed market shares.

From an empiricist's perspective, it should be noted that the approach in this thesis creates a set of synthetic individuals *as if researchers had the data at the individual level.* Any policy simulations like merger analyses or welfare changes in consumer surplus can be conducted in a similar manner as in existing literature[3].

[2]The proposed algorithm is applicable to not just for observed portfolio shares, but for other aggregated outcomes if explanatory variables are product attributes.

[3]The drawback of the approach is to assume the number of simulated individuals are representative for the studied population.

Also, the advancement in empirical method has flexibility in application not only in the Marketplaces, but also in other sectors of empirical industrial organization. The model performance of the proposed Bayesian BLP, the simulation and estimating algorithm are discussed is evaluated in Chapter 4. And the empirical applications are in Chapter 5 and Chapter 6.

Chapter 3

Institutional Background

The ACA created State-based or Federally Facilitated Marketplaces in which eligible individuals have an option to purchase health insurance plans. In the Health Insurance Marketplaces (Marketplaces), an insurance company sells a product that consists of multiple insurance plans that share the same network design. Yet the plans differ by their actuarial fair values and are marketed by metal levels levels corresponding to actuarial values, including – bronze (60%), silver (70%), gold (80%) andor platinum (90%) [1]. More generous plans tend to have higher premiums and lower levels of cost-sharing. Marketplace insurers set premiums by geographic rating areas (GRAs) that either consist of sets of counties or three-digit ZIP within a state. Insurers set their plans' premiums in a given GRA. They are then multiplied by an age-rating factor for each household according to household members' ages and household income. While insurers must set their premiums uniformly within each GRA, they can vary which plans they offer within a given GRA. For example, if a GRA comprises of a county with multiple three-digit ZIP codes, then an insurer might offer different menus of products or product portfolios to consumers in each three-digit ZIP code.

Marketplace plans are subject to guaranteed issue and must offer Essential Health

[1]In addition to those four metal levels, a catastrophic plan is also available under the same product but could only be purchased by individuals under 30 years old.

Benefits. To further encourage take-up, the ACA included an individual mandate[2] that imposed taxes on uninsured eligible individuals. Individuals purchasing a Marketplace plan qualify for premium tax credits if their family incomes are between 100% and 400% of the Federal Poverty Level. The premium tax credit subsidy is calculated on the second-lowest premium silver plan available to the consumer, but the tax credit is applicable to all available plans. One implication of this subsidy design is that it will not change the differences between premiums for plans within markets.

In addition to premium tax credits, the eligible individuals whose income is less than 250% of the Federal Poverty Level can receive further cost-sharing reduction subsidies (CSRs) if they select a silver plan. These subsidies reduce the cost-sharing amounts consumers will pay, thereby increasing actuarial value. Marketplace plans' benefit designs, therefore, will not only vary by metal level, but by consumers' income levels. This unique structure might indicate why silver plan tends to have higher market shares than plans in other metal levels for the CSR-eligible population (DeLeire et al., 2017).

The individual mandate was repealed in 2019. The federal government is no longer transferring funds to insurers to cover CSRs, which they are legally obligated to provide. The latter policy change had a large impact on premiums in 2019 and 2020 (Drake & Anderson, 2020), as many states allow insurers to increase premiums to compensate their loss. Interestingly, as premium tax credits, by design, are calculated by premiums of second-lowest silver plan in a market, the premium increase resulting from the defunding of CSR subsidies has resulted in a phenomenon known as "silver loading" that has led to increased enrollment (Drake & Anderson, ibid.). That is, the increase in premiums have led to higher subsidies individuals can receive. And

[2]The mandate has become abolished in 2019 and no longer applies to uninsured individuals afterwards.

given that premium tax credits are applicable to all plans in individual markets, many consumers could have zero out-of pocket premiums net of the tax credit if they select less costly insurance plans. In my empirical study in Chapter 6, I also take the effect of silver loading into account, along with the repeal of individual mandate.

Chapter 4

Bayesian Demand Estimation for Multiple-Product Firms

4.1 Data Generating Process and Demand for Portfolio Market Shares

This section discusses the conceptual model of multi-product firms and the estimation involved in my Bayesian approach given a collection of data within each market and time. The context specific example is the demand for health insurance. Yet the approach can be extended to estimating demands in hospitals, primary care providers, purchasing new cars or other industries.

First, suppose consumers are selecting health insurance plan from a set of Marketplace insurers. A choice in their consideration set is defined at the level of product-firm, or product, for simplicity. In the Marketplaces the product-firm can be a particular health plan-insurer. The distinction of plans matters, especially in the context of this thesis, as many plans are often categorized by one set of plan attributes. And, in certain policy question the choices are aggregated by metals over multiple insurers, but not across plans within the same insurer.

Next suppose researchers have conducted an experiment on a group of consumers selecting plans. In the thought experiment, the researchers can observe one of the four types of dependent variables, creating four discrete choice model scenarios as shown in Table 4.1. In the example there are five health plans plus one outside option available to consumers in an oligopolistic market in which each health insurer offers multiple plans. If the experimenters had observed the data generating process and known choices at individual level by plans, they would be able to record the choice decision as the outcome variable for each subject as $y_{ijtm} = I(\text{product j is selected by a consumer i at time t in market m}) = 1; = 0$, otherwise. This is scenario one, the most standard case in empirical applications of discrete models when both individuals, their characteristics and choices at the product level are observed along with product specific attributes.

Table 4.1: Data Structures of Dependent Variables in Scenarios One and Two

ID	Income	Choices	Firm Ownership	Firm Choices
1	high	0	0	0
2	high	1	a	1
3	high	2	b	2
4	low	3	a	1
5	low	4	b	2
6	high	5	b	2
7	high	1	a	1
8	high	2	b	2
9	low	3	a	1
10	low	4	b	2

Note: Assume 60% of the individuals have high versus low income. The individuals have five options of plans plus being uninsured. The each of the plan is owned by an insurer, and the ownership is known to researchers. The table shows the data structure in scenario one and two when researchers have granular individual-level data, and the model is for product choice in the 3rd column, or brand choice in the fifth column.

Second, I define scenario two as the case in which researchers have detailed

Table 4.2: Data Structures of Dependent Variables in Scenarios Three and Four

Product ID	Product shares	Firm Ownership
0	10%	0
1	20%	a
2	20%	b
3	20%	a
4	20%	b
5	10%	b

(a) Data for Scenario 3

Firm ID	Portfolio Shares
0	10%
1	40%
2	50%

(b) Data for Scenario 4

Note: In 4.2a and 4.2b, the joint distribution of income and choices is often unknown when researchers have aggregated data in scenario three or four. Yet the main difference is in scenario three the dependent variable is product shares, while in scenario four the observed market-level shares is at the brand level, or as portfolio shares. In all of four scenarios, product IDs, firm ownership, firm ID, marginal distributions of income, as well as product attributes are known to researchers.

individual-level data but model brand choices. That is, the research question is about choices of insurers given product-level attributes and observed consumers' selection of insurers. One more realistic example is in marketing research. Researchers observe detailed product-level choices and product-level attributes by stock keeping units (SKUs). Then a brand choice is aggregated over the portfolio of SKUs under the same name of a brand.[1] In health insurance industry the consumers could select a "metal level" that consist of multiple plans with same actuarial fair value but different product attributes and each of the plans of the same group is owned by different insurers[2]. The data structure of the dependent variables in the first two scenarios are

[1] An alternative approach to model brands or firms is to project the space of available products or stock keeping units onto some common distinct attributes among the consumer goods. See Fader and Hardie (1996), for an example that reduces the dimensions of stock pile units of 56 fabric softeners into five attributes by sales, size, formula, form and brand. In their analysis the number of choices in to consideration was hence reduced to 22 types of softeners.

[2] See, for example, Tebaldi (2017) that estimates demands by using individual choices and predicts

shown in Table 4.1.

In the third scenario, the dependent variable is aggregated at the product level as shown in Berry et al. (1995) and BLP-like studies. In this scenario, the observed dependent variable is the collection of aggregated product-level market shares that are averaged out by individuals, while the independent variables are product-level attributes. The dependent variable is shown in Table 4.2, panel 4.2a. Notice that scenario three and scenario one both have the same six dependent variables.

Lastly, the fourth scenario is the data structure used in this thesis and is defined by using dependent variables *aggregated at the group level* while the independent variables are product-level attributes. The groups, similar to those in scenario two, can be geographic locations, or by product types, portfolios of multiple firms, or brands. Here, I focus on the application of multi-product firms and defined the aggregated market share at the firm level: portfolio market shares of a firm or brands. The dependent variables are defined as the sum of product-level market shares by the portfolio of firms[3]. A particular empirical application of portfolio market shares is in sufficient statistics for merger analyses for multi-product firms (see Nocke & Schutz, 2018a). Empirically, the data structure in scenario four is comparable to the standard BLP approach in the third scenario in that both study markets using aggregated dependent variables. From theoretical perspectives on the supply side of a market economy, both approaches are based on modeling multi-product firms competing in Betrand Nash games by setting product prices. Also Noche and Schultz (2018b) have proved that the pricing strategy of multi-product firms can have unique market equilibria when the utility function follows multinomial logits. However, they indicate the firms' pricing strategies might have multiple equilibria when consumers heterogeneity is considered in empirical application.

price elasticities by mental tiers in California.

[3]The definition can alternatively be the average probabilities of firm choices over individuals as described in scenario two.

I propose a Bayesian approach that focuses on understanding and modeling the demand side of markets in which firms own multiple products. The particular difficulty of the modeling strategy is that researchers can only observe portfolio market shares, rather than product-level aggregated market shares. The atypical feature of scenario four, as shown in Table 4.2, panel 4.2a, is the number of dependent variables is the same as counts of firms plus outside option of no purchase, while the exact choices are not observed but aggregated as in scenario three.

Table 4.3: Comparing Four Scenarios

	Scenario			
Observed Data	1	2	3	4
Individual choices				
– directly for research questions	v	v		
– indirectly from surveys			v	v
Individual demographics				
– directly for research questions	v	v		
– indirectly from surveys			?	?
Use Aggregated choices		v	v	v
Product attributes	v	v	v	v

Note: Checks denote the case when the types of data are available for discrete choice models. Question marks denote that using the data would be optional in research questions. In all scenarios, researchers observe the same set of product-level characteristics that are independent of demographics of consumers.

Table 4.3 summarizes the four cases and distinguish the use of data at individual or aggregated level explicitly. Occasionally, demographic characteristics are used to account for observed differences in choice probabilities or market shares. In my classifications I assume, for simplicity, in both scenario one and two using individual-level data is equivalent to observing choices and demographic characteristics for each of decision makers. While in scenario three and four, a pool of surveyed and de-identified individuals can be sampled or collected through public use files like the Current Population Survey. Alternatively, researchers might draw some household incomes from wealth distributions and label the sample as observed individuals. Then, demographic attributes in scenarios three or four can be either at the individual level in surveyed data or hypothetical individuals sampled from aggregated data. The dependent variables, however, are at market level and aggregated by products or firms.

The most important distinction of the four scenarios is the observation of one type of outcome variables. Researchers thus need different estimation approaches. Specifically, when researchers could collect very granular data for choice decisions at the individual level as in scenario one and two, they can estimate a discrete choice model using existing approaches.[4] When aggregated product-level market shares are the observed outcomes, researchers often adopt the approach from Berry et al. (1995) and so called BLP estimation approach when outcome variables at the product level and product attributes are available as in scenario three.

The major distinction in scenario four in this study is that researchers could have collected product-level attributes, but don't observe aggregated product-level market shares. Instead, they have observed portfolio market shares generated by multi-product firms. The portfolio market shares are aggregated over product-level market shares by the portfolio or ownership index J_f for each firm. Alternatively they are the averages over all subjects, if the individual choices of brands could be observed. Recognizing this, I develop an estimation procedure to estimate parameters of interest on the demand side when the dependent variable is aggregated at the firm level by creating a synthetic set of individuals whose choice decisions are consistent

[4] See, for example, Train (2009) for comprehensive surveys for available approaches.

with observed portfolio market shares. The process is based on Musalem et al. (2009), who couple both insights of BLP and Bayesian data augmentation approaches to fill in the gap of unobserved individual choices.

4.2 Model Setup and Parameter Identification

The demand model developed throughout the thesis has the following form

$$u_{ijtm} = \beta_{imt} x_{ijtm} + \alpha_j + \xi_{jtm} + \epsilon_{ijtm}, \forall j \setminus \{0\}, \quad (4.1)$$

$$u_{i0} = \beta_{imt} x_{i0tm} + \epsilon_{i0tm}. \quad (4.2)$$

where $i = 1, ..., N$ stands for individuals, $j = 1, ..., J$ stands for in-market products plus an outside option $j = 0$. The index of $t = 1, ..., T$ stands for time, and $m = 1, ..., M$ is the geographic location. A market will be denoted as a pair of (t, m) in my thesis. The parameters of β_{imt} is a K-dimensional vector of individual-specific preference weights and x_{ijtm} is K-dimensional vector of observed product-level attributes that may vary by individuals. For example, premiums of health insurance plans may vary because of individual's family status and age and is a factor of some base price of a product. Notice that, in addition, the specification above for the outside option has a term of x_{i0tm} in the utility function. In the literature this term is deducted from all other options and hence the utility of outside option is normalized. Yet my specification includes the un-normalized term explicitly, as in the Marketplaces it can stand for the total costs of being uninsured. From economic theory, the coefficient of premiums in β_i is $\beta_{i,prem}$.

It should be noticed that in the context of multi-product firms, a product j can be owned by a firm f through the ownership indicator $1 = I(j \in J_f)$, where J_f is the product portfolio of the firm. In other words, the products, with a slight abuse of notation, stands for a pair (j, f), or a product can be interchangeably denoted by a product-firm combination. This distinction is important, particularly when some product characteristics offered by different firms can be classified into the same product category in the health care sector. One example is that a silver-metal-plan is a product labeled by actuarial-fair value by multiple insurers in Marketplaces. Many health insurers can market silver plans, yet those silver-insurer plans can have varying product attributions like different benefit designs or networks of providers.

The variable, α_j stands for a vector of fixed effects or mean utility at the product level independent of individuals, markets and years[5]. The inclusion of fixed effects may help capture some unobserved categorical attributes among different products. For example, when α_j could stand for the fixed effects of firms, a binary dummy of all inside goods, or types of networks of a plan etc. Another important feature in the model is the inclusion of unobserved (to researchers) product-, time- and market-specific demand shocks ξ_{jtm}. This term absorbs the unexplained part of utility, and hence variations in choice probabilities or market shares, that are not captured by observed attributes. In insurance markets, the effects could be customer service for a product, qualities of insurance plans or plans' broadness of provider networks. The last quantity, ϵ_{ijt}, represents other unobserved individual heterogeneity for each individual, year, market and product. Lastly, it should be noticed that there is always an out-side option of not purchasing any good for consumers. There are $J+1$ options in total. I index the outside option as good 0. For good 0, its observed and unobserved product attributes are zeros but the individuals have random utility from the good

[5]Since only differences of a consumer's utility between two goods matter, any constant for the same individual in the same market and time period won't be identified. This is particularly important in some specifications when "fixed effects" at market-level are included. For example, if one specifies that a common state fixed effect exists for both insurance plans and being uncovered by health insurance for a consumer in a county of that state, the fixed effect becomes unidentifiable. But an alternative approach is to assume being insured in that state brings some utilities compared to being uninsured in the same state. In this case the state fixed effect could be identified when the effect is indeed interacted with the inside goods.

ϵ_{i0tm}.

I assume the individual errors follow type-one extreme value distribution so that $\epsilon_{ijt} \sim F(\epsilon_{jt}) = 1 - exp(-exp(\epsilon_{ijt})), \forall j \in J \cup \{0\}$. This assumption allows me to derive individual choice probabilities for each product analytically $p_{imt}(j|\beta_{imt}, \alpha_j, D_i, x_{ijmt}, \xi_{jmt})$ as multinomial logits by conditioning on the set of $(\beta_{imt}, x_{ijtm}, \alpha_j, D_i, \xi_{jtm})$ in the technical appendix A. For notational simplicity I am about to omit the index (m, t), unless otherwise stated, to simplify notations in following discussions.

I adopt the following specification for my main modeling strategy as McFadden and Train (2000) have shown that this kind of random coefficient discrete choice models can rationalize any random utility framework. The random coefficient model has the first stage prior

$$\beta_i \sim N(\mu_\beta + \delta D_i, \Sigma_\beta). \tag{4.3}$$

One of the main objectives is to estimate the mean and variance of this first-stage prior, $(\mu_\beta, \delta, \Sigma_\beta)$. The constant μ_β is a $kx1$ vector, while the set of δ is a kxl vector multiplied by a lxd matrix, D_i, of individual attributes. This specification can thus include demographic characteristics to account for observed heterogeneity for preference weights. The specification is particularly important in health care sector since many times choice decisions in health insurance, for example, depend on age, income or family status in Marketplaces. Finally, Σ_β is a kxk variance-covariance matrix to account for unobserved heteregeneity of preference weights. Parameters of interests, $(\mu_\beta, \delta, \Sigma_\beta, \alpha_j)$, and β_i will also be augmented using Bayesian approach in estimation section using Monte Carlo Markov Chain (MCMC).

The last component in the utility function are the products-specific demand shocks, ξ_j. The inclusion of this term has two important roles. First, the term absorbs any unobserved omitted product-specific variables that could have been important in consumers' choice decisions. Second, in many application one of the product attributes, for example, premiums or prices, could be correlated with other important attributes that are unobserved to researchers. Alternatively, prices are endogenous because of firms' multi-product pricing decision and market competition. See Appendix B for details for the pricing equation in the literature and specific applications in health care. Without accounting for endogeneity, the estimated preference weights for prices can be biased.

In this paper I address endogeneity using the control function approach from Petrin and Train (2010). One advantage of this approach is I can simply add a control variable in the utility function and test the endogeneity after estimation, if I can gather a set of instrument that exogenously predict the endogenous variables. An alternative strategy, also discussed in Appendix B, is to fully specify equations in supply side to address the source of endogeities in utility function as in Yang et al. (2003). However, in many empirical applications, researchers might be only interested in estimating preference weights. Using control functions in my random utility model hence simplifies the estimation difficulty and computational burden. In the estimation section I will describe how to conduct the Bayesian estimation.

Identification of Parameters.

The identification of parameters in discrete choice models has two meanings. First, as shown in Appendix A, individuals make decisions by comparing different utilities of goods. This implies any fixed effects in utilities constant across all goods for the same time and within the same market will not be identified. Instead, the fixed effects that can be identified are the difference of a good between the reference option. Another example is the alternative-specific constant of the outside option is not identified in the discrete choice model, while the alternative-specific constants of each inside option are identified with respect to the a normalized mean utility of outside option. That is, estimates of ξ_j are the difference between the alternative specific constant of good j and good 0. In the paper I follow the convention of literature and normalize $\xi_0 \equiv 0$.

Second, the identification of the parameters can mean whether parameters of interest can be recovered if sufficient data variations exist. The means of individual preference weights, μ_β, are identified if there exist differences among product attributes. For example, preference weights for price can be identified by the difference between prices of inside options and the outside option.

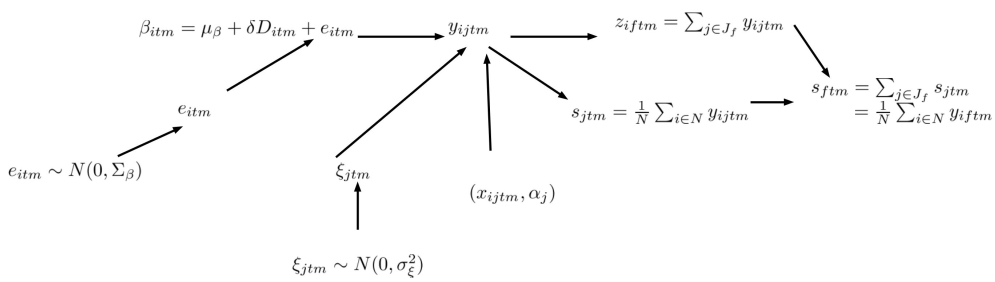
Parameters of interest can be identified not only by variations among attributes within the same market and year, but due to different market structures. Specifically, the variance of individual preference weights, Σ_β can be identified if researchers can observe two identical markets in all dimensions except sizes of product portfolios. Or the researchers observe different sets of firms selling the same set of competing products in two markets but the second market has a third firm in competition. Indeed, those two cases imply the choice probabilities for any two firms might not decrease proportionally within the same market if a third competing firm is present – the independent irrelevant alternative (IIA) property of multinomial logit choice models. Yet observing violations of IIA by observed portfolio shares over different markets can be accounted for by including heteregeneous preference weights: hence, IIA violation becomes a source of the identification of the distribution of heteregeneous preferences. In the example above, the first IIA violation results from an additional product choice while in the second case IIA is violated in the presence of a third multi-product firm.

Finally, the variations across markets are especially important for identifying the effects of observed heteregeneity, δ. As Nevo (2000b, p. 529) indicates "To tie demographic variables to observed purchases [...], several markets, with variation in the distribution of demographics, have to be observed." Hence δ can be identified if researchers have multiple markets of similar products and firms competing. In those markets, *ceteris paribus*, observed portfolio market shares differ because the demographic distributions vary.

4.3 Estimation Procedures

Estimation Using Acyclin Graphs

Figure 4.1: Data Generating Process



A particular advantage of Bayesian statistics is its ability to simplify the estimation of complicated models in previous section. More specifically, the conceptual model is seen in Figure 4.1 in which the observed choices by products or firms are seen as y_{ijtm} or z_{iftm} at individual level. If the data are aggregated, then the choices are available as s_{jtm} or s_{ftm} on the very right hand side of the graph. Recall the parameters of interest consist of $(\mu_\beta, \delta, \Sigma_\beta, \{\alpha_j\})$ given observed data of product-level attributes x_{ijtm}, and demographic distributions D_{itm}, but I only observe a vector of market shares, s_{ftm}, for each firm in a market.

The first step of the estimation procedure is to rewrite the utility function in Section 4.2 as two sets of parameters - one with random coefficients and the other constant for all individuals.

$$u_{ijtm} =(\mu_\beta + \delta D_i + e_i)x_{ijtm} + \alpha_j + \xi_{jtm} + \epsilon_{ijtm}, \quad (4.4)$$

$$=(\mu_\beta + e_i)x_{ijtm} + (\delta D_i x_{ijtm} + \alpha_j + \xi_{jtm}) + \epsilon_{ijtm}, \quad (4.5)$$

$$=(\mu_\beta + e_i)x_{ijtm} + (\delta D_i x_{ijtm} + \alpha_j + \phi\eta_j)$$

$$+\sigma_{\xi|\eta}\tilde{v}_j + \epsilon_{ijtm}, \quad (4.6)$$

$$\phi \equiv \frac{cov(\xi,\eta)}{\sigma_\eta^2}. \quad (4.7)$$

The first equation is derived by substituting the first stage prior of β_i into the utility function by rewriting $\beta_i \equiv \mu_\beta + \delta D_i + e_i$ and $e_i \sim N(0, \Sigma_\beta)$. The last step is to replace the endogenous ξ by the conditional distribution derived for the control function.

Next define $b_i \equiv (\mu_\beta + e_i)$. The part of individual specific preference weight is $b_i \sim N(\mu_\beta, \Sigma_\beta)$. The introduction of e_i is used to facilitate model derivation. By the control function approach the potentially endogenous demand shock ξ_j is written as the sum of conditional mean on η, and its coefficient $\frac{cov(\xi,\eta)}{\sigma_\eta^2}$, as well as an exogenous of product specific shock that normally distributes as $v_j \equiv \sigma_{\xi|\eta}\tilde{v}_j \sim N(0, (1-\rho_{\xi,\eta}^2)\sigma_\xi^2) = N(0, \sigma_v^2)$. When one is willing to assume all attributes are exogenous, the distribution of unobserved shocks to products, ξ_j, is independent of η_j.

In Bayesian statistics, it is necessary to specify a distribution fo ξ_j. As Jiang et al. (2009) indicates the performance of Bayesian BLP models is not sensitive to the distributional assumption of ξ_j, I assume it follows a normal distribution $\xi_{jtm} \sim N(0, \sigma_\xi^2)$. This simplified specification of shocks suggests that product-specific unobserved effects are independently drawn from the same distribution over time, across products, markets or within the same multi-product firms.

The major advancement of this thesis is to estimate the model from Figure 4.1 as if researchers would estimate a random-coefficient discrete choice model at individual level for firm choices. In comparison to Musalem et al. (2009), the choice is product, while in this thesis the choice is a brand or multi-product firms. The estimation uses Monte Carlo Markov Chain with Gibbs sampling for $(\mu_\beta, \delta, \Sigma_\beta, \{\alpha_j\})$ and is a very standard Bayesian analysis for random coefficient model as in Train (2009, chapter 12). Below I discuss how the choices of brands are simulated and updated through MCMC iterations. Readers interesting in the details of estimation of other parameters are referred to appendix C for the technical discussions and pseudo-codes.

First, to apply the discrete choice model, I have R individual in each market-period and define the choice indicator $y_{ijtm} = I(\text{ if j is purchased by i})$, where $I(.) = 1$ when the condition is true, otherwise it is zero. The BLP model for portfolio market shares is based on transforming the estimation procedure for choice decisions of products into choices of firms. Specifically, I let choice variable of firms be $z_{iftm} \equiv \sum_{f \in J_f} y_{ijtm}$. The choices will always satisfy observed market share constraints

$$\frac{\sum_i z_{iftm}}{R} \equiv s_{ftm}{}^6. \tag{4.8}$$

It can be shown, had I observed the vector of individuals' choices for firms, z, the likelihood function in this model for $m = t = 1$ is[7]

$$\begin{aligned} L_{tm}(z|(\mu_\beta, \Sigma_\beta, \sigma_v^2), \xi, x, \phi, \delta, \eta_j) =& \Pi_{i=1}^R Pr(z_{iftm} = 1|b_{itm}, x, v, \phi, \delta) f(b_i|\mu_\beta, \Sigma_\beta) \\ & \times f(v_j|0, \sigma_v^2). \end{aligned} \tag{4.9}$$

$$Pr(z_{iftm} = 1|b_{itm}, x, v, \phi, \delta) = \Pi_f \left(\sum_{j \in J_f} \frac{exp(V_{ijtm} + b_{itm} x_{ijtm})}{\sum_j exp(V_{ijtm} + b_{itm} x_{ijtm})} \right)^{z_{iftm}}, \tag{4.10}$$

$$V_{ijtm} \equiv \delta D_{itm} x_{ijt} + \phi \eta_j + v_j + \alpha_j.$$

[7]With some slight abuse of notations, I drop subscripts for a collection in the variables.

It can be seen that a major distinction of the classical maximum likelihood function and Bayesian statistics with augmented parameters is in equation (4.9) the unobserved individual parameters and product specific errors are parameters sampled from a known parametric distribution. In the classic approach the unobserved individual parameters are averaged out, while v_j could also be integrated out if treated as fixed effects, or it can be estimated as parameters of fixed effects.

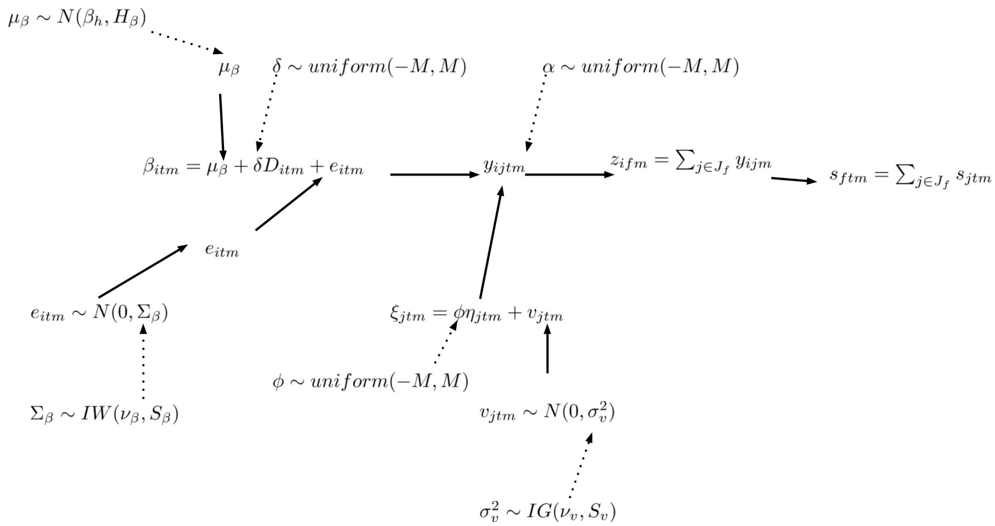
Notice that by conditioning on parameters and assuming the estimates of η_j are imputed from a reduced form pricing equation and treated as data, equation (4.10) follows a multinomial probability in which the number of categories is the size of multi-product firms and has one "successful draw" of a firm for each individual. And assuming individuals are mutually independent the joint likelihood of observed choices is the product of individual probabilities of (4.10). The full likelihood function is thus

$$LL \equiv \Pi_{tm}\Pi_{i=1}^{R} Pr(z_{iftm} = 1|b_{itm}, x, v, \phi, \delta) f(\{b_i\}|\mu_\beta, \Sigma_\beta) f(\{v_j\}|0, \sigma_v^2). \tag{4.11}$$

over markets. Alternatively, one might account for the sampling variations in η_j and specify the joint likelihood of $(\{z_i\}, \eta_j)$ for estimation. In this case I adopt a limited information likelihood approach from Yang et al. (2003) so that I could write $\eta_j \sim N(x_j - Z\Gamma, \sigma_\eta)$. Then the joint likelihood is $LL \times N(x_j - Z\Gamma, \sigma_\eta))$. Below I first describe the estimation procedure by treating η as a set of fixed data. In Appendix C I also discuss the benefit if η is updated in the MCMC iterations.

The last piece of applying Bayesian approach is the assumption of priors. In the acyclic graph in Figure 4.2 I show the flow of data-generating process in solid arrows and prior distributions on parameters in dashed arrows for a particular market and time. The advantage of the Bayesian approach using the acyclic graph is researchers can easily write down the full conditional posterior distribution of equation (4.11)

Figure 4.2: Estimation with Priors



and priors following three rules in Rossi et al. (2005, p. 69) using Gibbs sampling. Intuitively, using Gibbs sampling in the acyclic graph often results in posterior distributions in a more trackable formula. For example, to estimate Σ_β, the information needed will consists of $(\{e_{itm}\}, \nu_\beta, S_\beta)$, rather than the complete set of all parameters, and all of the other priors. It is because, from the flow chart, once e_{itm} is known, it has all information to estimate Σ_β. And it can be shown that the full conditional posterior distribution of Σ_β given $(\{e_{itm}\}, \nu_\beta, S_\beta)$ follows inverse Wishart distribution.

Notice that the collection of observed data is $\{(s_{ftm}, x_{ijtm}, D_{itm})\}$ and a set of valid instruments that will predict ξ_j[8]. Researchers lack the information of $(\{y_{ijtm}\}, \{z_{ijtm}\}\{\beta_{itm}\}, \{v_{jtm}\})$. I adopt similar strategy from Musalem et al. (2009) and treat them as extra parameters. For example, in equation (4.10), the individual coefficient with unobserved heterogeneity b_i and product specific shocks v_j will be treated as extra augmented parameters as in Tanner and Wong (1987) and drawn from their corresponding full conditional posterior distributions. The details can be found in Train (2009) in Chapter 12, while the posterior draws of $\{v_{jt}\}$ can be sam-

[8]When instruments are not needed, $\xi_{jmt} = v_{ijm}$ in the model.

pled from the likelihood function using Metropolis Hasting algorithm with Gibbs as shown in appendix C. Next I discuss how to simulate the choices from aggregated market shares.

Simulating Choices from Aggregated Market Shares

Musalem et al. (ibid.) simulate a sequence of $\{y_{ijtm}\}$ from observed market shares of products for a set of R simulated individuals for each market such that researchers would have observed choices of products at individual level. Then the simulated vector of $\{y_{ijtm}\}$ is treated as data in the case of a repeated cross sectional model. In the case of modeling panel data, each individual with a preference of $\beta_i = \mu_\beta + e_i$ has a positive chance to select each of the simulated product index from $\{y_{ijt}\}$ through estimation and MCMC samplings. Thus given a market and year, each iteration in MCMC can be regarded as a hypothetical experiment of choice occasions at individual level in which every experiment results in the shares of products consistent with observed data[9]. Analogously, I didn't observe the product-level shares, but I can draw a sequence of $\{z_{iftm}\}$ from observed portfolio market shares. And the average of $\{z_{iftm}\}$ over those R individuals in each market equals to the observed market share, s_{ftm}.

In the rest of the thesis I follow Musalem et al. (ibid.) and draw the indexes by first randomly assigning the sequence to the synthetic individuals[10]. Then the sequence

[9] In comparison to the original paper of BLP, v_{jtm} can be imputed by contraction mapping proposed by Berry (1994). The vector of v_{jtm} will be adjusted so that predicted market shares by the data-generating process, given x_{ijtm}, equals observed product-level market shares $\{s_{jtm}\}$. The missing information on $\{\beta_{itm}\}$ will be simulated directly by drawing a set of R individuals from $N(\mu_\beta, \Sigma_\beta)$. The sample is then averaged out such that mean predicted choice probabilities from those individuals are as close as the actual observed market shares. The estimating procedure of BLP does not involve in estimating the indicator variables $\{y_{ijtm}\}$.

[10] Alternatively, assignments of z_{ftm} can be modeled as drawing from a multinomial distribution which has the probability vector $\{s_{ftm}\}$ and R realizations for $length(s_{ftm})$ categories. Then any draw from the multinomial distribution can result in an augmented individual choice vector consistent with observed portfolio market shares. Or given a drawn vector of $\{z_{ftm}\}$, its permutation will also be consistent with the observed data of portfolio market shares. The complexity of all combination of choice sequences, however, seems to make estimation infeasible.

is treated fixed for R simulated individuals when data is repeated cross-sectional without observed heterogeneity of δ. Intuitively, without observed demographics, all simulated individuals are the same if researchers assign them a random number or labels in the statistical model[11]. Specifically, suppose $R = 2$ and we label those two individuals as $(1, 2)$ and assign augmented choices (z_1, z_2) sequentially. Then without observed demographics, we re-label those two individuals as $(2, 1)$ with the same set of choices, but (z_1, z_2) has the same implication of the model. In other words, a permutation of the simulated choice sequence will result in a classic label switching problem (Stephens,2000) since the augmented choice arrangements are not identified.

In Table 4.4, I provide another example that shows a hypothetical scenario in which five simulated individuals are drawn. The observed portfolio shares are 40% for selecting any health insurance plan while 60% for being un-insured. Without demographic attributes such as the column of income, either those five individuals are assigned to choice seq1 or choice seq2 will produce the same posterior distribution – the arrangement of ID is random and provides no extra information.

While in the same table, researchers collect demographic variables from different markets, then those demographic characteristics are identification constraints for the label switching problem: each simulated individual is indexed by her own observed demographic variables. Here I assume I have no conditional distribution of the portfolio shares on those demographic data[12]. In Table 4.4, I assume in the market 60% of the individuals have high income and the rest 40% individuals have low income. The individuals have two options. The index of one means no health insurance and two stands for being covered by any health insurance plans. The joint distribution of income and the coverage rate of insurance is not unknown while the marginal distri-

[11]But they are not the same for their unobserved characteristics that need to be simulated from Σ_β given an iteration. Thus the algorithm requires updates β_i for each iteration.

[12]In some applications researchers could have conditional portfolio shares by demographic groups. Then in those application it is necessary to fix the simulated choices by each group in estimation.

butions of income level and insured population are observed. Given 40% of insurance rate, either augmented sequence in choice seq1 or choice seq2 is possible. But the assignments of seq1 can result in different estimates of δ_{income}. Thus when observed demographics are used, it is necessary to account for either assignments of augmented choices.

In the classic BLP approach, this issue is addressed as the classic solution is to average out choice probabilities over individuals for both observed and unobserved heterogeneity. In the Bayesian BLP approach using augmented choices, one of the solutions is to re-assign the data consistent choice sequences through MCMC iterations while the model treats the draw of observed demographics fixed. Alternatively, it would be possible to fixed a vector of augmented choices but the model iterates through observed demographics to improve the precision of estimates. For example, Polyakova and Ryan (2019) match the moments of predicted aggregated market shares to the observed aggregated shares conditional on demographic groups. Similar strategy can be seen as adding additional moments, or market share constraints, as in Petrin (2002) with purchases of minivans with income groups.

However, in my empirical setting I use marginal distributions of demographics and treat them as data. I augment the choices of brands using the marginal distributions of market shares and update them through iterations. Augmenting choices using the conditional distributions of market shares on demographics could require the size of simulated individuals to expand exponentially by the groups of conditional demographics. Specifically, if a total of R individuals is drawn for each demographic group, then the number of simulated individuals could exponentially increase as R^g, where g is the count of a conditional distribution of market shares for a group. In other words, I need to draw augmented choices consistent with conditional market shares for each sub-group of observed demographics. And each of conditional distributions of sub-group is an extra market share constraint that the algorithm needs to satisfy.

Table 4.4: Simulating Choices with Demographic Distributions on Products Sold by Multi-product Firms

ID	Income	Choice Seq1	Choice Seq2
1	high	1	1
2	high	2	1
3	high	1	1
4	low	1	2
5	low	2	2

Note: Assume in a market 60% of the individuals have high income, and the true choice sequence is seq1. So 60% of individuals select good 1 (uninsured) and the rest select good 2 (insured). If the joint distribution of income and coverage rate is unknown while the marginal distributions of income level and insured population are observed, either choice sequences in choice seq1 or choice seq2 is possible are data consistent.

Using more constraints in the model, however, implies the algorithm would stick to a particular region of parameter space longer, as the chances of parameters to move freely in the parameter space are inversely related to the number of market shares constraints. For the interest of model development, however, I adopt a more simplified approach and allow for augmented choices to be consistent with marginal distributions of aggregated market shares. Future research should allow for adding additional constraints in the algorithm to improve precision of parameter estimates.

Hence to account for the assignments of augmented choice indexes using portfolio market shares, I follow Musalem et al. (2009) and treat $\{z_{iftm}\}$ as parameters and modify their Gibbs sampling for drawing posterior of $\{z_{iftm}\}$. In the algorithm, the values of $\{z_{iftm}\}$ are first randomly assigned such that market share constraints, $\sum_i z_{iftm} = Rs_{ftn}$, are satisfied. Then conditional on $(\{\beta_{imt}\}, \alpha, D_i, \{v_{jtm}\}, \{x_{ijtm}\})$, I can calculate the probabilities of purchasing any product for each individual. The product-level choice probabilities are aggregated into choice probabilities of firms by the set of an index function J_f. The updates of choice sequences using Gibbs sampling is based on the fact that, since market share constraints need to be satisfied, the update of choice indexes can depend on pairing any two individuals in the same market. That is, given any random pair of two individuals who make different choices on selecting multiple product firms, one consumer always has a positive chance to purchase a product from the other firm selected by the other paired consumer. Thus, the pair of consumers can switch their draws of $\{z_{iftm}\}$ with a positive probability and the result is also data consistent. The distributions of $\{z_{iftm}\}$ can be sampled by repeating this switching process through all pairs of consumers for each MCMC iteration as shown in equation using (4.10) [13] [14].

Setting Priors and Initial Values

In the following sections I use the same uninformative priors as shown in Figure 4.2 in next subsection of simulation, and Chapter 5 and Chapter 6 in estimating Marketplaces. Table 4.5 summaries the hyper parameters in those models. I choose the values so that prior distributions are not informative and diffuse enough for $(\mu_\beta, \Sigma_\beta, \sigma_v^2), \sigma_\eta^2)$. The priors of Γ, δ, and α are assumed to follow an improper prior of uniform distribution of $uniform(-M, M)$. That is, they have a prior distribution of the whole space of real line[15] By assuming the uniform distribution the posterior of those parameters would be similar to the fixed effects in classic statistics.

However, there are two particular concerns in MCMC. The first is whether re-

[13] Conceptually, a long run of MCMC with permissible choice indexes updated in each iteration is similar to average out the choice probabilities as in the classic BLP approach estimated by simulated likelihoods.

[14] An alternative strategy involves in proposing fully augmented sequences from multinomial distributions as in Chen and Yang (2007). The draws are accepted using Metropolis Hasting algorithm. However, Musalem et al. (2009) indicate that the alternative algorithm has low acceptance rates and can significantly increase computational time.

[15] The integral over the real line sums to infinity. Hence the prior is improper and not a distribution.

Table 4.5: Hyper-parameters for Priors

Parameters	Prior
β_h	0_K
H_β	$diag(100, K)$
ν_β	K + 5
S_β	diag(1,K)
ν_v	1.001
S_v	1
ν_η	leng(η) + 5
S_η	diag(1,leng(η))
M	$\sim$ inf

Note: Here $diag(x, y)$ is a diagonal matrix whose diagonal elements are x and its dimension is y. And the function $leng(.)$ returns the dimension of a vector. The variable K equals to the dimension of β, while η is the source of endogenous attributes that are accounted for when the first stage equation is updated in MCMC. Finally, M is assumed to be close to infinity, or the improper uniform distribution has zero density for any real number.

searchers could incorporate extra information from existing literature and use it as priors. For example, many studies have estimates of price coefficients from State-based Marketplace like CoveredCA. Then in my application in estimating Federally Facilitated Marketplace in Pennsylvania, it might be plausible to use the price coefficients of CoveredCA to improve the precision of price estimates in PA. While reasonable, the strategy implicitly assumes the eligible consumers in both Marketplaces would have similar tastes over premiums. Or even if consumers in PA have different tastes of premiums compared to consumers in CoveredCA, the researchers are willing to report the price coefficients from PA as a weighted average of evidence from consumers of PA using the informative prior. Yet the informative prior can bias the estimates of price coefficients in PA toward the estimates in CA, if those estimates are not exactly the same. In my empirical sections I instead remain agnostic and hence the prior distributions are uninformative.

Another closely related question is whether the researchers might use the estimates from CA as initial values of the MCMC to estimate the demand model in PA. In many applications, it is possible that researchers start with a set of initial values estimated by a multinomial logit model of choices. Then the set of initial values are supplied for the MCMC of the random coefficient mixed logit model. A particular advantage of this approach is the MCMC would start with a set of parameters potentially near the true joint posterior distribution. Thus the computation time for MCMC could be greatly reduced. On the other hand, if the posterior distribution has multi-modals, then the MCMC could stay in one modal of the true posterior distribution for a significant amount of time and draws.

In the rest of sections, I set the initial values of all fixed effect parameters to zeros, and random coefficients are drawn from the first stage prior with zero mean and with low or large variances. The estimates of η of the control function are calculated by ordinary least square with instruments. The sensitivity of my models would be potentially impacted by the initial values of variances in the first stage prior. For example, in my experiences when the variance is relative large, the MCMC algorithm tends to remain stuck for hours or even days due to numerical issues. If the variance of first stage prior is small, however, then the numerical issue seems to be a lesser concern and the MCMC can proceed with less computation time for a pre-determined number of iterations compared to the time used with larger variances.

Yet still another concern is using initial values might give evidence that whether MCMC converges. In my simulation and empirical estimations, I instead spot the necessary condition of convergence using trace plots– the posterior draws of parameters against iterations. If the MCMC fails to converge, the trace plots would likely show a trend of moving upwards or downwards. Yet observing no trends from trace plots wouldn't guarantee model convergence, either.

Finally, a distinction between initial values and priors is that the effect of initial value might dissipate when MCMC runs long, while the prior information is used to calculate the weighted average of parameters for all of the iterations in MCMC.

4.4 Simulation and Model Comparisons

To study the performance of the Bayesian demand model for portfolio market shares, I simulate a few data sets in the following ways. First, I provide two cases for the importance of unobserved heterogeneity of individual preferences – one with higher values for the variance covariance of Σ_β, and the other is with nearly zero variance-covariance. And in either high or low Σ_β, the observed demographics of D will be included or excluded for different data generating processes. So there are a total of four combinations of individual heterogeneity in my simulation study. In the case of low Σ_β, the random coefficient mixed logit model is close to a standard multinomial logit model regardless of inclusion of observed demographic characteristics. In comparison, a demand model with high unobserved heterogeneity in Σ_β might be associated with model miss-specifications. That is, some observed demographic characteristics could play essential roles in choices but become ignored or they are only partially observed by researchers due to limited access to data.

In all of the four cases, I generate a data set that has 25 markets and each market is observed twice. In each market-time there are always one outside option and its attribute matrix is normalized to be all zero, while four firms compete with multiple-product lines in all of those 50 market-periods[16]. The dimension of products in each market-time period is generated randomly as $1 + 2 *$ # of firms*round(1+random uniform). This ensures that the firms have at least two multiple products in each market period. As shown in Table 4.6, the number of products sold in each market ranges from 10 to 20, while a multiple product firm offers from at minimum two to 9 products. The portfolio market shares, when aggregated by the indexes of portfolio for each firm, ranges from 0.24% to 91.96% in those markets.

In each market-time 10,000 individuals with demographic attributes select a product by maximizing their utilities. Then I sample 50 to 200 individuals in the estimation procedure. The number of R is 50, 100, or 200. The demographic attributes are also observed if the observed heterogeneity exists in true data generating processes. In my simulation I have only one observed attribute and it stands for standardized age. The distributions of age follow $N(\mu_{mt}, \sigma^2_{mt})$, $\mu_{mt} \sim N(0, 1)$, and $\sigma_{mt} \sim |N(0, 1)|$. That is, the demographic distribution is drawn from normal distributions whose mean and variance differ by market and time. In the context of modeling demand in health insurance, potential enrollees across age distributions could have face different levels of premiums– thus potentially different premium elasticities of demand. The potential

[16]For simplicity I restrict the number of firms to be constant, but it is not required in applications. Instead, allowing the numbers of firms to vary by markets will help identify heterogeneous preference weights as IIA will not hold.

enrollees often can select a product of health insurance plan offered by one of health insurers or opting out and become uninsured. The attributes in my simulation models include premiums (prem) and maxmimum out-of-pocket (MOOP). Consumers in different ages have varying preference weights for premiums but not for out-of-pocket. Their unobserved heterogeneity for premiums and MOOP is drawn from bivariate normal distribution with non-zero correlations, $Sigma_\beta$. Consumers also have specific preferences for insurers and products. The former is captured on average by the fixed effect $\alpha_{f|j}$ while the latter by ξ_{jtm}. The whole economy is specified similarly in earlier section. Health insurers endogenously set premiums for their products in monopolistic competition. The endogenous problem is addressed if researchers find a vector of instruments Z that help predict premiums. Then a vector of residuals from predicted premiums enters the utility function by the control function approach in estimation procedures.

Here I re-iterate the model:

$$\begin{aligned}
u_{ijtm} =& \beta_{i,prem} prem_{jtm} + \beta_{i,MOOP} MOOP_{jtm} + \xi_{jtm} + \alpha_{f|j} + \epsilon_{ijtm}, \\
\begin{pmatrix} \beta_{i,prem} \\ \beta_{i,MOOP} \end{pmatrix} \sim & N \left[\begin{pmatrix} \beta_{prem} + \delta age_i \\ \beta_{MOOP} \end{pmatrix}, \begin{pmatrix} \sigma^2_{prem} & \sigma_{(prem,MOOP)} \\ \sigma_{(prem,MOOP)} & \sigma^2_{MOOP} \end{pmatrix} \right], \\
\begin{pmatrix} \xi_{jtm} \\ \eta_{jtm} \end{pmatrix} \sim & N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\xi} & cov(\xi,\eta) \\ cov(\xi,\eta) & \sigma^2_{\eta} \end{pmatrix} \right], \\
prem_{jtm} =& Z\Gamma + \eta_{jtm}.
\end{aligned}$$

In the simulation case of no demographics, $\delta = 0$. And in the case of low unobserved

Table 4.6: Summary Statistics for Simulated Markets

	min	median	mean	max
# of products	10	20	15.4	20
# of products per firm	2	4	3.85	9
Outside shares (%)	0.1	1.900	6.25	76.18
Portfolio shares (%)	0.24	11.48	20.00	91.96

Note: statistics are reported from 50 markets (25x2 market periods). The outside shares are the percentages of individuals that opt out purchasing any product. Portfolio shares are report at firm-market-time level. Summary statistics are reported for the simulation case of high unobserved heterogeneity and no demographic attributes.

heterogenous preferences,

$$(\sigma^2_{prem}, \sigma_{(prem,MOOP)}, \sigma^2_{MOOP}) = \begin{cases} (10^{-198}, 0, 10^{-198}), \text{if low unobserved heteroneity} \\ (1.106561, 1.05720, 1.037895), \text{otherwise} \end{cases}$$

And I let $(\beta_{i,prem}, \beta_{i,MOOP}) = (-1.5, -0.5)$. The fixed effects of four firms are randomly generated at $(0.5205918, 0.7990129, 0.2073603, 0.9276937)$. Both variables of Z and $MOOP$ are independently generated from a standard normal distribution, while the source of endogenous variables is $((\sigma^2_\xi, cov(\xi, \eta), \sigma^2_\eta)) = (0.41, -0.12, 0.09)$, and $\Gamma = 1.5$ in all of simulations. Hence the premium is an endogenous explanatory variable through the correlation of η and product specific attribute ξ.

In next section, I estimate the demand function using random coefficient mixed logit at individual level for choices at either product level or brand level, Bayesian BLP for product-level market shares as in Musalem et al. (2009), and my approach for portfolio market shares for multi-product firms.

Table 4.7: Low Unobserved Heterogeneity and No Observed Demographics

									Firm Fixed Effects			
Scenario		prem	MOOP	ϕ	σ^2_{prem}	$\sigma_{prem,moop}$	σ^2_{moop}	σ^2_v	1	2	3	4
1	2.5%	-1.96	-0.67	-1.57	0.05	-0.01	0.05	0.26	0.14	0.45	-0.32	0.45
	mean	-1.72	-0.53	-1.34	0.16	0.03	0.18	0.35	0.41	0.71	-0.02	0.73
	median	-1.71	-0.52	-1.34	0.14	0.02	0.15	0.34	0.40	0.71	-0.01	0.73
	97.5%	-1.59	-0.44	-1.10	0.40	0.13	0.51	0.47	0.69	0.97	0.26	0.97
	sd	0.10	0.06	0.12	0.09	0.04	0.12	0.05	0.14	0.14	0.15	0.14
	Bias	0.05	0.00	0.00	0.03	0.00	0.03	0.01	0.01	0.01	0.05	0.04
2	2.5%	-2.35	-0.84	-1.44	0.05	-0.04	0.05	0.33	-0.31	0.08	-0.96	0.03
	mean	-1.83	-0.62	-0.92	0.22	0.04	0.30	0.67	0.29	0.54	-0.26	0.61
	median	-1.79	-0.61	-0.91	0.17	0.02	0.19	0.54	0.32	0.56	-0.23	0.63
	97.5%	-1.61	-0.47	-0.45	0.76	0.26	1.31	2.09	0.64	0.88	0.13	0.96
	sd	0.19	0.09	0.25	0.18	0.08	0.33	0.42	0.23	0.20	0.26	0.22
	Bias	0.11	0.01	0.17	0.05	0.00	0.09	0.17	0.06	0.07	0.22	0.10
3	2.5%	-1.85	-0.64	-1.43	0.05	-0.02	0.05	0.17	0.50	0.77	0.09	0.81
	mean	-1.66	-0.55	-1.21	0.15	0.02	0.17	0.22	0.77	1.04	0.37	1.07
	median	-1.65	-0.54	-1.20	0.14	0.01	0.14	0.22	0.76	1.04	0.37	1.06
	97.5%	-1.54	-0.46	-1.01	0.35	0.07	0.44	0.30	1.02	1.31	0.64	1.32
	sd	0.08	0.04	0.11	0.08	0.02	0.10	0.03	0.14	0.14	0.14	0.14
	Bias	0.02	0.00	0.02	0.02	0.00	0.03	0.00	0.06	0.06	0.03	0.02
4	2.5%	-2.28	-0.88	-1.58	0.05	-0.06	0.06	0.24	0.16	0.57	-0.36	0.63
	mean	-1.83	-0.67	-1.08	0.25	0.02	0.31	0.49	0.62	0.93	0.15	1.03
	median	-1.79	-0.66	-1.08	0.19	0.01	0.22	0.40	0.63	0.93	0.16	1.04
	97.5%	-1.58	-0.51	-0.59	0.77	0.18	1.05	1.20	0.98	1.25	0.54	1.35
	sd	0.18	0.09	0.26	0.20	0.06	0.28	0.26	0.20	0.17	0.23	0.19
	Bias	0.11	0.03	0.06	0.06	0.00	0.10	0.06	0.01	0.02	0.00	0.01
	True value	-1.50	-0.50	-1.33	0.00	0.00	0.00	0.25	0.52	0.80	0.21	0.93

Note: Number of sampled or augmented individuals $R = 200$. Total MCMC iteration is 700,000 and convergence is observed after 400,000th iteration, after which every 500th draw is saved for the summary table. See Appendix D.2 for trace plots and marginal distributions. Bias is calculated by the squared difference between the posterior mean and the true parameter value.

Table 4.8: Low Unobserved Heterogeneity and with Observed Demographics

				age &						
Scenario		prem	MOOP	prem	MOOP	ϕ	σ^2_{prem}	$\sigma_{prem,moop}$	σ^2_{moop}	σ^2_v
1	2.5%	-1.66	-0.69	0.06	-0.11	-1.39	0.05	-0.01	0.04	0.51
	mean	-1.40	-0.58	0.10	-0.06	-1.13	0.20	0.02	0.17	0.65
	median	-1.38	-0.57	0.10	-0.06	-1.13	0.18	0.02	0.14	0.64
	97.5%	-1.24	-0.48	0.14	-0.01	-0.87	0.46	0.10	0.46	0.85
	sd	0.11	0.06	0.02	0.03	0.14	0.10	0.03	0.12	0.09
	Bias	0.01	0.01	0.16	0.00	0.04	0.04	0.00	0.03	0.16
2	2.5%	-2.34	-1.22	0.04	-0.26	-1.11	0.05	-0.06	0.05	0.74
	mean	-1.66	-0.77	0.14	-0.11	-0.41	0.32	0.07	0.36	1.62
	median	-1.60	-0.73	0.14	-0.10	-0.44	0.21	0.02	0.22	1.29
	97.5%	-1.37	-0.56	0.25	0.01	0.46	1.31	0.55	1.45	4.50
	sd	0.24	0.16	0.06	0.07	0.38	0.35	0.16	0.44	0.99
	Bias	0.03	0.07	0.13	0.01	0.85	0.10	0.01	0.13	1.88
3	2.5%	-1.57	-0.72	0.06	-0.10	-1.39	0.06	-0.01	0.04	0.34
	mean	-1.35	-0.61	0.09	-0.06	-1.16	0.22	0.05	0.16	0.43
	median	-1.34	-0.61	0.09	-0.06	-1.16	0.21	0.04	0.15	0.43
	97.5%	-1.20	-0.51	0.13	-0.01	-0.94	0.44	0.13	0.40	0.55
	sd	0.10	0.05	0.02	0.02	0.12	0.10	0.10	0.23	0.05
	Bias	0.02	0.01	0.17	0.00	0.03	0.05	0.00	0.03	0.03
4	2.5%	-2.27	-1.25	0.05	-0.23	-1.40	0.06	-0.01	0.05	0.51
	mean	-1.65	-0.79	0.13	-0.10	-0.83	0.36	0.12	0.39	1.06
	median	-1.58	-0.76	0.12	-0.09	-0.83	0.23	0.05	0.25	0.84
	97.5%	-1.36	-0.56	0.23	0.00	-0.28	1.34	0.60	1.63	2.84
	sd	0.24	0.17	0.05	0.06	0.29	0.34	0.19	0.41	0.60
	Bias	0.02	0.09	0.14	0.01	0.25	0.13	0.02	0.15	0.66
	True Value	-1.50	-0.50	0.50	0.00	-1.33	0.00	0.00	0.00	0.25

Note: Number of sampled or augmented individuals $R = 200$. Total MCMC iteration is 700,000 and convergence is observed after 400,000th iteration, after which every 500th draw is saved for the summary table. See Appendix D.2 for trace plots and marginal distributions. Bias is calculated by the squared difference between the posterior mean and the true parameter value.

Table 4.9: High Unobserved Heterogeneity and No Observed Demographics

									Firm Fixed Effects			
Scenario		prem	MOOP	ϕ	σ^2_{prem}	$\sigma_{prem,moop}$	σ^2_{moop}	σ^2_v	1	2	3	4
scenario 1	2.5%	-1.74	-0.59	-1.42	0.87	0.81	0.86	0.22	0.29	0.54	-0.11	0.66
	mean	-1.53	-0.47	-1.22	1.18	1.08	1.27	0.28	0.52	0.77	0.13	0.88
	median	-1.52	-0.47	-1.22	1.15	1.06	1.24	0.28	0.52	0.76	0.13	0.88
	97.5%	-1.39	-0.36	-1.03	1.61	1.41	1.90	0.36	0.76	1.01	0.39	1.12
	sd	0.09	0.06	0.10	0.19	0.15	0.27	0.03	0.12	0.12	0.13	0.12
	Bias	0.00	0.00	0.01	0.01	0.00	0.05	0.00	0.00	0.00	0.01	0.00
scenario 2	2.5%	-2.84	-1.29	-1.44	0.14	0.04	0.12	0.42	-0.98	-0.60	-1.61	-0.63
	mean	-1.71	-0.66	-0.88	2.09	1.44	1.72	1.29	0.25	0.45	-0.21	0.54
	median	-1.61	-0.62	-0.89	1.50	1.13	1.17	1.08	0.36	0.54	-0.09	0.64
	97.5%	-1.14	-0.28	-0.23	6.87	4.57	7.12	3.34	0.86	1.02	0.54	1.17
	sd	0.46	0.26	0.31	2.18	1.30	1.98	0.79	0.45	0.42	0.57	0.46
	Bias	0.04	0.02	0.20	0.96	0.15	0.47	1.08	0.07	0.12	0.18	0.15
scenario 3	2.5%	-1.66	-0.57	-1.40	0.88	0.82	0.86	0.14	0.61	0.90	0.26	0.97
	mean	-1.51	-0.46	-1.22	1.17	1.06	1.21	0.17	0.85	1.14	0.49	1.19
	median	-1.51	-0.46	-1.21	1.15	1.05	1.18	0.17	0.84	1.14	0.50	1.19
	97.5%	-1.39	-0.36	-1.03	1.55	1.36	1.76	0.22	1.11	1.39	0.76	1.43
	sd	0.07	0.05	0.09	0.17	0.14	0.23	0.02	0.12	0.12	0.12	0.12
	Bias	0.00	0.00	0.01	0.00	0.00	0.03	0.01	0.11	0.12	0.08	0.07
scenario 4	2.5%	-2.92	-1.29	-1.56	0.45	0.29	0.35	0.41	-0.42	0.02	-1.23	-0.08
	mean	-1.92	-0.71	-0.87	2.60	1.96	2.34	1.31	0.43	0.75	-0.08	0.81
	median	-1.86	-0.69	-0.87	2.14	1.71	1.89	1.19	0.49	0.78	-0.01	0.86
	97.5%	-1.29	-0.28	-0.18	6.85	4.94	6.60	2.93	1.05	1.28	0.68	1.44
	sd	0.42	0.27	0.35	1.80	1.23	1.73	0.68	0.39	0.35	0.50	0.39
	Bias	0.18	0.05	0.21	2.22	0.81	1.70	1.12	0.01	0.00	0.08	0.01
	True value	-1.50	-0.50	-1.33	1.11	1.06	1.04	0.25	0.52	0.80	0.21	0.93

Note: Number of sampled or augmented individuals $R = 200$. Total MCMC iteration is 700,000 and convergence is observed after 400,000th iteration, after which every 500th draw is saved for the summary table. See Appendix D.2 for trace plots and marginal distributions. Bias is calculated by the squared difference between the posterior mean and the true parameter value.

Table 4.10: High Unobserved Heterogeneity and with Observed Demographics

			age &							
	Scenario	prem	MOOP	prem	MOOP	ϕ	σ^2_{prem}	$\sigma_{prem,moop}$	σ^2_{moop}	σ^2_v
1	2.5%	-1.63	-0.64	0.21	-0.06	-1.35	0.72	0.62	0.62	0.35
	mean	-1.42	-0.50	0.26	-0.01	-1.13	1.09	0.91	1.08	0.43
	median	-1.41	-0.50	0.26	-0.01	-1.13	1.07	0.90	1.03	0.42
	97.5%.	-1.25	-0.36	0.31	0.05	-0.94	1.57	1.28	1.73	0.53
	sd	0.10	0.07	0.02	0.03	0.11	0.22	0.17	0.30	0.05
	Bias	0.01	0.00	0.06	0.00	0.04	0.00	0.02	0.00	0.03
2	2.5%	-3.04	-1.33	0.11	-0.36	-0.91	0.17	0.05	0.13	0.76
	mean	-1.83	-0.66	0.30	-0.16	-0.33	2.49	1.67	1.78	2.45
	median	-1.77	-0.62	0.29	-0.15	-0.37	1.84	1.28	1.31	2.29
	97.5%	-1.11	-0.24	0.60	0.01	0.50	8.06	5.04	6.15	5.05
	sd	0.50	0.29	0.13	0.10	0.37	2.22	1.39	1.62	1.17
	Bias	0.11	0.03	0.04	0.03	1.01	1.92	0.37	0.55	4.84
3	2.5%	-1.56	-0.60	0.20	-0.06	-1.38	0.83	0.69	0.64	0.19
	mean	-1.41	-0.48	0.24	-0.02	-1.20	1.18	0.95	1.01	0.24
	median	-1.41	-0.48	0.24	-0.02	-1.19	1.18	0.95	0.99	0.24
	97.5%	-1.27	-0.37	0.29	0.03	-1.01	1.61	1.25	1.49	0.29
	sd	0.08	0.06	0.02	0.29	0.29	0.24	0.15	0.22	0.04
	Bias	0.01	0.00	0.07	0.00	0.02	0.01	0.01	0.00	0.00
4	2.5%	-2.61	-1.14	0.13	-0.23	-1.22	0.20	0.09	0.16	0.48
	mean	-1.70	-0.64	0.28	-0.11	-0.68	1.83	1.32	1.60	1.04
	median	-1.62	-0.61	0.26	-0.10	-0.71	1.45	1.10	1.24	0.86
	97.5%	-1.17	-0.32	0.52	0.01	-0.08	5.71	3.57	4.87	2.40
	sd	0.37	0.22	0.10	0.06	0.29	1.41	0.98	1.31	0.51
	Bias	0.04	0.02	0.05	0.01	0.42	0.52	0.07	0.32	0.62
	True value	-1.50	-0.50	0.50	0.00	-1.33	1.11	1.06	1.04	0.25

Note: Number of sampled or augmented individuals $R = 200$. Total MCMC iteration is 700,000 and convergence is observed after 400,000th iteration, after which every 500th draw is saved for the summary table. See Appendix D.2 for trace plots and marginal distributions. Bias is calculated by the squared difference between the posterior mean and the true parameter value.

Table 4.11: Firm Fixed Effects with Observed Demographics

		Low Unobserved Heterogeneity				High Unobserved Heterogeneity			
Scenario	Firm	1	2	3	4	1	2	3	4
1	2.5%	0.21	0.46	-0.24	0.65	0.33	0.56	-0.09	0.77
	mean	0.48	0.72	0.04	0.92	0.56	0.78	0.21	1.00
	median	0.48	0.72	0.05	0.92	0.56	0.79	0.21	1.00
	97.5%	0.74	0.96	0.31	1.18	0.81	1.02	0.47	1.24
	sd	0.14	0.13	0.15	0.14	0.12	0.13	0.14	0.13
	Bias	0.00	0.01	0.03	0.00	0.00	0.00	0.00	0.01
2	2.5%	-0.77	-0.34	-1.46	-0.27	-1.45	-0.88	-2.18	-1.02
	mean	0.13	0.36	-0.49	0.55	-0.13	0.12	-0.59	0.36
	median	0.19	0.41	-0.41	0.6	-0.04	0.15	-0.51	0.42
	97.5%	0.57	0.80	0.05	1.01	0.77	0.92	0.38	1.23
	sd	0.32	0.29	0.39	0.31	0.56	0.47	0.61	0.54
	Bias	0.15	0.19	0.48	0.15	0.42	0.46	0.63	0.32
3	2.5%	0.51	0.79	0.11	0.97	0.68	0.96	0.35	1.11
	mean	0.78	1.09	0.41	1.24	0.92	1.20	0.59	1.36
	median	0.79	1.09	0.42	1.24	0.92	1.20	0.59	1.36
	97.5%	1.04	1.35	0.69	1.49	1.15	1.45	0.87	1.60
	sd	0.14	0.14	0.15	0.14	0.12	0.13	0.13	0.13
	Bias	0.07	0.09	0.04	0.1	0.16	0.16	0.15	0.19
4	2.5%	-0.12	0.28	-0.79	0.44	-0.15	0.25	-0.75	0.33
	mean	0.51	0.82	-0.03	1.00	0.63	0.90	0.20	1.13
	median	0.56	0.84	0.04	1.02	0.68	0.94	0.28	1.18
	97.5%	0.90	1.23	0.41	1.39	1.15	1.38	0.80	1.68
	sd	0.26	0.24	0.31	0.25	0.35	0.31	0.4	0.34
	Bias	0.00	0.00	0.06	0.01	0.01	0.01	0.00	0.04
	True Value	0.52	0.8	0.21	0.93	0.52	0.80	0.21	0.93

Note: Number of sampled or augmented individuals $R = 200$. Total MCMC iteration is 700,000 and convergence is observed after 400,000th iteration, after which every 500th draw is saved for the summary table. See Appendix D.2 for trace plots and marginal distributions. Bias is calculated by the squared difference between the posterior mean and the true parameter value.

4.5 Simulation results: Comparison of Posterior Estimates

The posterior distributions for all of the four scenarios are compared in Table 4.7 to Table 4.11. By comparing four different types of modeling strategies due to data availability on the dependent variables, we can see that most true parameters for preferences weights and other "fixed effect" parameters can be covered across all levels of (un)observed heterogeneity of individuals. Most estimates of parameters do fall in 95% credible intervals (confidence intervals in Bayesian estimation) even if researchers are limited to observe portfolio-level market shares.

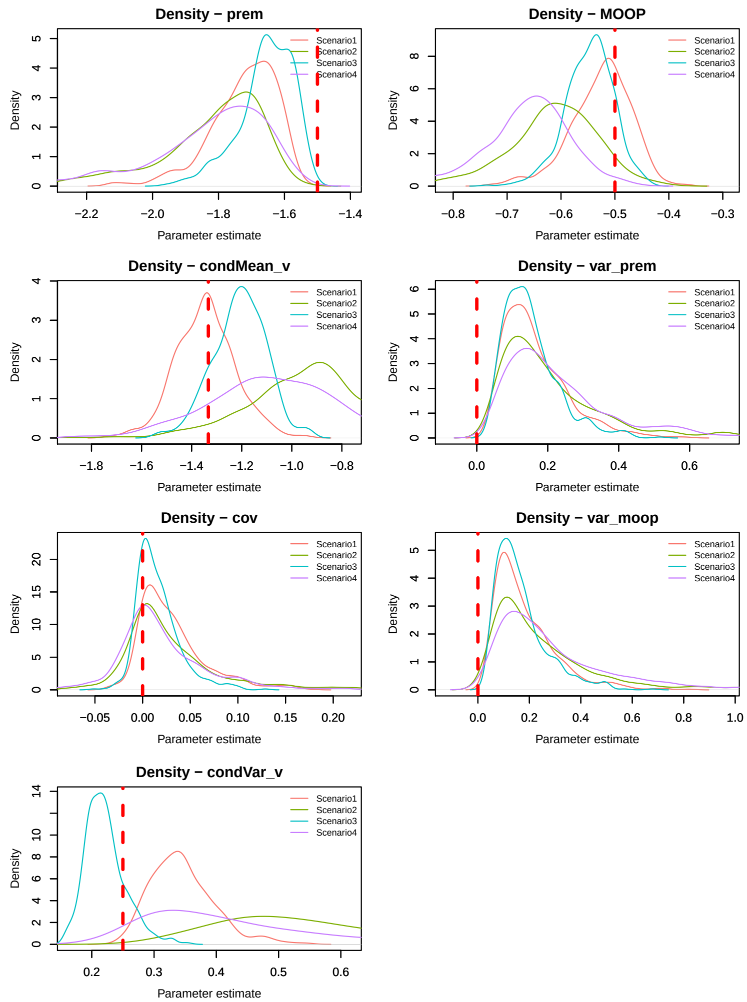
One exception is δ, however, in both high and low levels of unobserved individual heterogeneity. The parameter estimate tends to be biased downward even if the dependent variable is observed individual-level choice of product. In a sensitivity analysis in appendix Appendix D.1 I estimate the same data but vary $R = 50$ or $R = 100$. The coverage of $\delta = 0.5$ is also biased downward across four different scenarios, implying the variation in generating the age distribution provides less information for identification. Also notice that across all scenarios the estimates of variance covariance Σ_β are biased upward if the true values of unobserved individual heterogeneity are nearly zero. Whereas when the true values of Σ_β are high, the estimates using product-level choices are nearly unbiased. The estimates of posterior means of Σ_β could be up to two times of the true value, even if individual choices are available in scenario two.

In addition, when the dependent variable is the choice of multi-product firms in scenario two or four, the bias of posterior mean and standard deviation are in general higher for all parameters compared to those estimates from product-level choices. Yet we might find mixed results by comparing the estimates of the posterior means and standard deviations in scenario 4 to scenario 2. Given that in scenario two we have the true observed choices of firms at individual level, a possible explanation is the same MCMC algorithm is less likely to explore some areas of the true posterior distribution if we have to augment individual choices using portfolio market shares. Indeed, in Figure D.2 to Figure D.4 in Appendix D, we could see the posterior distributions for the estimates in scenario four are more concentrated.

The parameters in all scenarios for the combination of individual heterogeneity are also plotted in Figure 4.3 to Figure 4.9. The results of density plots suggest there are two groups of overlapped densities – one set of scenario 1 and 3, and the other for scenario 2 and 4. The finding is consistent with the fact that scenario 1 and 3 both have product-level choices as the dependent variable, while in scenario 2 and 4 the dependent variable is grouped or aggregated to multi-product firms. The figures also show that the posterior distributions estimated by the choices of firms spread more widely if unobserved individual heterogeneity becomes higher. Lastly, even if we estimate a Bayesian discrete choice model with observed product-level choices, we could see multiple modes of parameters. For example, the density plot of MOOP coefficients in scenario one of Figure 4.5 has two modes– one at the true value of -0.5 and the other around -0.4. In comparison, we can see in Figure D.2 the densities have a peak near the true value and another mode around -1.2. It's possible that given the same data generating process, the posterior distributions have different modes as in scenario two or four the likelihoods are aggregated for multi-product firms.

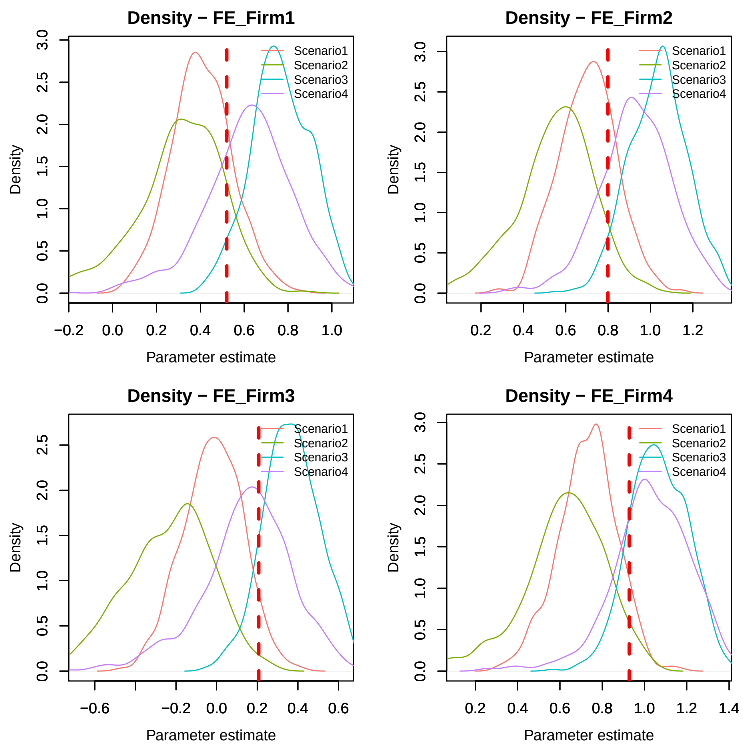
In summary, by simulation we can compare the performance of empirical models. When the dependent variable is the true individual choice of firm aggregated by the firm's multiple product portfolio, the posterior distributions could be biased and have widely standard deviations compared to those estimated by product-level choices at individual level. Yet when only aggregated portfolio market shares are available, we could see the posterior distributions behaves similarly to those estimated by true choices of firms at individual level. In addition, the simulation result for scenario four, main advancement of empirical estimation strategy in this dissertation, could cover the true value of parameters in different degrees of (un)observed individual heterogeneity. Yet the cost is, as expected, larger variances and more biased estimates of posterior means compared to those calculated from known choices at the individual level, aggregated or not. Yet in many empirical applications the exact individual choices at the product level, or aggregated shares at the product level are not available. In next section I demonstrate the use of the Bayesian estimation procedure using portfolio market shares with real world data.

Figure 4.3: Posterior Distributions, Low Variance, $\delta = 0, N = 200$



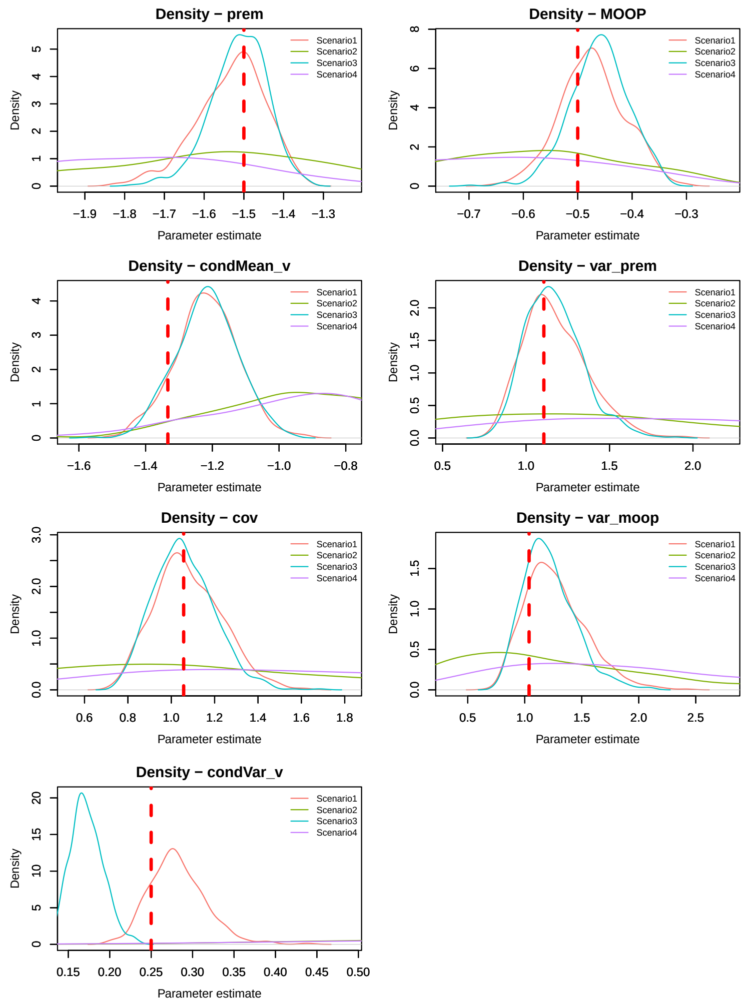
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for low unobserved heterogeneity of preference weights.

Figure 4.4: Posterior Distributions of Firm Fixed Effects, Low Variance, $\delta = 0, N = 200$



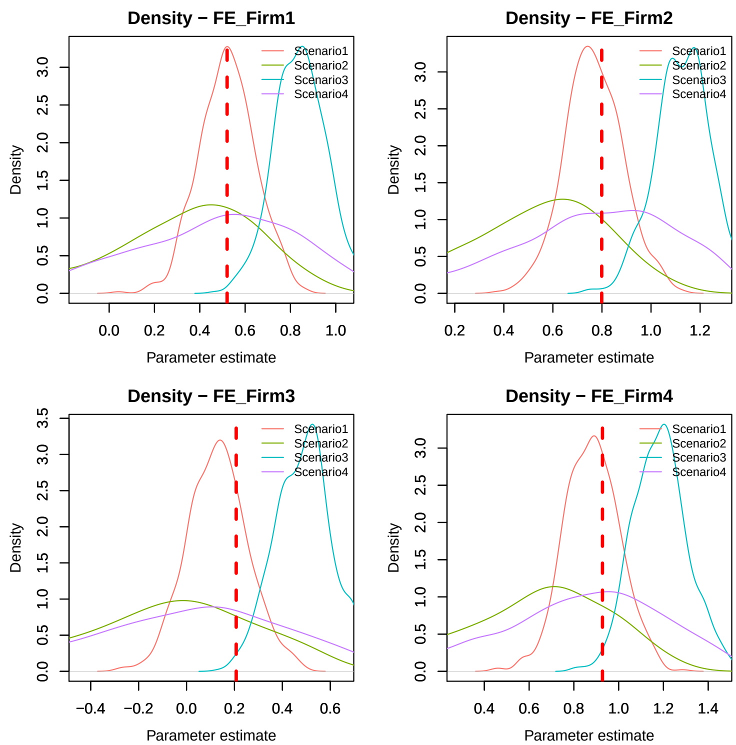
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $N = 200$.

Figure 4.5: Posterior Distributions, High Variance, $\delta = 0, N = 200$



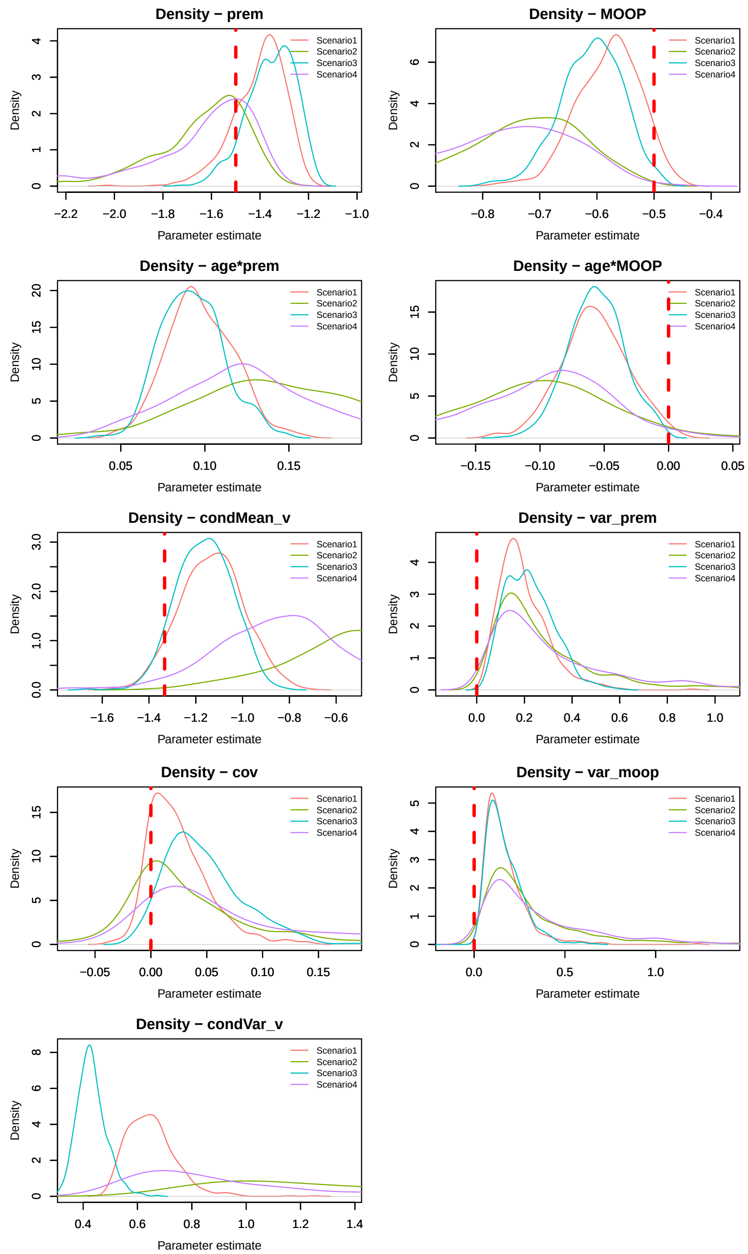
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights.

Figure 4.6: Posterior Distributions of Firm Fixed Effects, High Variance, $\delta = 0, N = 200$



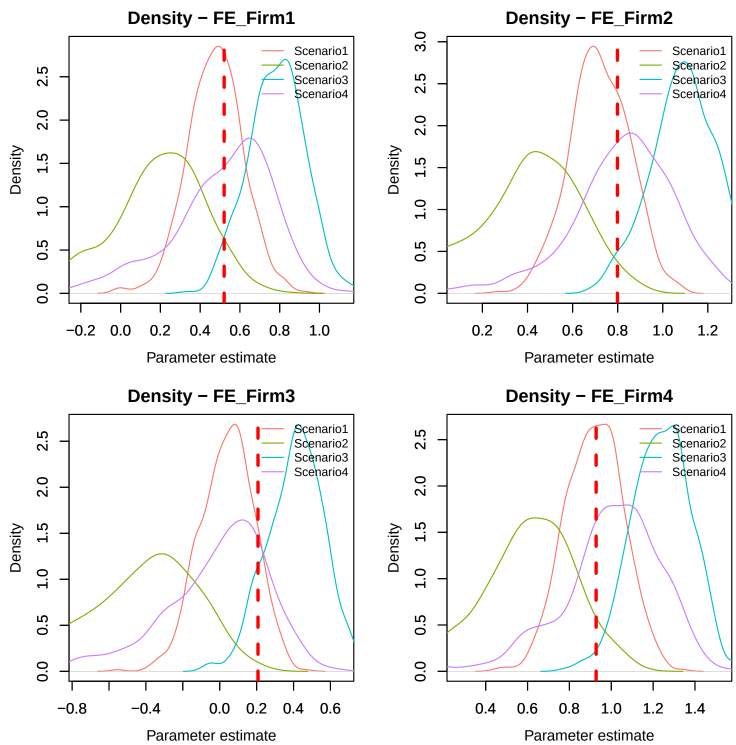
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $N = 200$.

Figure 4.7: Posterior Distributions, Low Variance, $\delta = 0.5, N = 200$



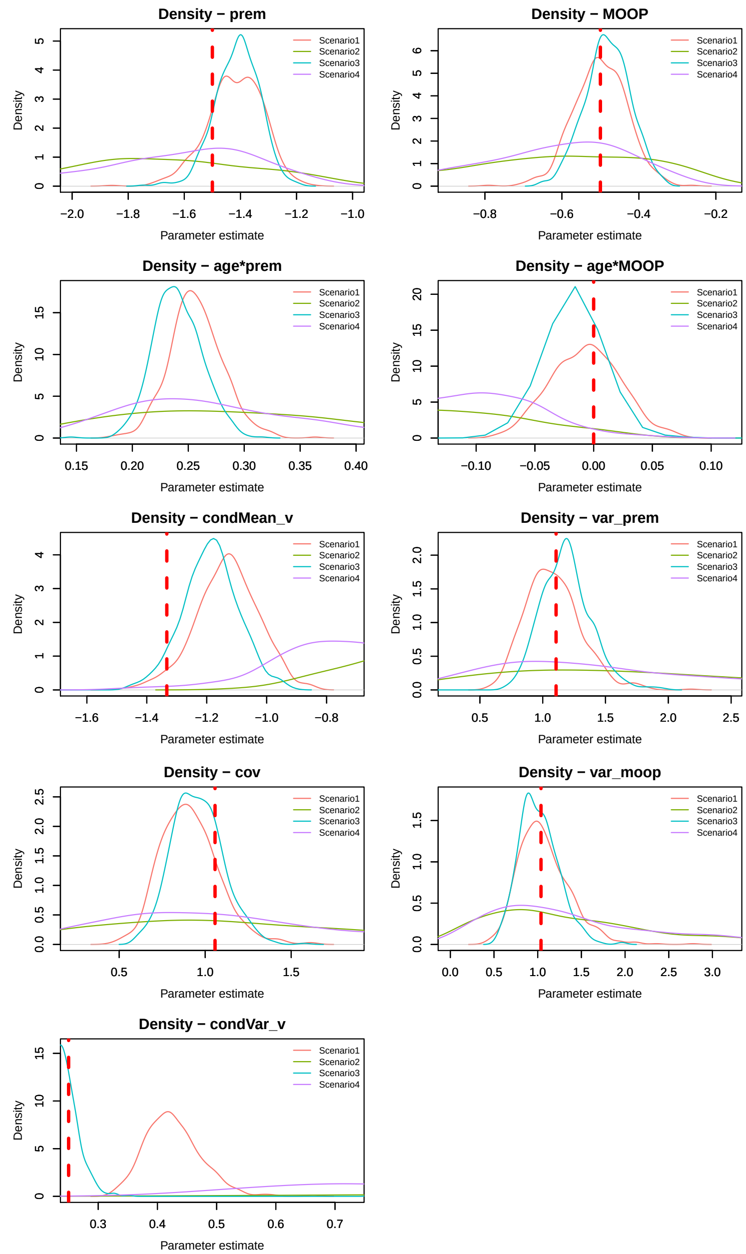
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for low unobserved heterogeneity of preference weights.

Figure 4.8: Posterior Distributions of Firm Fixed Effects, Low Variance, $\delta = 0.5, N = 200$



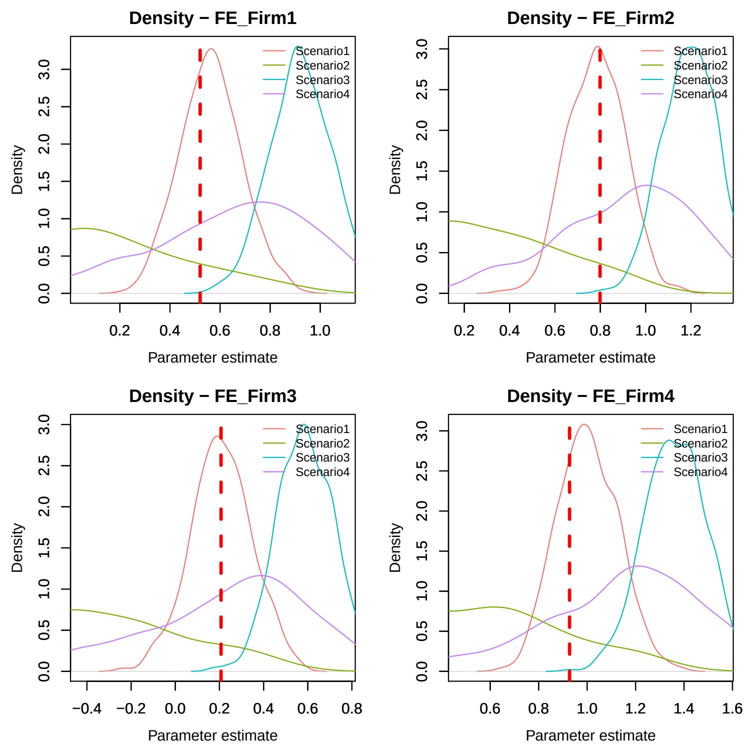
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $N = 200$.

Figure 4.9: Posterior Distributions, High Variance, $\delta = .5, N = 200$



Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights.

Figure 4.10: Posterior Distributions of Firm Fixed Effects, High Variance, $\delta = 0.5, N = 200$



Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $N = 200$.

Chapter 5

Bayesian Demand Estimation in CoveredCA

5.1 Estimating Demand in Health Insurance Marketplaces

The model specified in chapter Section 4.4 can be used to estimate the consumers' demand of health insurance in Federally Facilitated Marketplaces (FFMs) when portfolio market shares are the only dependent variables available to researchers. The market-level data of enrollments in FFMs can be downloaded from a unified system, HealthCare.gov, along with characteristics of qualified health plans (QHPs) in those states. In this chapter, however, I replicate estimates of demand function using CoveredCA and compare the results in literature to demonstrate the potential of my new empirical method. In next chapter I estimate demand of health insurance in Pennsylvania using data from HealthCare.gov website.

In general, a potential buyer of health insurance in a state can choose to purchase a QHP from the state's Marketplace online or through brokers. Or the individual can choose to opt out and get uninsured. The option of being uninsured is indexed as $j = 0$. When the potential enrollee stays uncovered, she incurs a cost and pays penalty due to individual mandate that is calculated by her family status and income level as required by ACA from 2014 to 2018.

If she chooses to get insured, she can shop standardized QHPs marketed by different metal-tiers in her Marketplace. A metal-tier represents actuarially fair value that insurers are expected to pay for the enrollee's medical expenses. In general, the more valuable a metal plan is, the higher is the premium of the QHP and more generous the benefits. One of the metal-tiers, silver plan, are often selected by the enrollee[1] because they provide additional cost-sharing reductions (CSRs) that decrease deductibles, co-payments and co-insurance rates. However, the CSRs are only applicable for the enrollee whose income level is less than about 250% of federal poverty level (FPL).

Regardless of metal-tiers, the enrollee faces a menu of multiple QHPs within her residential areas. The geographic location along with her age, family status and smoking status jointly determines her premium for each health insurance plan. In addition, if her income level, is less than 400% of the FPLs, the potential enrollee will have a premium tax credit (PTC) that can be applied to any plan available the Marketplace. Assume for simplicity the potential buyer maximizes her utility given income level and PTC, τ_i, she maximizes her utility by choosing one of plans $j = 1, ...J$ in her residential area during the open enrollment period below [2],

$$
\begin{aligned}
u_{ijtm} =& \beta_{prem,itm}(prem_{ijtm} - \tau_{itm}) + \beta_{MOOP} MOOP_j + \alpha_{f|j} + \xi_{jtm} + \epsilon_{ijtm}, \\
u_{i0tm} =& \beta_{price,itm} mandate_{itm} + \epsilon_{i0tm}, \\
\beta_{prem,itm} =& \mu_{prem} + \delta_{inc} inc_{itm} + \delta_{age} age_{itm} + \delta_{family} family_{itm} + e_{itm},
\end{aligned}
$$

where inc is income in terms of FPL and t stands for a year while m is a market.

[1] See HelathCare.gov (Accessed July 15, 2020).

[2] An open enrollment period lasts from Nov. 1st in the previous year to Jan. 31st of the enrollment year.

Notice that Affordable Care Act allows premiums to vary by individual demographics while PTCs and mandate are constant for all QHPs available to the enrollee. The utility function can be rewritten by normalizing the outside option. Thus I will estimate

$$\begin{aligned} u_{ijtm} =&(\mu_{prem} + e_{itm})p_{ijtm} + \delta_{age} age_{itm} p_{ijtm} + \delta_{family} family_{itm} p_{ijtm} + \\ &\delta_{inc} inc_{itm} p_{ijtm} + \beta_{MOOP} MOOP_j + \alpha_{f|j} + \xi_{jtm} + \epsilon_{ijtm}, \qquad (5.1) \\ p_{ijtm} =&(prem_{ijtm} - \tau_{itm} - mandate_{itm}), \\ u_{i0tm} =&\epsilon_{i0tm}. \end{aligned}$$

The relative price of a plan is the premium adjusted for the demographic characteristics minus the PTC and individual mandate. The two together are the opportunity cost of being uninsured. By rewriting the utility function I can estimate the set of δ and β_{MOOP} together as a block in the same step of MCMC using Metropolis-Hasting algorithm.

Let $y_{ijtm} = 1$ if $j = 0, 1, ...J_{tm}$ is chosen. By aggregating all consumers' choices, we can derive market shares at either product or firm level by

$$s_{jtm} = \sum_{i \in m}^{M} y_{ijtm}, \text{ Aggregated product-level shares.} \qquad (5.2)$$

$$s_{ftm} = \sum_{j \in J_f} s_{jtm}. \text{ Aggregated firm-level shares.} \qquad (5.3)$$

When the researcher has individual-level choices $\{y_{ijtm}\}$ directly observed for each consumer, demand function with unobserved consumer heterogeneity is often estimated by random-coefficient logit at individual level. In some case when the researcher observes product-level shares, or if she aggregates individual choices to product levels to study market structures, a BLP-type model at the product-level shares is often applied. For example, Tebaldi (2017) studies insurers' competition and the premium designs in Covered California (CoveredCA) by metal-level shares. He equates metal-level shares to product-level shares as QHPs in California are highly standardized by actuarially fair values. As his utility specification is much simple and data for all but the enrollments at individual level are available, I am about to compare the demand estimates using portfolio market shares in CoveredCA as my dependent variables.

5.2 Data

The eligible population of potential enrollees in all Marketplaces including CoveredCA can be constructed by following steps. First, I collect surveyed individual-level data with demographic characteristics from IPUMS-USA (2020) from 2012 to 2016. The integrated survey data harmonize households from American Community Survey (ACS) and Current Population Survey (CPS). Using the database, I have the information such as: an identifier of household heads, number of dependents in households, ages, income levels in terms of federal poverty lines (FPLs), citizenship, being insured by employer sponsered health insurance, public insurance or not, and the household weights to represent the whole populations. These demographic characteristics allow me to select the individuals that are potential ineligible for the Marketplaces if they are covered by Medicare, Medicaid[3], or if they are not US citizens. Furthermore, I can model the choice decisions for getting insured and plan selections at household level. That is, the coverage of health insurance is determined by the head of a household as a representative agent. And since one of the special features in Marketplaces is how premiums of plans are adjusted by ages, I can calculate the total premiums for all plans in a household by multiplying a corresponding age factor. In addition, by using IPUMS-USA, I can have individual income in terms of continuous federal poverty lev-

[3] I exclude individual eligible for Medicaid state-wise if a state decides to expand Medicaid.

els. This information helps me compute not only individual mandate per household, but the corresponding cost sharing reductions that makes silver plans more generous if the household's income is below 250% FPL. The construction of eligible households and premium tax credits follows Drake and Abraham (2019) and is described below.

After collecting the household data [4], I convert the data from the level of Public Use Microdata Areas (PUMAs) to five digit ZIP codes using the crosswalk of PUMAs to counties by the most recent MABLE database from Missouri Census Data Center[5]. Then households in five digit county codes are created by the weights of total populations in the crosswalk. The population in five digit county code, combing with geographic rating areas (GRAs) from ACA, delineates a market of interest in my empirical application in CoveredCA.

Next I collect from HIX Compare of Robert Wood Johnson Foundation [6] for premiums and impute the expected subsidy per household as used in Drake and Abraham (2019). The data in HIX Compare has harmonized plan level characteristics in 50 states and DC by GRAs. By that I can extract from the data the premiums of the second lowest silver plans, the necessary information to calculate premium tax credits.

I also collect data from American Community Survey for demographics, from CoveredCA for plan attributes and enrollments in GRAs. There are 19 GRAs in 58 counties and 59 county-GRAs as markets in CoveredCA[7]. The enrollments in those 59 county-GRAs are available from 2014 to 2018.

For demographics, I draw non-repeated samples of 200 individuals from ACS for each of the sampling years. This strategy allows me to have more variations in demographics distributions by markets– an important source of identifying the

[4]Special thanks to Coleman Drake for his suggestions and codes.

[5]http://mcdc.missouri.edu/applications/geocorr.html

[6]https://hixcompare.org/individual-markets.html

[7]Los Angeles has two rating areas. The number of enrollments will be split by the population weights from MABLE.

coefficients of observed heterogeneity on income, family size and age.

The eligible population is constructed through the following steps. First I exclude the individuals whose household income is at and below 138% of federal poverty level as the low income would be covered by Medicaid in CA. Then the expected subsidy follows Drake and Abraham (ibid.) and is imputed by the second lowest silver plan in each county-GRA and adjusted by household sizes and income levels. Those households whose incomes less than 250% FPL are further subsidized by cost sharing reductions if they select silver plans. Therefore the attribute matrices of their silver plans are adjusted to the corresponding actuarially fair value. For example, if a potential enrollee has 150% FPL and her cost sharing reduction would be 80% AV for her silver plans. Then her choice set will have $2,500 deductibles rather than $6,500 in 2017 according to the design of CSRs in CA.

Table 5.1 and Table 5.2 show the demographics within markets and over markets by different time periods, while Table 5.3 shows the market structure by GRAs. The population of interest is individuals eligible for purchasing plans through CoveredCA. That is, an individual can opt out not buying any plan and be uninsured, or she can select one of many plans sold in her residential county-GRA.

An average consumer faces 16 plan choices offered by 3.3 insurers. The averaged premium tax credit for sampled individuals is $351 dollars per month. The potential enrollee expects to pay $99 dollars for individual mandate if she remains uninsured. Of the sampled individuals, 26% of them would be eligible for cost sharing reductions, 53% of them have family, slightly less than 64% would have premium tax credits upon enrollments and three quarters of them are younger than 50 years old.

Table 5.3 also summarizes data available in CoveredCA from 2014 to 2018 included realized portfolio shares by insurers by counties and by years[8]. The percentages of

[8]The major distinction between my thesis and existing literature is, while it is possible to request individual level data in CoveredCA, my approach will estimate demand of insurance in CA using aggregated portfolio market shares.

the uninsured are estimated by the total population in the county-rating areas from ACS. By definition, the counts of the uninsured is the total eligible buyers deducted from observed counts of insured population in CoveredCA. The estimates indicate that, while CoveredCA has been one of the most stable Marketplaces, an estimate of less than 80% of individuals in CA would still remain uncovered by health insurance. Finally, the last three columns in Table 5.3 show the range and mean of the 20% covered individuals. The data variations in the distributions of portfolio shares by multiple product firms across those county-rating areas over time will help identify model parameters.

Table 5.1: Summary Statistics of Demographics, CoveredCA

Rating Area	# of counties	Avg. PTC ($100/month)	Ind. Mandate ($100/month)	Elig. for CSR (%)
1	22	4.16	0.93	29.18
		(0.79)	(0.29)	(6.60)
2	4	2.98	1.09	22.15
		(1.10)	(0.38)	(12.40)
3	4	3.24	1.10	25.30
		(0.70)	(0.41)	(5.48)
4	1	1.83	0.98	15.90
		(0.31)	(0.34)	(3.49)
5	1	2.65	1.10	18.70
		(0.53)	(0.36)	(2.05)
6	1	2.75	1.04	20.50
		(0.40)	(0.35)	(2.00)
7	1	2.60	1.09	16.90
		(0.35)	(0.35)	(3.73)
8	1	2.45	1.11	15.90
		(0.14)	(0.35)	(2.30)
9	3	4.33	1.03	28.10
		(1.14)	(0.37)	(11.66)
10	5	3.51	1.00	29.12
		(0.75)	(0.36)	(9.60)
11	3	3.24	1.04	28.63
		(0.96)	(0.38)	(10.11) -

Rating Area	# of counties	Avg. PTC ($100/month)	Ind. Mandate ($100/month)	Elig. for CSR (%)
12	3	2.84	1.07	21.67
		(0.49)	(0.36)	(3.61)
13	3	4.17	0.88	25.03
		(1.17)	(0.29)	(5.99)
14	1	3.49	0.99	29.20
		(0.70)	(0.33)	(2.89)
15	1	1.97	0.97	26.00
		(0.16)	(0.36)	(3.35)
16	1	2.08	0.98	24.60
		(0.14)	(0.34)	(2.79)
17	2	2.39	0.98	27.00
		(0.31)	(0.33)	(4.44)
18	1	2.25	1.11	17.50
		(0.19)	(0.38)	(4.30)
19	1	2.56	1.06	24.30
		(0.24)	(0.40)	(2.68)
Total	58	3.51	0.99	26.29
		(1.06)	(0.33)	(8.13)

Note: Standard deviation are in parentheses. Statistics are calculated over 200 sampled individuals for each rating-area-county-year. Five year data from 2014 to 2018 are observed.

Table 5.2: Summary Statistics of Demographics, CoveredCA, Continued

Rating Area	# of counties	Family Sz (>1 ppl,%)	Elig. for PTC Subsidies (%)	Senior (≥50 y.o., %)
1	22	54.13	71.02	31.08
		(10.98)	(8.62)	(7.68)
2	4	51.87	49.07	30.90
		(9.16)	(12.63)	(5.55)
3	4	57.15	55.35	24.57
		(12.73)	(10.13)	(7.42)
4	1	31.10	45.20	20.40
		(4.25)	(4.78)	(4.38)
5	1	53.10	48.50	21.50
		(5.21)	(4.60)	(1.41)
6	1	44.90	53.70	18.60
		(5.75)	(3.58)	(2.72)
7	1	54.10	48.70	23.90
		(5.45)	(3.29)	(2.77)
8	1	46.40	42.20	28.80
		(3.25)	(3.09)	(6.23)
9	3	51.63	60.73	26.00
		(8.42)	(11.81)	(3.85)
10	5	58.24	68.16	22.70
		(7.95)	(8.82)	(5.92)
11	3	55.93	68.70	26.30
		(11.36)	(13.11)	(13.73)

Rating Area	# of counties	Family Sz (>1 ppl,%)	Elig. for PTC Subsidies (%)	Senior (≥50 y.o., %)
12	3	53.33	55.00	24.43
		(5.28)	(7.53)	(4.95)
13	3	48.67	70.17	27.90
		(10.40)	(6.50)	(8.13)
14	1	55.50	71.90	20.40
		(8.13)	(4.16)	(4.11)
15	1	43.00	61.80	16.70
		(4.18)	(7.62)	(4.22)
16	1	44.00	59.10	17.90
		(3.22)	(1.39)	(2.53)
17	2	51.85	64.90	20.35
		(4.46)	(4.63)	(4.04)
18	1	52.70	50.00	20.10
		(2.08)	(3.91)	(3.13)
19	1	50.90	59.60	17.10
		(4.52)	(4.76)	(4.17)
Total	58	52.99	63.65	26.76
		(10.22)	(12.27)	(8.18)

Note: Standard deviation are in parentheses. Statistics are calculated over 200 sampled individuals for each rating-area-county-year. Five year data from 2014 to 2018 are observed.

Table 5.3: Summary Statistics of Market Structure, CoveredCA

Rating Area	# of Insurers	# of Plans	Uninsured (%)	Portfolio Shares (%) max	mean	min
1	2.2545	11.9818	0.8008	0.1549	0.0935	0.0348
	(0.5656)	(3.3537)	(0.0775)	(0.0640)	(0.0486)	(0.0535)
2	4.8000	22.4000	0.7917	0.1214	0.0439	0.0022
	(0.4104)	(1.3917)	(0.0557)	(0.0545)	(0.0131)	(0.0014)
3	4.3500	23.2000	0.8073	0.0983	0.0446	0.0034
	(0.5871)	(3.7219)	(0.0579)	(0.0523)	(0.0131)	(0.0034)
4	5.2000	24.2000	0.7289	0.1033	0.0522	0.0039
	(0.4472)	(3.6332)	(0.0487)	(0.0348)	(0.0093)	(0.0041)
5	4.0000	19.0000	0.8016	0.1235	0.0522	0.0021
	(0.7071)	(2.0000)	(0.0399)	(0.0408)	(0.0184)	(0.0005)
6	2.8000	14.2000	0.7413	0.1520	0.0961	0.0438
	(0.4472)	(1.7889)	(0.0401)	(0.0516)	(0.0312)	(0.0242)
7	4.8000	24.6000	0.7808	0.1054	0.0462	0.0080
	(0.4472)	(2.6077)	(0.0372)	(0.0144)	(0.0102)	(0.0088)
8	4.8000	22.2000	0.7994	0.1199	0.0425	0.0047
	(0.4472)	(1.3038)	(0.0455)	(0.0444)	(0.0125)	(0.0022)
9	2.6667	13.6667	0.7982	0.1396	0.0824	0.0286
	(0.8165)	(4.0473)	(0.0377)	(0.0292)	(0.0308)	(0.0505)
10	3.5200	17.2800	0.7760	0.1715	0.0674	0.0036
	(0.7703)	(3.9950)	(0.0727)	(0.0715)	(0.0281)	(0.0063)
11	3.2667	17.0667	0.8508	0.0791	0.0496	0.0196
	(0.8837)	(4.9348)	(0.0524)	(0.0302)	(0.0256)	(0.0273)

Rating Area	# of Insurers	# of Plans	Uninsured (%)	Portfolio Shares (%) max	mean	min
12	2.6000	13.6000	0.8157	0.1169	0.0743	0.0282
	(0.6325)	(3.5617)	(0.0408)	(0.0492)	(0.0235)	(0.0339)
13	2.5333	12.0667	0.6922	0.2224	0.1386	0.0797
	(1.0601)	(4.5586)	(0.1549)	(0.1452)	(0.0912)	(0.1107)
14	3.8000	18.4000	0.8816	0.0571	0.0319	0.0119
	(0.4472)	(2.1909)	(0.0272)	(0.0141)	(0.0103)	(0.0125)
15	6.0000	30.4000	0.6454	0.1137	0.0591	0.0131
	(0.0000)	(2.6077)	(0.0434)	(0.0127)	(0.0072)	(0.0093)
16	6.4000	32.0000	0.7290	0.0866	0.0424	0.0050
	(0.5477)	(3.7417)	(0.0367)	(0.0104)	(0.0058)	(0.0044)
17	4.8000	25.4000	0.8376	0.0511	0.0344	0.0123
	(0.4216)	(1.5776)	(0.0354)	(0.0140)	(0.0095)	(0.0057)
18	4.8000	25.6000	0.8322	0.0666	0.0354	0.0088
	(0.8367)	(5.5498)	(0.0238)	(0.0112)	(0.0050)	(0.0081)
19	5.8000	29.4000	0.8454	0.0398	0.0269	0.0070
	(0.4472)	(1.6733)	(0.0137)	(0.0032)	(0.0038)	(0.0063)
Total	3.3051	16.8203	0.7937	0.1337	0.0735	0.0238
	(1.3281)	(6.5884)	(0.0796)	(0.0723)	(0.0472)	(0.0473)

Note: Data from 2014 to 2018 in CoveredCA. Standard deviations are in parentheses.

5.3 Model Specification and Identification

I specify the empirical model similar to Tebaldi (2017) for demand estimation on enrollments in CoveredCA. To perfectly replicate his results, however, I would have used actuarially fair values as one of the important explanatory variable. This is because in CoveredCA plan characteristics are highly standardized so that two plans of the same metal tier are equivalent in all but health insurers and unobserved networks (to me). The low variations in plan characteristics when I estimate the model using AV could result in difficulty in identifying related preference weights for the demand model using aggregated portfolio market shares. Instead, I use maximum out-of-pocket expenditure in my specification – the MOOP is standardized among plans but varies over time. In addition, low income individuals would qualify cost sharing reductions that change MOOP of silver plans.

The specification of the demand system consists of choices of an insurance plan j if a potential enrollee gets insured or opts out the insurance ($j = 0$). In the latter case she pays individual mandate and losses the opportunity to get subsidized.

While plan characteristics have low variations from the standardized plan design, the market structure in CA does provide extra data variations for identifying some parameters. For example, given that multiple counties are in the same rating area with the same base premiums, the variations in portfolio market shares by health insurers in those counties can be explained by preference weights for product attributes. That is, holding products offered and the participation of health insurers constant, the variations in portfolio shares will be explained by the data variations in demographic distributions. Also, within the same GRA, it is likely that some insurers can choose to enter local markets. For example, I observe that there is only one monopoly insurer in some counties in rating area one, while other counties of the same rating area have multiple insurers. The variations in the size of firms within the same rating area also can contribute to model identifications. Specifically, the unobserved heterogeneity of premium can be identified if the portfolio market shares violate the independent irrelevant alternative assumption of multinomial logistic models.

Yet another difficulty using portfolio market shares is I might not observe the joint distribution of demographic attributes and portfolio market shares, or the exact choices of insurers at individual level by demographic groups. To solve these I have to augment data of individual choices with demographic characteristics and address label switching issues (Stephens, 2000) in Bayesian literature. That is, given a set of fixed demographic draws from ACS, the assignments of augmented choices to the individuals have different implications on demand estimates. If two markets have the same set of insurers and products but there exist difference in the distribution of potential enrollees' income levels, then the coefficients on the income levels can be identified if I observe variations in portfolio shares. Yet by assigning the labels of chosen portfolios, low income group and high income group could have selected products differently, if the assignments are fixed.

Two possible solutions are available for the label switching problem when portfolio shares and demographics are independently observed. First is to draw a sequence of choices consistent with portfolio market shares. Next I can treat the sequences as fixed data, and in each step of the MCMC a new set of demographic attributes consistent with empirical distributions are proposed and accepted by a MH algorithm. While this strategy is feasible, the algorithm requires updating all demographic characteristics at once and whose computational time can increase significantly when the dimension of demographic variables and the counts of simulated population increase.

Alternatively, I follow the panel data approach in Musalem et al. (2009) that treat augmented choices as extra parameters while I treat a set of eligible individuals with demographics drawn from ACS as known data. Then in each iteration of MCMC only a subset of individual choices in the same market are updated by

Gibbs sampling. This method potentially can speed up the computational time as the dimension of simulated data only grows linearly with total simulated individuals.[9] However, it should be noticed that the label switching problem occurs in particular in the Bayesian framework that augments individual choices. Berry et al. (2004), Petrin (2002) and Nevo (2000a) all estimate demand models and impose extra moment conditions for demographic characteristics. The label switching is alleviated by integrating out the probabilities over all drawn demographic characteristics. That is, a set of predicted averaged probabilities are computed to match the observed market shares. A hybrid method in the Bayesian approach that uses both augmented choices with demographics and averaged probabilities can be seen in Yang and Chen (2003).

5.4 Estimates from CoveredCA

I estimate a demand model specified in Section 5.1 and address the issue of potentially endogenous premiums using Hausman instruments (Hausman, 1996). That is, I assume the premium vector of a firm is correlates with competitors' pricing decisions. Hence the premium of a plan can be predicted by competitors' premiums of similar products. The instruments are constructed using the average premiums of silver plans and include dummies of year and rating areas. Next I run a first stage linear regression of premiums on the instruments and predict a vector of η_j by the residuals. The predicted residuals are treated as data and plugged into the demand function. While it's possible to treat η_j as data and let the residuals fixed over all iterations of MCMC, the vector can also be updated iteratively once the demand estimates are drawn. See Appendix C for the detailed algorithm.

[9] If I have the joint distributions of demographics and portfolio market shares, then I could assign the sequences by the joint distributions and treat both demographics and choices as fixed data. This is a strategy similar to Musalem et al. (2008) that impose extra constraints for data augmentations. The main disadvantage of this approach, however, is it restricts MCMC to search in narrower parameter space. That could potentially limit the ability of model convergence of MCMC.

The fixed effects in the demand specification includes four large insurance companies, Anthem, Blue Cross in California, Kaiser, and Health Net. The rest insurers are grouped together as a category, as is handled similarly in Tebaldi (2017) [10]. The demand of health insurance plans is estimated by the Bayesian algorithms in Appendix C. A total of 800,000 iterations are conducted to draw the joint posterior distribution and the first batch of 500,000 iterations were discarded for burn-in periods. The rest 300,000 draws were kept with every 100th draw used in final analyses to reduce autocorrelations among estimated MCMC parameters.[11]

In Figure 5.1 and Figure 5.2 we can observe the trace plots and posterior densities for the parameters using MCMC. From the trace plots, it seems the draws are mixed relatively well as each parameter is not demonstrating an upward or downward trend. The parameters do show multiple modals because of augmented individual choices for insurers from portfolio shares. For example, the mean premium coefficient in Figure 5.1 has two possible modes at -1.95 and -2.05.

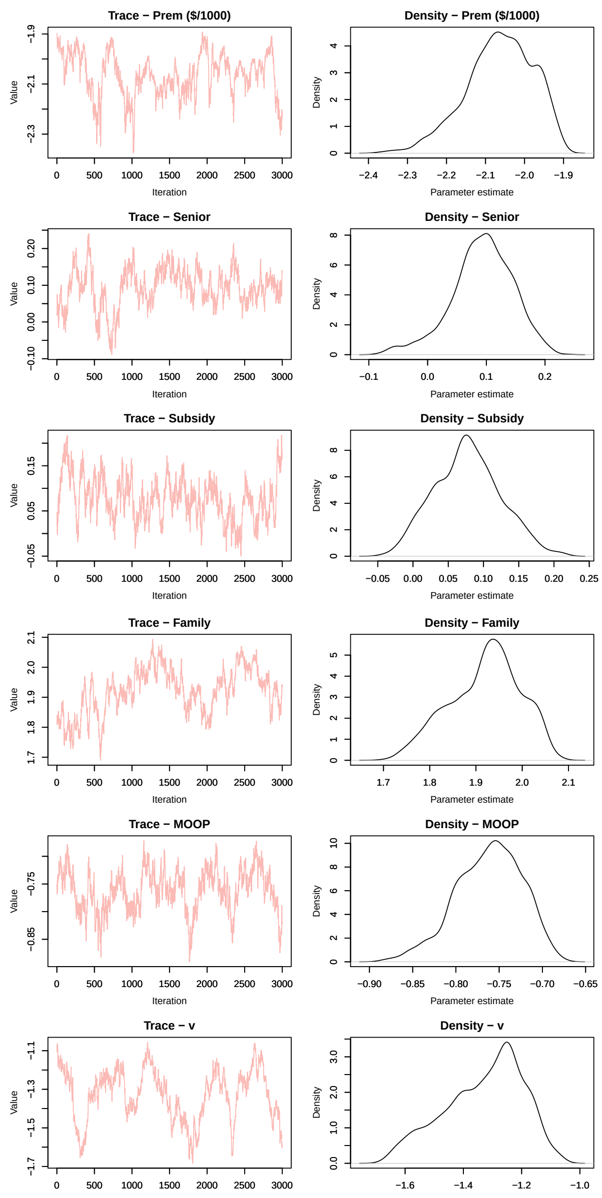
The signs of parameter estimates are mostly consistent with those in literature. The premium coefficient has a negative effect on utility, while the effect becomes smaller in magnitude if an individual is senior or has family. The estimate of subsidy status seems paradoxical as literature suggests it has a strong negative effect (Saltzman, 2019). While in the density plot of the subsidy only a small fraction of the draws has negative coefficients.

The effect of MOOP is negative and suggests the individual dislike higher maximum out-of-pocket expenditure. And the effect on predicted residue is negative, im-

[10] However, I exclude interactions of insurer fixed effects with demographic groups for estimation convenience

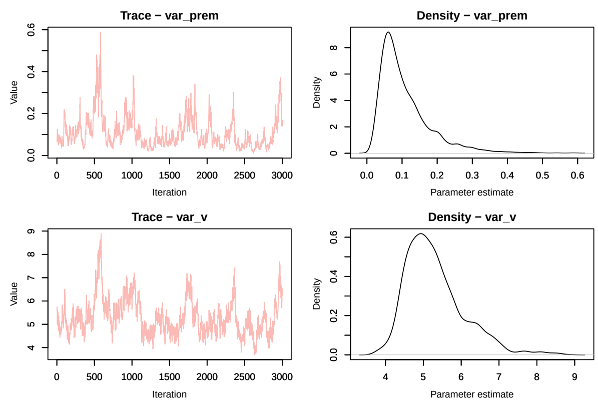
[11] Alternatively, Carlin & Louis (2008) argue thinning draws would instead throwing away useful estimates. This is in particular a concern if researchers want to reduce the variances of MCMC draws. But practically thinning draws requires much less physical memories and hence might potentially speed up computational time. The run time is especially important as it took me at least one week for 800,000 iterations of a single chain MCMC on a local machine with 16 GB physical memory for 5 years, 59 county-GRAs, and 200 simulated individual each market-year in CoveredCA.

Figure 5.1: Posterior Distributions for Demand Estimates in CoveredCA using Portfolio Market Shares



plying that the endogenous part of premiums correlates negatively with unobserved product attributes. Finally, the variance of unobserved heterogeneity in premium is about 0.1. Interestingly, the magnitude is similar to Ericson and Starc (2012) who

Figure 5.2: Posterior Distributions for Demand Estimates in CoveredCA using Portfolio Market Shares, Continued



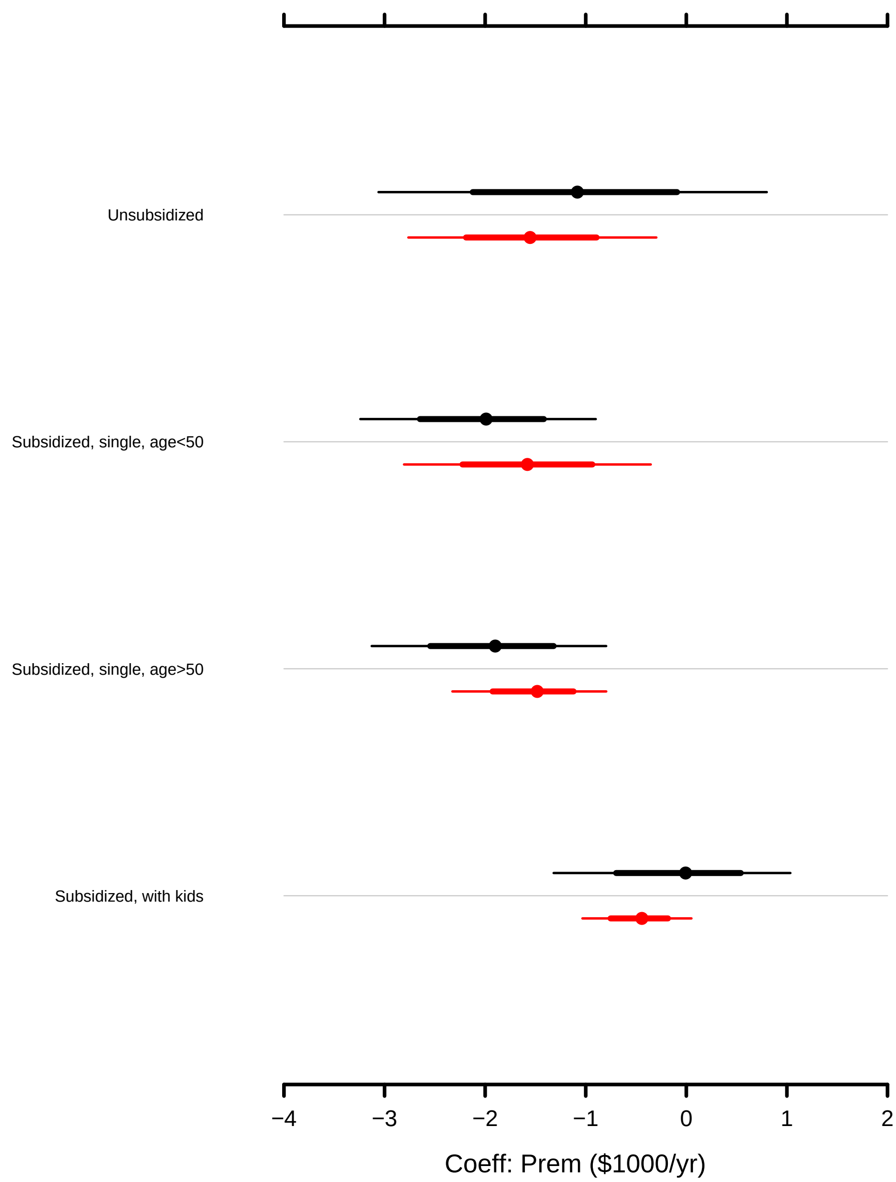
Left: Trace plots for thinned MCMC draws; Right: Sampling Distributions.

indicate unobserved heterogeneity plays an essential role in Marketplaces.[12] The estimated variance of the premium coefficient would be smaller if I add more demographic characteristics so that observed heterogeneity of premiums are accounted for.

The posterior draws of the premium coefficients are calculated for each of the four demographic groups as shown in Tebaldi's (2017, table 9) for comparison. In Figure 5.3, the black lines are 90% confident intervals from predicted premium coefficients in estiming the demand using portfolio market shares while the red lines are the 90% confidence intervals from Tebaldi's study that uses granular individual-level data. Interestingly, although the 90% confident intervals across all demographic groups have good overlaps, the median estimates for subsidized, single population, regardless of age groups, seem to be almost larger in absolute value using the proposed Bayesian approach. My premium coefficients for the unsubsidized and subsidized family population tend to attenuate toward zero in comparison, however.

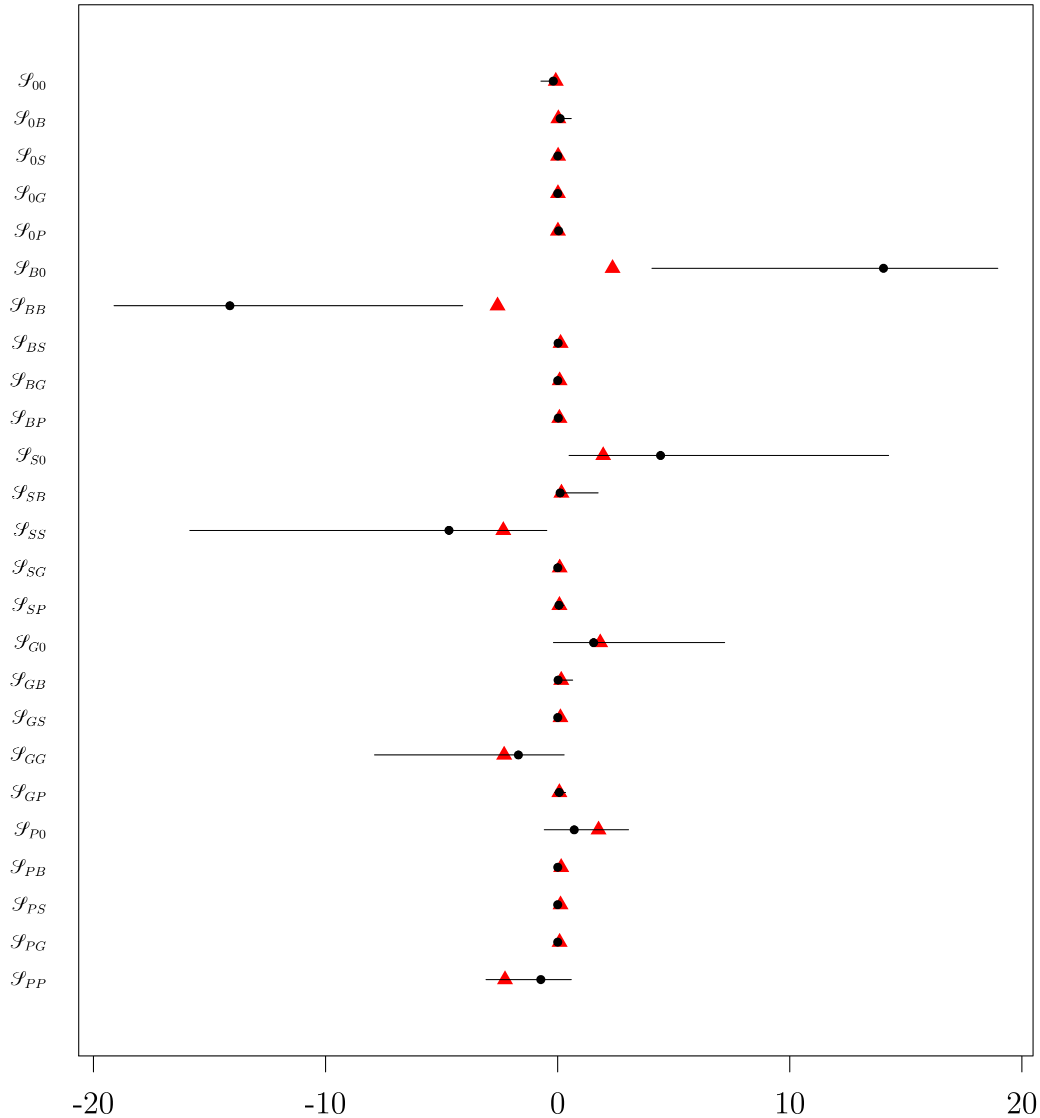
[12] As in my estimation the unit is \$1,000, this translates to a standard deviation 3.16 per \$100 ($=\sqrt{0.1} \times 10$) and my estimate would be close to the estimate of Ericson and Starc (ibid).

Figure 5.3: Distributions of Predicted Premium Coefficients by Demographics



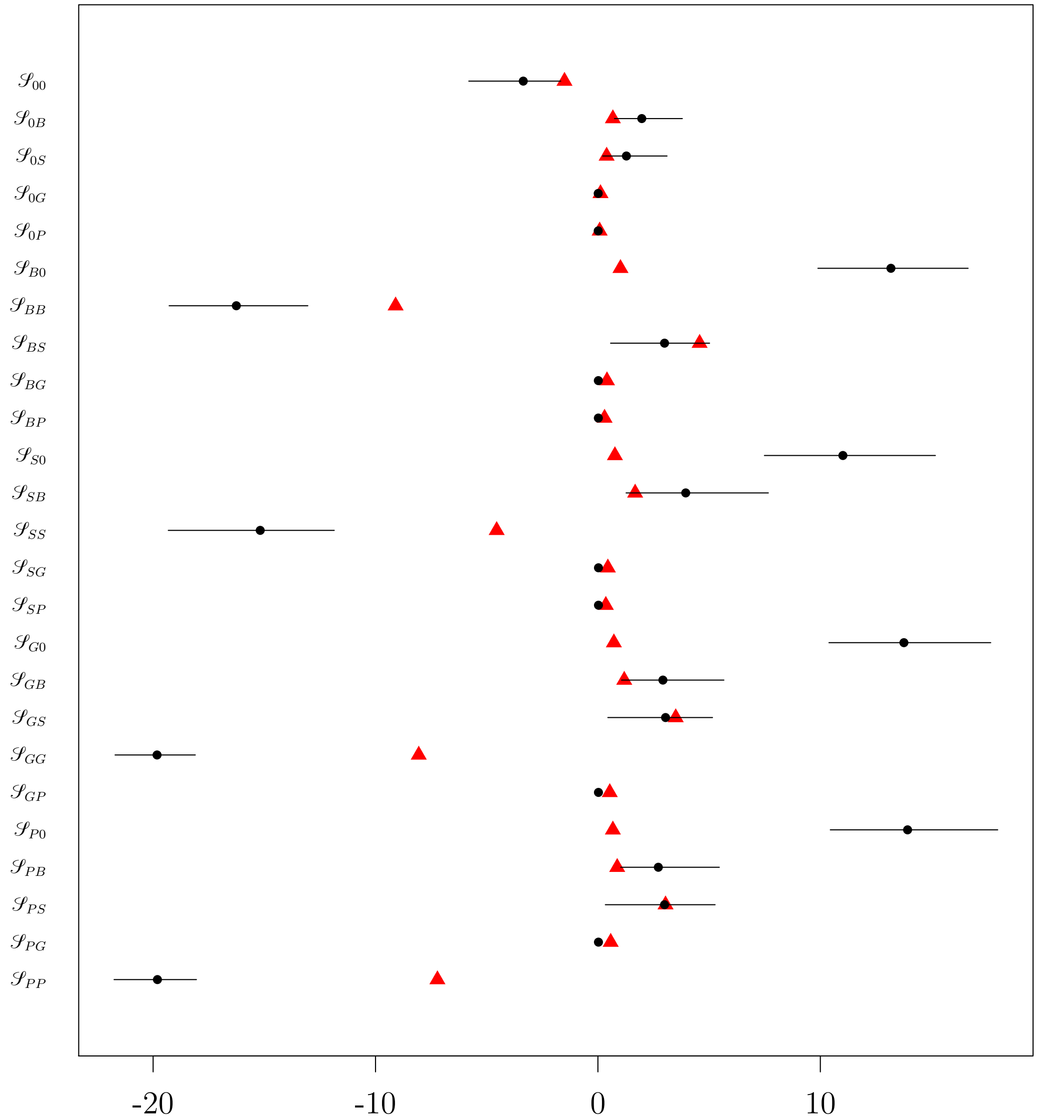
Black horizontal line: predicted 90% confidence (credible) interval estimated by portfolio market shares. Red horizontal line: predicted 95% confidence interval in Tebaldi (2017). Solid circles are medians and thicker intervals are 50% confidence set.

Figure 5.4: Premium Semi-Elasticities, Unsubsidized Population



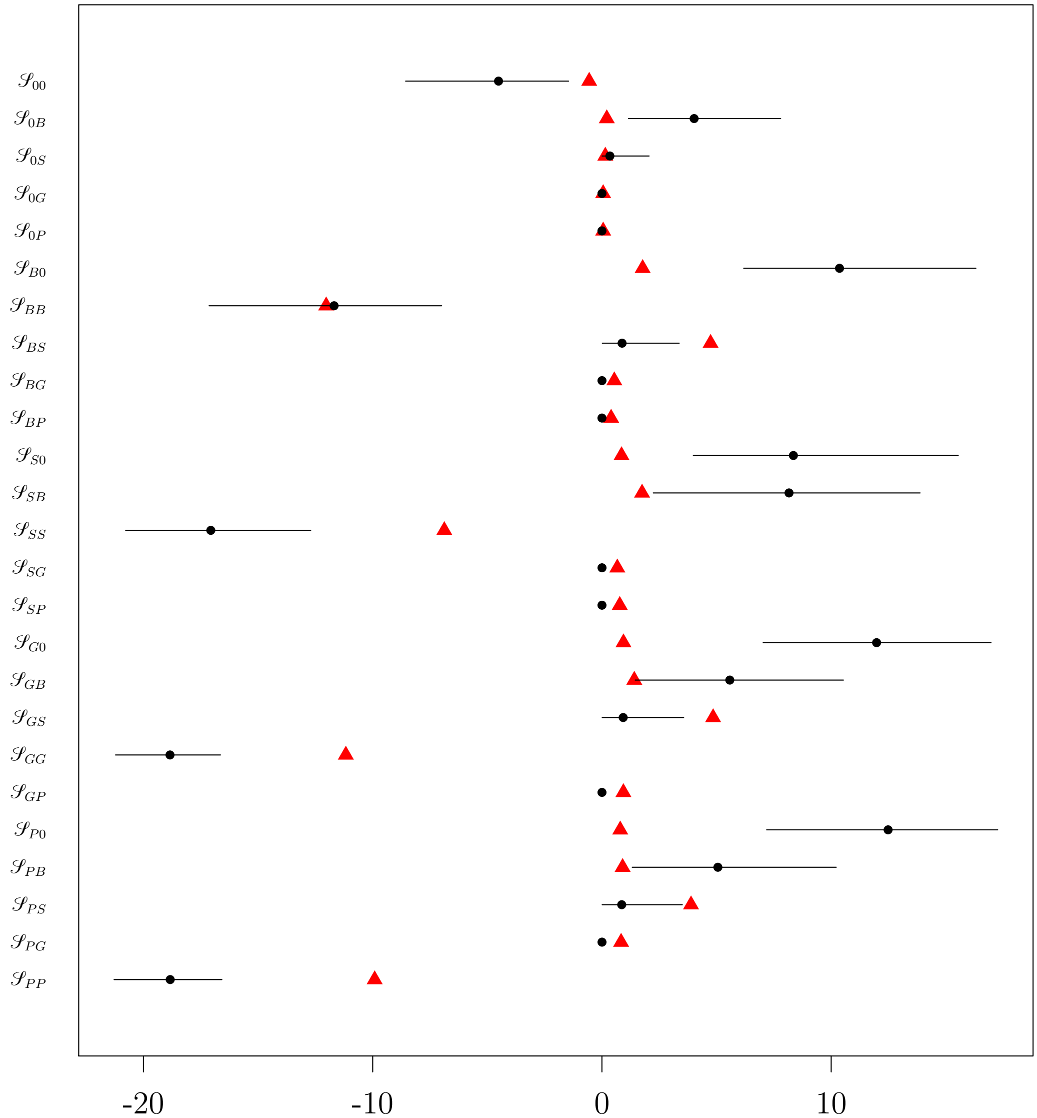
Note: Semielasticities for \$100 change in premiums. $\mathscr{S}_{ab}$ is the % change in predicted shares for good a with respect to \$100 change in good b. The unsubsidized have income greater than 400% federal poverty line. Abbreviations are 0 : Uninsured, B : Bronze, S : Silver , G : Gold, P : Platinum. Black lines are estimates of 95% confidence intervals from models using portfolio market shares; black circles stand for medians while red triangle is the median estimate from Tebaldi (2017, table 10).

Figure 5.5: Premium Semi-Elasticities, Subsidized, Single and Young Population



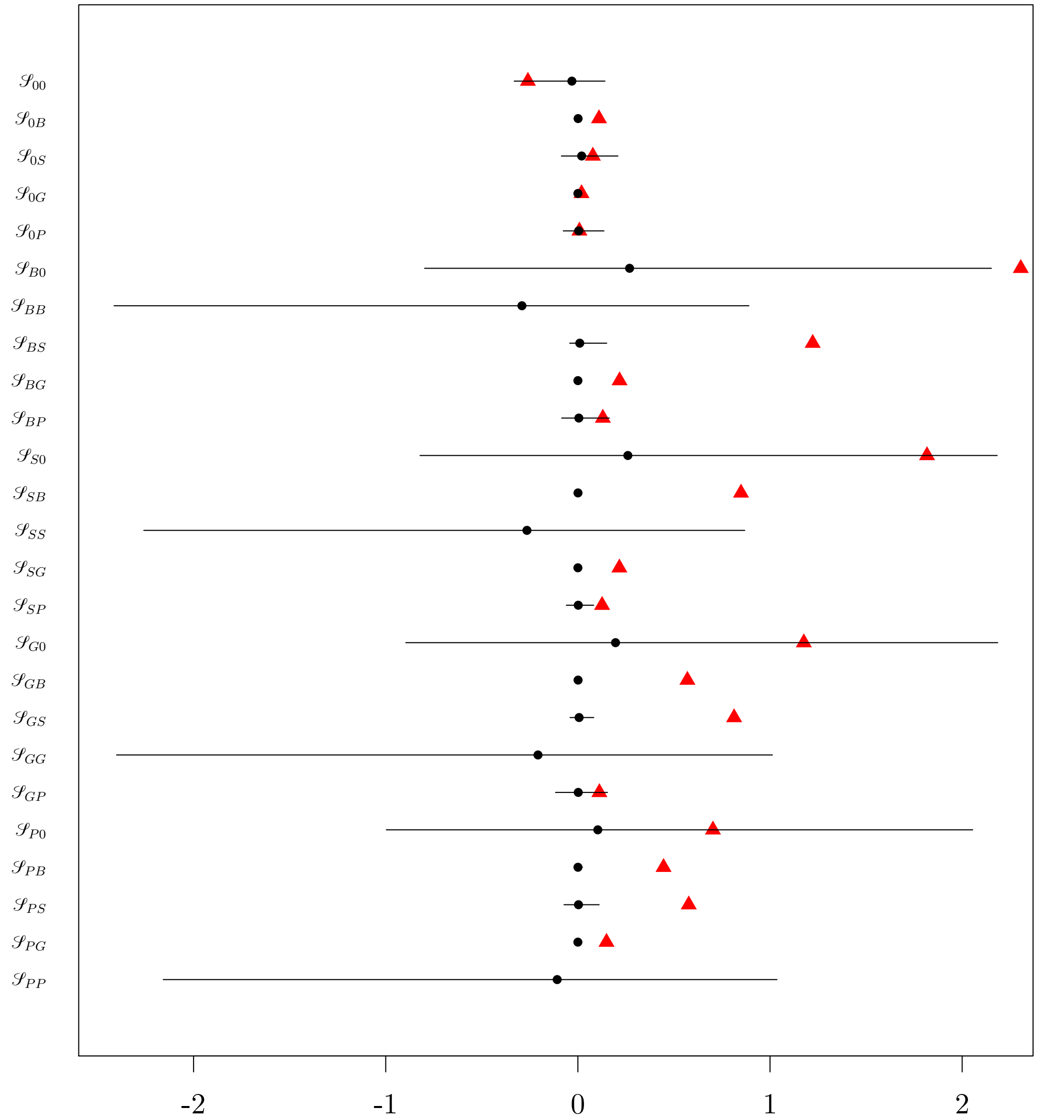
Note: Semielasticities for \$100 change in premiums. $\mathscr{S}_{ab}$ is the % change in predicted shares for good a with respect to \$100 change in good b. The eligible individuals have premium tax credits, are single and less than 50 years old. Abbreviations are 0 : Uninsured, B : Bronze, S : Silver , G : Gold, P : Platinum. Black lines are estimates of 95% confidence intervals from models using portfolio market shares; black circles stand for medians while red triangle is the median estimate from Tebaldi (2017, table 10).

Figure 5.6: Premium Semi-Elasticities, Subsidized, Single and Senior Population



Note: Semielasticities for \$100 change in premiums. $\mathscr{S}_{ab}$ is the % change in predicted shares for good a with respect to \$100 change in good b. The eligible individuals are those have premium tax credits, are single households and age at or over 65 years old. Abbreviations are 0 : Uninsured, B : Bronze, S : Silver , G : Gold, P : Platinum. Black lines are estimates of 95% confidence intervals from models using portfolio market shares; black circles stand for medians while red triangle is the median estimate from Tebaldi (2017, table 10).

Figure 5.7: Premium Semi-Elasticities, Family Households



Note: Semielasticities for \$100 change in premiums. $\mathscr{S}_{ab}$ is the % change in predicted shares for good a with respect to \$100 change in good b. The eligible individuals are households with more than one member. Abbreviations are 0 : Uninsured, B : Bronze, S : Silver , G : Gold, P : Platinum. Black lines are estimates of 95% confidence intervals from models using portfolio market shares; black circles stand for medians while red triangle is the median estimate from Tebaldi (2017, table 10).

Finally, while the estimates for premium coefficients shown in Figure 5.3 suggest that I recover the premium coefficients using portfolio market shares, it is the predicted changes in the market shares that often matter for policy analyses. By that I calculate the semi-elasticities of shares with respect to changes in premiums by $100 by metal levels. In Figure 5.4 to Figure 5.7 I report and compare the estimates of four demographic groups to those from Tebaldi (2017). As I have 3,000 draws of from the joint posterior distributions, I utilize all of the draws to calculate the distributions of the semi-elasticities. The black horizontal lines indicate the 95% confidence sets for each semi-elasticity using estimates from demand for portfolio market shares. The red triangles are the corresponding medians from Tebaldi (ibid.).

Interestingly, in Figure 5.4 we can see almost all of semi-elasticities of the unsubsidized population estimated by the Bayesian method cover the median estimates from granular individual-level data in the literature. Two exceptions are the own semi-elasticities of bronze plans and the changes in premiums of platinum plans on the shares of the uncovered. The coverage of the Bayesian semi-elasticities for the subsidized groups seem to perform quite well for cross-premium elasticities. The medians of own premium semi-elasticities in the subsidized groups, however, are mostly about twice larger in absolute values compared to the median estimated by individual-level data. We do see a mixed pattern for the estimates for family enrollees in Figure 5.4, however. The own premium semi-elasticities tend to be more centered around zeros in this case.

Chapter 6

Bayesian Demand Estimation in a Federal-facilitated Marketplace

6.1 Introduction: Estimating Demand of Health Insurance in the Marketplace of Pennsylvania

The uninsured individuals in the United States have dropped from 41.1 millions in 2013 (Collins et al., 2015) to about 27 millins in 2020 – about a 6% increase of the insured adult Americans during this period (Collins & Gunja, 2019). One of the main factor behind the increased rate of insurance coverage is the creation of Health Insurance Marketplaces (Marketplaces) required by the Affordable Care Act. By law, state governments can adopt one of the following model for the Marketplaces. First is a model of purely State-based Marketplaces (SBMs) in which the state governments create their websites, self administrate insurance companies, insurance plans and advertising open enrollment periods. Second is a model of Federally Facilitated Marketplaces (FFMs) and the federal government does the centralized managements and consumers in those FFMs enroll through HealthCare.gov website. States can also adopt a mixed model in which the governments manage their own markets but use the same federal website for enrollments. In the latter two cases insurance companies in states adopting the FFM platform are required to pay a 2% to 3.5% user fees to the federal government for administration (Keith, 2018). States of their Marketplaces might have laws requiring certain fees to fund the markets. For example, California also charges 3.5% user fees (CoverdCA, 2020), while the state of Washington lowers it fees from 3.46% to 2.56% from 2018 to 2020 (WASBE, 2019).

Recent literature on the differences in the models of Marketplaces suggest that premiums grow faster and enrollments drop more in FFMs compared to those in SBMs (Zhu et al., 2018). In addition, the user fees coupled with recent cut of budget in advertising FFMs by the federal government has temporarily made the states of New Jersey and Pennsylvania (PA) switch to their own State-based but Federally Facilitated business models in 2020. The state of Pennsylvania, for example, will have its own fully functional SBM in 2021 so that it could implement policy tools to encourage individuals to purchase health insurance. Yet little is known in the consumer demand of health insurance in PA, whereas a new branch of emerging literature in consumer health insurance demand has been studying the states of California, Washington or Colorado (See, Tebaldi, 2017; Saltzman, 2019; Drake, 2019 and Panhans, 2019).

The difference can be partly also explained in types of Marketplaces those states adopt– their SBMs have been able to provide comprehensive individual enrollments to researchers. The enrollments in FFMs, however, often are available at market level aggregated by counties or multi-product health insurers. The data aggregations of enrollments have caused some empirical difficulties in understanding economic parameters through existing empirical methods. To address the issue while understanding the economic parameters, researchers in very recent studies have attempted to provide very stylized economic models for policy analyses. For example, using aggregated enrollments in FFMs, Polykaova and Ryan (2019) estimate an equilibrium model of insurance demand and supply to study the welfare loss of the mean-testing subsidy design in Marketplaces, while Aizawa and Kim (2020) estimate the impact of defunding advertisements in FFMs on enrollments.

This chapter contributes to the new branch of literature studying FFMs using aggregated enrollments at market levels. Rather than defining a stylized model to address a particular question in health policies, I adopt a BLP-type aggregated demand model for multi-product firms and study the willness to pay for plan benefits. Next I estimate the cut of user fees in a particular FFM – the Marketplace in Pennsylvania. I find that if carriers fully pass the cut of 3% user fees through the reduction of premiums, the median gain of consumer surplus would increase by $70 per enrollee, while the amount increases to $150 for a median individual eligible for the Marketplace once the change of subsidy is also accounted for. The saving in user-fees can provide a guidance when PA fully switches its Marketplace from FFM to SBM in 2021.

6.2 Data and the Marketplace in PA

About 11% of total individuals in Pennsylvania are eligible in Health Insurance Marketplace [1]. The Marketplace consists of nine unique geographic rating areas (GRAs) in 67 counties of PA. In each of the GRAs a health insurance company can bid an array of benchmark premiums for a plan by actuarial fair value. Then the premiums of the plan varies by enrollees' family size, ages of all covered individuals in the same household, and smoking status, but not the health status of the enrollees in the same rating area.

Aggregated enrollments in the Marketplace of PA are reported yearly by CMS and available for public use. I retrieved market-level enrollments aggregated at HIOS plan-county level from 2014 to 2018 from the page of issuer level enrollment data

[1] Source:https://www.insurance.pa.gov/Coverage/Pages/Health-Insurance.aspx. Retrieved July 8th, 2020.

on the website of CMS[2]. It should be noticed that, however, a unique HIOS-plan is a product line of qualified health plans that include one or more of the following metal tiers that a potential enrollee could select: catastrophic, bronze, silver, gold and platinum. A typical firm of health insurance company, instead, could own multiple unique identifiers of HIOS-plans in the same geographic rating area, or in the same state. For example, Highmark BlueCross BlueShield in PA in 2018 owns three different HIOS identifiers in seven different GRAs, and it has two unique HIOS identifiers in GRA 4 and five counties. Those two plans, however, are sold separately and in a non-overlapped way: one plan (unique HIOS) has different aggregated market shares at two counties while the other is sold in the rest three counties. In each of the five counties aggregated market shares are reported as an HIOS-plan that is summed over multiple metal levels over the same HIOS number. This unique feature could help me identify parameters of interest in the demand model. That is, given an HIOS is the same in the same GRA but multiple counties, the variation in market shares at aggregated HIOS level can be explained by the variations in demographic distributions in the same set of counties.

Since a health insurer can own multiple unique HIOS numbers and the aggregated market shares are reported at HIOS county level, in the following I define a unique HIOS number as a plan or an "issuer," the same terminology used by CMS. I am also defining a unique multi-product firm as a unique identifier of HIOS. A product of this analysis is a metal plan in an HIOS of an issuer. As catastrophic plans account for less than 1% of total market shares in the Marketplace, and the plan is only available to individuals under 30 years old, I exclude catastrophic plans from my analysis. Lastly, a health insurance company that owns multiple HIOS numbers is called a carrier by the terminology of CMS.

As product-level attributes in the Marketplace of PA is not standardized, I collect

[2]See link https://www.cms.gov/CCIIO/Resources/Data-Resources/issuer-level-enrollment-data

the attributes from Health Insurance Plan Data Set (HIX) that offered structured plan benefit designs and networks from Robin Wood Johnson Foundation. The top panel of Table 6.1 below shows the market structure of the Marketplace over five years and 67 counties as well as the product portfolio of each issuer in PA. There are on average 6.2 carriers statewide in a year, while the level of competition varies greatly. A carrier could enter the Marketplace for one year but left afterwards, or started offering any metal of insurance plans since 2014 and continued. An average carrier would exist in the Marketplace of PA for about 3.88 years, and owns about 1.9 issuers over time. The number of metal plans per issuer also varies– not all issuers offer all of those four metals at once. In Table 6.2, the issuers always offer at least one silver and gold plans, but only about 83% of the issuers have bronze plans, while slightly more than 40% of the issuers offer plans at the most generous platinum level. The premium, ordered by metal level and current year US dollars for a typical enrollee at 21 years old vary from $2.5 to $4.62 thousands. The types of networks of a metal plan under an issuer also has large variations. A slightly more than majorities of metal plans have preferred provider networks (PPO), followed by health management organization and exclusive provider organization[3]. Less of than 10% of the metal plans have network of point of service.

[3]The summation over network types exceeds 100% as some issuers offer more than one type of networks for the same plan in the same geographic rating areas.

Table 6.1: Market Structure in the Marketplace of Pennsylvania

	min	25%	Mean	75%	max	s.d.
# of Carrier in a Year	5	5	6.20	7	8	1.30
# of Years per Carrier	1	2.5	3.88	5	5	1.64
# of Issuers per Carrier per Year	1	1	1.90	2	4	0.91
# of Plans per Issuer per Year	2	3	3.32	4	4	0.68

Table 6.2: Firm Charcteristics in the Marketplace of Pennsylvania

	Mean	s.d.
Network(%/Issuer)		
HMO	37.67	48.56
PPO	54.26	49.93
EPO	11.21	31.62
POS	8.97	28.64
Metal (%/Issuer)		
Bronze	82.96	37.68
Silver	100.00	0.00
Gold	100.00	0.00
Platinum	42.15	49.49
Premium ($1,000/Year)		
Bronze	2.50	0.73
Silver	3.24	1.16
Gold	3.69	1.12
Platinum	4.62	1.37
# No of Issuer	223	

Note: Summary statistics are calculated by 223 unique issuer across five years and nine geographic rating areas. Premiums by metals are calculated for a single 21 years old individual.

Table 6.3: Characteristics of Sampled Individuals in the Marketplace of Pennsylvania

	2014	2015	2016	2017	2018
Age	42.39	42.75	42.51	42.70	43.03
	(12.72)	(13.08)	(12.65)	(13.04)	(13.07)
FPL (%)	321.33	308.72	321.12	313.17	320.02
	(125.70)	(123.37)	(126.14)	(124.13)	(124.62)
Family (%)	29.85	31.16	31.70	32.99	33.70
	(45.77)	(46.32)	(46.54)	(47.02)	(47.28)

(a) Characterisitcs of Eligible Individuals by Years

	min	25%	Mean	75%	max	s.d.
Mandate & Subsidy	0.93	1.78	6.60	9.09	42.98	6.50
Policies						
1: changes in subsidy	0.00	0.00	0.20	0.30	1.34	0.21
2: changes in subsidy w/o mandates	0.76	0.98	1.45	1.73	3.61	0.59

(b) Changes in Subsidies and Mandate (per 1,000$), Year 2018

Note: Characteristics in both panels are averaged by 50 sampled individual for each of 67 counties. In bottom panel, the policy of cutting user-fees in the Marketplace leads to a second order effect of reducing subsidies (policy one). The second policy also accounts for the repeal of individuals mandate in 2019.

Table 6.4: Portfolio Market Shares in the Marketplace of Pennsylvania (%)

Carrier	Issuer	2014	2015	2016	2017	2018
Uninsured		67.73	51.73	48.61	53.29	51.43
		(9.87)	(14.93)	(12.91)	(13.08)	(13.91)
Aetna	33906	0.79	1.02			
		(0.52)	(0.60)			
Aetna	64844	0.63	5.81	3.80		
		(0.43)	(3.72)	(3.98)		
Aetna	91303	0.31	0.63			
		(0.21)	(0.43)			
Assurant	19068		0.14			
			(0.12)			
Capital BC	45127		0.27	6.53	16.41	13.27
			(0.18)	(3.97)	(7.89)	(9.81)
Capital BC	53789	0.38	0.56	9.29		
		(0.36)	(0.59)	(8.85)		
Capital BC	82795	0.15				
		(0.14)				
Geisinger	22444	5.85	6.39	10.44	17.29	21.26
		(4.83)	(6.08)	(10.09)	(11.98)	(15.97)
Geisinger	75729	0.79	0.75	2.56		
		(0.45)	(0.56)	(1.91)		
Highmark BCBS	33709	23.74	32.28	18.47	3.72	1.76
		(7.45)	(10.19)	(5.45)	(2.55)	(1.28)
Highmark BCBS	36247			16.65		
				(1.67)		

Carrier	Issuer	2014	2015	2016	2017	2018
Highmark BCBS	38949					2.36
						(0.83)
Highmark BCBS	55957	20.18	39.40	40.57		
		(4.92)	(14.60)	(7.00)		
Highmark BCBS	70194	3.29	2.49	3.91	4.55	9.35
		(2.15)	(1.26)	(2.94)	(5.78)	(4.26)
Highmark BCBS	83731				31.47	17.51
					(8.18)	(9.03)
Independence BC	31609	11.91	10.12	8.76	16.03	12.63
		(2.17)	(2.20)	(2.54)	(3.57)	(1.38)
Independence BC	33871	38.28	36.19	45.32	53.30	55.10
		(4.25)	(5.21)	(4.63)	(13.18)	(12.49)
UPMC	16322		9.50	33.40	42.22	46.15
			(6.03)	(9.73)	(9.39)	(13.89)
UPMC	16481	0.10				
		(0.07)				
UPMC	52899	0.45				
		(0.44)				
UnitedHealth	24872		5.13	7.41		
			(4.66)	(4.36)		
# of Issuer-County		431	496	393	213	200

Note: Standard deviation in parentheses. BC is the abbreviation of Blue Cross and BCBS stands for Blue Cross Blue Shield.

To estimate a discrete choice model of insurance demand, I also need to construct eligible population in the Marketplace of PA. The eligible enrollees are computed using America Community Survey and I rule out the subset of individuals qualifying for Medicaid, Medicaid expansion in 2015 in PA, or Medicare. In the survey I also observe demographic characteristics such as income, age and family status at individual level. Thus I can calculate the expected premium tax credit (PTC) for the subsidized individuals as well as individual mandates, the penalty of remaining uncovered, per household. Those two quantities are also used in my counter-factual analysis when I impute the change of consumer surplus. Specifically, the state of Pennsylvania has proposed to change its model of managing Marketplace to a purely State-based Marketplace in 2021. Given this change, carriers might be able to cut premiums of all plans by at most 3% user-fees that they pay when selling plans through the Federal platform. Another possibility is when PA adopts a Stated-based Marketplace, the relative prices of all plans in Marketplaces remain the same but the relative price of any plan versus no-coverage drops. My first policy analysis can quantify this effect[4]. In this chapter I focus on the demand side estimation and do not explicitly model the cost function of issuers. Instead, I vary the possible set of user fees by an increment of 1% with a range of 0% to 3% possible cut of premiums after the Marketplace clears. One possible explanation is conditional on PA switching to a pure State-based Marketplace, if premiums didn't change then either the state of PA would keep 3% user fees to maintain the full function of its Marketplace, or the marginal cost function of carriers could be perfectly inelastic if no user fees existed. Another extreme scenario in market equilibrium after a proposed cut of user fees from the state government is a cut of 3% premiums by carriers means the marginal cost curve is perfectly elastic and no adverse selection exists. While a cut of 1% or 2% of premiums by carriers indicates an equilibrium effect due to cut in user fees or shifts of negatively sloped

[4] However, this passing through effect assume marginal costs of health insurers are perfectly elastic and imply no adverse selection in the health insurance market (Einav and Finkelstein 2011). An alternative modeling strategy is to allow adverse selection in the cost function as in Tebaldi (2017), Panhans (2019) or Polykaova and Ryan (2019) if claims are available.

marginal cost curves in the health insurance market.

When premiums of all plans change, however, there could exist a second order effect known as "silver loading" in Marketplaces (Drake and Anderson, 2020). In other words, because the PTC is linked to the second lowest premiums of a silver plan in a geographic rating area, the decrease in premiums of that plan also lowers the amount of subsidies. Thus my second policy analysis will take silver loading into account to calculate the net change in enrollments and consumer surplus.

Notice, however, the changes in consumer surplus in both policy proposals intend to provide approximations to the adoption of State-based Marketplace model in the 2021. In realty, however, is that the current Administration repealed individual mandates in 2019– a part of the opportunity cost of remained uninsured by any health insurance coverage. Thus my third counterfactual policy analysis also takes the repeal of individual mandate into account. In the top panel of Table 6.3, I summarize the characteristics of 50 sampled individuals for all county-years in my estimation procedure as well as the costs of being uninsured through the proposed policy changes in 2018. The majority of sampled individuals are over 40 years old, eligible for premium tax credit as the income level is about 308% to 320% of federal poverty line, and more than 56% of the sampled individuals from ACS are single households.

In the bottom panel of the same table we can see that the opportunity cost of being uninsured as the sum of individual mandate and forgone premium tax credit is about $6,600 USD per year in 2018 for an average household. The amount of change in the silver loading through a cut of 3% in user fees and fully passed to premiums of all health plans is $200 (=0.2*$1,000) dollars in 2018. The effect of silver loading plus repeal of individual mandate in 2018 could lead to a drop of $1,450 of the opportunity cost for the average household.

It should be noticed that in Table 6.3 the summary statistics are calculated from a random sampled households from the pool of eligible population from ACS. An underlying assumption is the randomly sampled individuals are representative. In my un-reported analysis, I tested my sampling strategy and found the random sample of 50 individuals per county-year is statistically no different than other batches of sampled individuals at 1% level. The pool of eligible population from ACS, in addition, is used to calculate the proportion of uninsured individuals in PA as in Table 6.4. Given the official report of PA health department, about 5% of the total population is uninsured and another 5% of total population is currently covered by individual Marketplace. My calculation of the uninsured and eligible population is thus consistent with the official report.

In Table 6.4 we also observe the variations of portfolio market shares of issuer-carrier over time. Interestingly, UPMC (University of Pittsburg Medical Center) was a late entrant in 2015 but gained a significant market shares over time, while Highmark Blue Cross Blue Shied (Highmark BCBS) was an early entrant in the Marketplace but it's market shares of different issuers were dropping over the observed periods. The variations of market shares for different issuers under the same carrier could help identify carrier fixed effects in my model. In addition, the entry and exit of different issuers under the same carrier could imply that the carrier redesigned the benefits of plans or networks. Those variations provide extra identifiablility of revealing consumers' preferences of plan characteristics as well as network fixed effects.

6.3 Model Specification and Results

The estimation procedure follows Appendix D and estimates demand in health insurance plans when portfolio market shares at the firm level are the dependent variable. The strategy extends Musalem et al. (2009) that provide a Bayesian demand estimation for highly adopted approach by Berry et al. (1995, hence BLP approach) for aggregated data. The key advancement in my BLP approach is to recognize that potential enrollees in Marketplaces make decision and purchase health insurance plans. The choices are summarized through the unobserved individuals (to health economists) and reported as enrollments of multi-product firms. Thus my Bayesian approach starts by simulating a set of heterogeneous but representative agents that make choices consistent with observed portfolio market shares at the firm level. Parameters in demand functions could be recovered given variations in portfolio shares if market structures in terms of competitions are similar, while demographic characteristics vary by markets.

The preferred specification is the following:

$$\begin{aligned} u_{ijtm} &= \beta_{imt} prem_{ijtm} + \beta_{AV} AV_j + \alpha_j + \xi_{jtm} + \epsilon_{ijtm}, \forall j \setminus \{0\}, \\ u_{i0} &= \beta_{imt,prem} \tau_{imt} + \epsilon_{i0tm}, \\ E(\beta_{imt,prem}) &= \beta_{prem} + \delta D_{itm}. \end{aligned}$$

where τ_{imt} includes both premium tax credits and individual mandates. The dummies of fixed effect variables of α_j include eight carriers in PA, a binary indicator for all plans that captures the mean utility of being insured, network dummies of HMO, EPS, and EPO. The slope terms of observed demographic characteristics in D_{itm} include dummies of family, subsidy status and being senior and over 50 years old. The specification is similar to Polykaova and Ryan (2019) and I use actuarial fair value (AV) of metal plans rather than maximum out-of-pockets or deductibles.

Notice that I account for the potential endogeneity of premiums with unobserved plan attributes ξ_{jtm} using control functions. The instrumental variables consists of dummies of five years and nine geographic rating areas, as well as Hausman instruments (Hausman, 1996). Intuitively, the premium of one metal plan is correlated with its competitors' premiums of the same metal. In the context of multi-product firms under the same carrier, the instruments of premiums are calculated by competing carriers to exclude both effects of cannibalizing the shares of other issuers and potentially endogenous pricing of the same carrier. The first stage of the control function approach is to project the unobserved plan attributes in the utility function into instruments and then error terms are predicted through a reduced form pricing equation. In my estimation algorithm, I update the residuals by running the first stage regression in each of MCMC iterations.

Table 6.5: Posterior Distributions of Demand Estimates in the Marketplace of Pennsylvania

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Prem ($/1000)						
Constant	-1.65	-1.28	-1.12	-1.15	-1.03	-0.84
Senior	0.16	0.44	0.51	0.52	0.61	0.89
Subsidy	-0.98	-0.74	-0.65	-0.66	-0.58	-0.37
Family	-0.07	0.15	0.23	0.24	0.33	0.55
Actuarial Fair Value	3.05	3.99	4.33	4.32	4.77	5.31
ϕ	-0.81	-0.37	-0.23	-0.25	-0.11	0.29
σ^2_{prem}	0.19	0.50	0.65	0.70	0.87	1.65
σ^2_v	6.33	7.45	7.95	8.05	8.64	10.30

Table 6.5 and plots Figure 6.1 to Figure 6.3 show the estimates by 50 sampled individuals per county-year in PA for 300,000 iterations of MCMC. After examining the trace plots of estimates, I discarded the first 200,000 as burn-in periods and used every 100-th draw of the rest iterations for my final analysis. The constant term of premiums has a median of -1.12, while the senior and family households are less price sensitive. And consistent with literature, the subsidized group with income less than 400% FPL is more price sensitive (in absolute value). Consumers also value health insurance plans. That is, the coefficient for actuarial fair value as the proportional

of medical expenses covered by health insurance plans is positive. In terms of willing to pay more one unit change in a more general plan design, median consumers are willing to pay for an extra $298 (90% C.I., (-872,1,602)).

The changes in consumer surplus under different proposed policies are calculated in Table 6.6[5]. The proposed changes in premium cut if PA has its own State-based Marketplace result in a gain of 70 dollars per consumer in 2018. Adding the effects of silver loading that reduces the relative price of uninsured cost almost doubles the amount of consumer surplus– an average consumer gains 150 dollars through the cutting of premiums as well as silver loading. Yet the repeal of individual mandate seems to be a major contributor of increase in consumer surplus– as about half of the eligible population would remain uninsured.

Table 6.6: Changes in Consumer Surplus in the Marketplace of Pennsylvania

Policy	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Cut 3% User Fees	62.15	68.12	69.96	69.96	71.65	77.06
Plus Silver Loading	142.68	148.51	150.32	150.32	152.10	159.08
Silver Loading and No Mandate	957.16	985.61	994.94	995.49	1005.32	1037.56

Note: Benchmark case is the status quo in 2018. Unite is dollar per person in 2018.

[5]To calculate the quantities of consumer surplus I assume the unobserved product specific attributes represent pure economic shocks. And the values are set at zero. This would be a reasonable assumption in particular as I have controlled for premiums, actuarial fair values and networks

Figure 6.1: Posterior Distributions for Demand Estimates in Pennsylvania

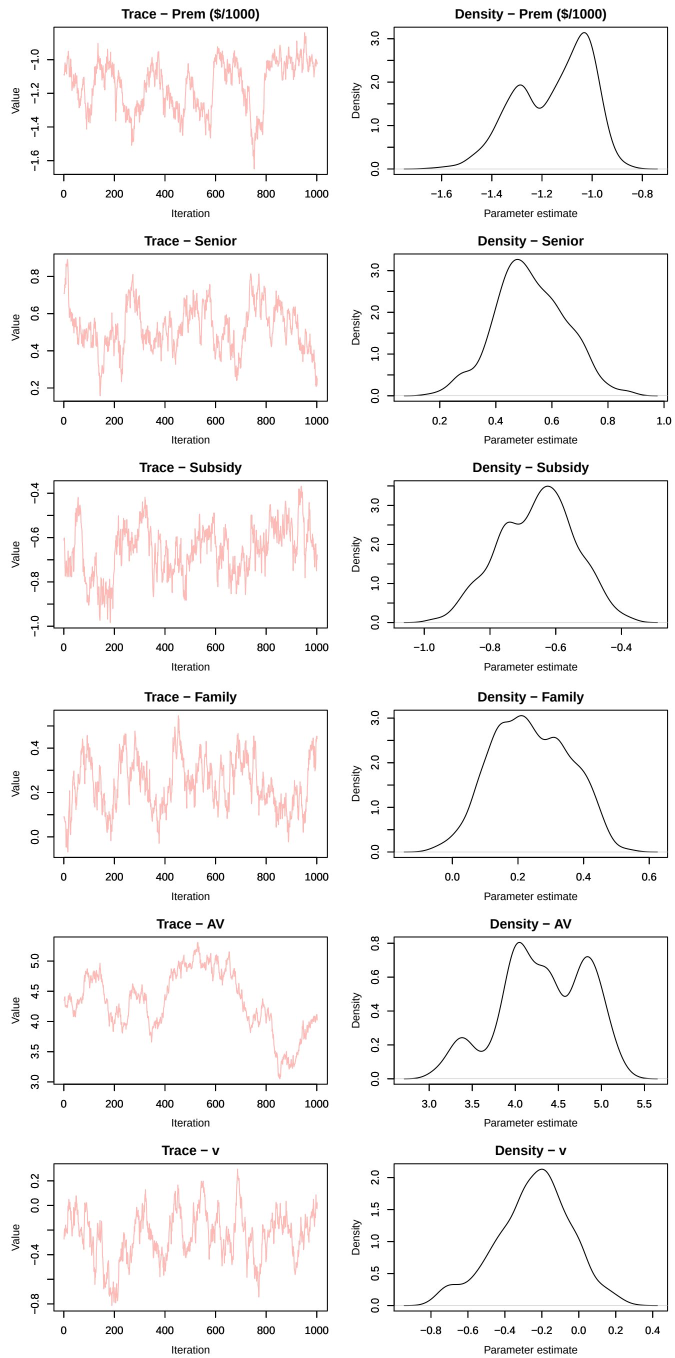
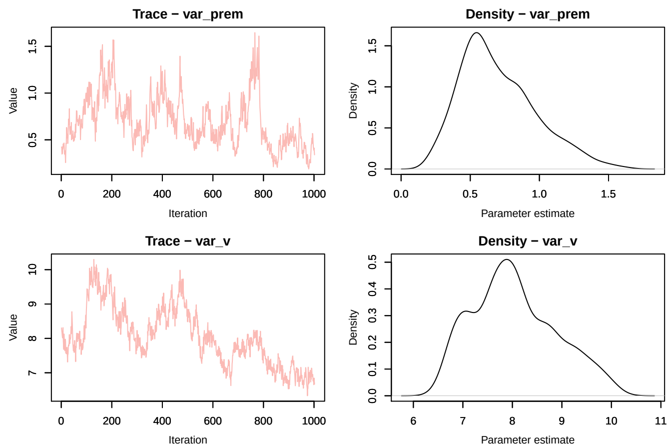
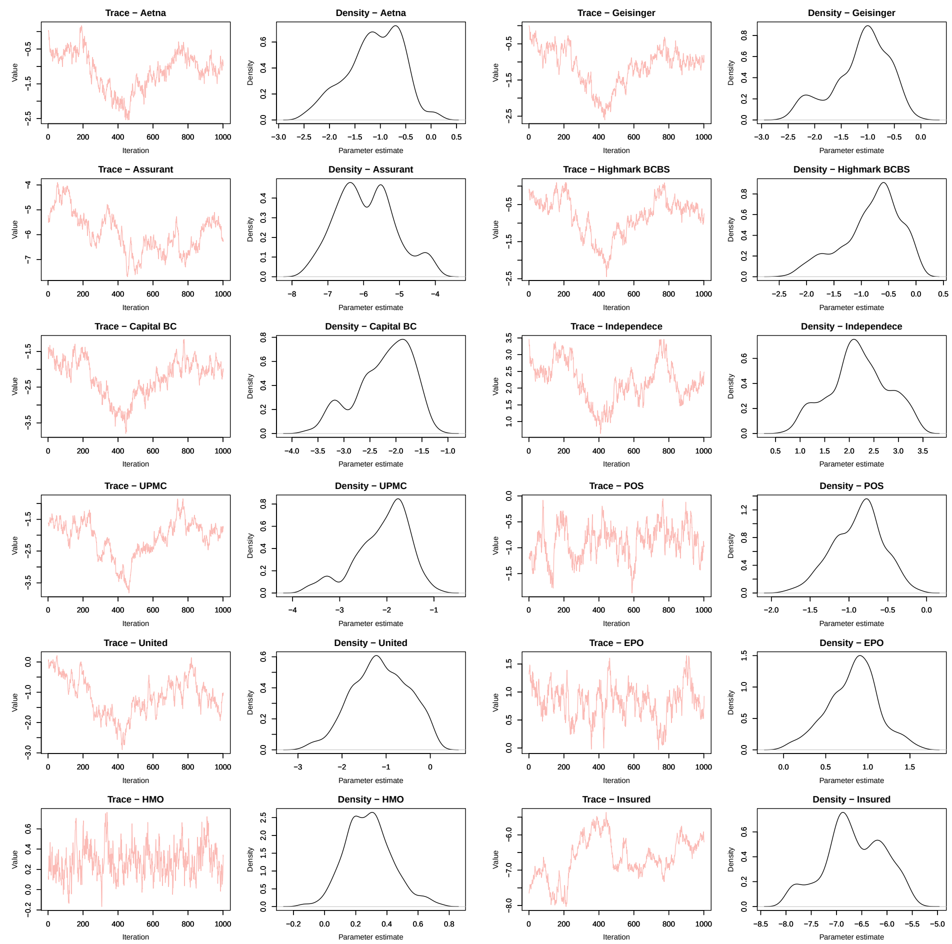


Figure 6.2: Posterior Distributions for Demand Estimates in Pennsylvania, Continued



Left: Trace plots for thinned MCMC draws; Right: Sampling Distributions.

Figure 6.3: Posterior Distributions for Demand Estimates in Pennsylvania, Fixed Effects



Left: Trace plots for thinned MCMC draws; Right: Sampling Distributions.

6.4 Summary and Limitation

In this chapter I contribute to the literature of health economics by studying a Marketplace using enrollments aggregated at multi-product issuers. My demand estimates are similar to Polyakova and Ryan (2019) whose demand elasticity of premiums decreases in absolute value with income and among younger potential enrollees, and increases in family size. In my study I also provide estimates of changes in enrollments when PA starts to use its own State-based Marketplace and pass all user fees to potential enrollees. I find that, further, if I include the effects of silver loading, the net change is about twice as large compared to a simple cut of all premiums by all carriers in 2018, which is the equal to 150 dollars increase in consumer surplus compared to no changes in premiums.

The effects of those economic policies, however, are based on partial equilibrium and would be likely an upper bounds of carriers' optimization problem than the equilibrium in 2021 when the State-based Marketplace opens. In addition, an implicitly assumption for the welfare calculation is all carriers are homogeneous in the marginal cost function. The assumption might not hold true, especially carriers have very different market shares over time in Table 6.4, and the carriers could often change their portfolios of products. From policy makers' point of view, knowing changes in consumer surplus might be different than choosing the optimal user fees for carriers. To evaluate the optimal policy, one might need in addition to know producer surplus, as well as the welfare function of the state government. For example, the state government might want to cut fees to increase consumer surplus. But cutting fees reduces the funding to maintain the quality of its Marketplace platform. An example is consumers are less likely to enroll in if the government didn't have sufficient funds to advertise open enrollment periods (Aizawa and Kim, 2020).

More detailed policy analyses in the future will need to include firms' bidding decisions explicitly for each year when markets clear. It should be noticed that as Marketplaces are very highly regulated, firms also face three different pricing constraints in 2014-2018. For example, if insurers bid premiums too high and make profits in excess of 80% of reimbursed claims, medical loss ratio requirement mandates that insurers refund the extra profits to consumers (Abraham et al., 2014) and recent evidence suggests the regulation makes firms more costly (Cicila et al., 2019). Another supply side regulation was the risk corridor program making insurers in positive profit subsidize insurers in profit loss– and some carriers could keep their premiums low due to the cross subsidies that distorted pricing incentives (Sacks el al., 2017). However, one might follow the strategy of estimating cost functions in Polykaova and Ryan (2019) or Tebaldi (2017), if the researcher has collected claims at carrier-level through the public use files of medical loss ratio.

Yet another concern for the future study is in 2018 the US federal government has terminated funding carriers for cost-sharing subsidy program that allows low income individuals purchase silver plans and enjoy more generous benefits without extra costs. Recent empirical analyses have shown the termination of funding CSR has driven up premiums of many plans (Dorn, 2019), while at the same time the termination increases subsidies levels through the effect of silver loading (Drake and Anderson, 2020). A detailed analysis in the future using data in those years would need to take all of those policy regulations into full account.

Chapter 7

Conclusion

This thesis complements the finding of demand estimates in Health Insurance Marketplaces and empirical literature of demand estimation in industrial organization using aggregated data by advancing Bayesian discrete choice models of multi-product firms. The specific Bayesian model for aggregated data can be used in empirical studies whenever the market shares are aggregated through firms, counties or product types.

The Bayesian approach is based on simulating a subset of representative individuals that could approximate the eligible population making the choices, and the choices are aggregated through products to firm level, for example. The current estimation procedure allows for the observed shares at the firm level independently observed by consumers' demographic characteristics, while the procedure iterates through all possible joint distributions of consumer characteristics and choices. It is possible, however, that researchers observe the conditional distributions of market shares on consumers' types such as ages and incomes. The conditional distributions impose additional conditions on the parameter space that MCMC could explore. One possible approach is to extend the augmented choices on observed characteristics while the choices can be updated within consumers of the same type.

Another data restriction is that one might observe current choices and lagged choices at the firm level as two marginal distributions in the same market at once. It is possible, indeed, to modify the existing approach and consider multiple augmented market constraints. See Musalem et al. (2008) for example. In addition, the discrete choice model in this thesis can be extended to accommodate a panel of consumers and include lagged choices. This application could have important role in explaining consumers' enrollment behavior in health insurance markets. For example, Drake et al. (2020) have shown enrollees have strong inertia in the Marketplace of California.

Additionally, the Bayesian approach using portfolio market shares has the following limitations. First, in the simulation I assume the model specification is correct up-to some omissions of observed heterogeneities. If important individual characteristics are miss-specified, the variance covariance matrix of unobserved heterogeneity of preference weights could increase. But given that researchers have data variations of portfolio market shares, demographics and products by markets, the Bayesian approach could provide confidence intervals that do cover the parameter values. The trade-off is the standard deviation would be larger compared to those with more granular market shares at the product level, or consumer level.

Second, while Bayesian estimation using MCMC can provide estimates of parameters of interests, the estimates are usually weighted averages of prior distribution and observed data. In case that the data have less variations in identification parameters, their posterior distributions would possibly behave similarly to the prior distributions. In application of my Bayesian approach, I find that while market shares are aggregated at the firm level, geographic variations by counties provide additional sources of variations for parameter identification. In addition, as Tebaldi et al. (2019) indicate, the structure of Marketplaces implies demand functions are non-parametrically identified as individuals are subsidized by income level and the premiums also vary by ages given the same level of income.

Thirdly, Bayesian demand estimation using MCMC could be very computationally intensive and time consuming, while the convergence results are not always guaranteed. In many cases, convergence of MCMC can be accessed using multiple chains with different initial values in parallel in a cluster of computers or on multiple cores of a local machine after long runs of MCMC. Researchers can evaluate the necessary conditions of convergence results using, for example, Gelman-Rubin statistics (1992), and find indications that convergence might fail. In my thesis I examine the results mostly by trace plots of graphs, as running multiple chains could take days or weeks for my particular analyses.

Finally, the Bayesian approach used in this thesis is a reduced form. That is, I do not explicitly specify the joint distribution of demand and supply and study the market equilibrium for policy analyses as in Yang et al. (2003). Interested readers can consult existing studies of market equilibria of Marketplaces, for example, like Tebaldi (2017) and Polyakova and Ryan (2019), and take the supply constraints into accounts for counter-factual analyses. Their approaches, however, require specifying many economic constraints on the supply side for insurers' pricing decisions. Alternatively, I discuss in Appendix B a simplified but general pricing model of multi-product firms. It should be noticed that, in an abstract economy in general the equilibrium analysis including firms' pricing model is well understood in the empirical literature– while the study of the multi-product firms conditional on the fact researchers observe aggregated portfolio market shares is not. Further studies in this branch of literature could explicitly incorporate the equilibrium analysis of demand and supply using portfolio market shares.

References

Abraham, J. M., Karaca-Mandic, P., & Simon, K. (2014). How has the affordable care act's medical loss ratio regulation affected insurer behavior? *Medical Care*, *52*(4), 370–377.
URL http://www.jstor.org/stable/24465982

Aizawa, N., & Kim, Y. S. (2020). Government advertising in market-based public programs: Evidence from the Health Insurance Marketplace. Working Paper 27695, National Bureau of Economic Research.
URL http://www.nber.org/papers/w27695

Altekar, G., Dwarkadas, S., Huelsenbeck, J. P., & Ronquist, F. (2004). Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics*, *20*(3), 407–415.
URL https://doi.org/10.1093/bioinformatics/btg427

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–90.
URL http://EconPapers.repec.org/RePEc:ecm:emetrp:v:63:y:1995:i:4:p:841-90

Berry, S., Levinsohn, J., & Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of political Economy*, *112*(1), 68–105.
URL https://www.jstor.org/stable/10.1086/379939

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, *25*(2), 242–262.
URL http://www.jstor.org/stable/2555829

Carlin, B. P., & Louis, T. A. (2008). *Bayesian methods for data analysis*. Chapman and Hall/CRC, 3 ed.
URL https://doi.org/10.1201/b14884

Chen, Y., & Yang, S. (2007). Estimating disaggregate models using aggregate data through augmentation of individual choice. *Journal of Marketing Research*, *44*(4), 613–621.
URL https://doi.org/10.1509/jmkr.44.4.613

Cicala, S., Lieber, E. M. J., & Marone, V. (2019). Regulating markups in us health insurance. *American Economic Journal: Applied Economics*, *11*(4), 71–104.
URL https://www.aeaweb.org/articles?id=10.1257/app.20180011

Collins, S. R., & Gunja, M. Z. (2019). Do Americans think about their health coverage ahead of the 2020 election? findings from the Commonwealth Fund Health Insurance in America Survey, March–June 2019.
URL https://www.commonwealthfund.org/publications/issue-briefs/2019/sep/what-do-americans-think-health-coverage-2020-election

Collins, S. R., Gunja, M. Z., & Beutel, S. (2015). New U.S. census data show the number of uninsured Americans dropped by 8.8 million. *The Commonwealth Fund Blog*.
URL https://www.commonwealthfund.org/blog/2015/new-us-census-data-show-number-uninsured-americans-dropped-88-million

CoveredCA (2020). Covered California comments on proposed HHS changes to benefit and payment parameters. https://www.coveredca.com/newsroom/news-releases/2020/02/03/covered-california-comments-on-proposed-hhs-changes-to-benefit-and-payment-parameters/.

DeLeire, T., Chappel, A., Finegold, K., & Gee, E. (2017). Do individuals respond to cost-sharing subsidies in their selections of marketplace health insurance plans? *Journal of Health Economics*, *56*, 71 – 86.
URL http://www.sciencedirect.com/science/article/pii/S0167629617301005

Dorn, S. (2019). Silver Linings For Silver Loading. *Health Affairs Blog*. DOI: 10.1377/hblog20190530.156427.
URL https://www.commonwealthfund.org/blog/2015/new-us-census-data-show-number-uninsured-americans-dropped-88-million

Drake, C. (2019). What are consumers willing to pay for a broad network health plan?: Evidence from Covered California. *Journal of Health Economics*, *65*, 63 – 77.
URL http://www.sciencedirect.com/science/article/pii/S0167629618304909

Drake, C., & Abraham, J. M. (2019). Individual market health plan affordability after cost-sharing reduction subsidy cuts. *Health Services Research*, *54*(4), 730–738. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-6773.13190

Drake, C., & Anderson, D. M. (2020). Terminating cost-sharing reduction subsidy payments: The impact of Marketplace zero-dollar premium plans on enrollment. *Health Affairs*, *39*(1), 41–49. PMID: 31905063. URL https://doi.org/10.1377/hlthaff.2019.00345

Drake, C., Ryan, C., & Dowd, B. E. (2020). Sources of inertia in health plan choice in the individual health insurance market. Working paper.

Einav, L., & Finkelstein, A. (2011). Selection in insurance markets: Theory and empirics in pictures. *Journal of Economic Perspectives*, *25*(1), 115–38. URL https://www.aeaweb.org/articles?id=10.1257/jep.25.1.115

Ericson, K. M., & Starc, A. (2012). Heuristics and heterogeneity in health insurance exchanges: Evidence from the Massachusetts Connector. *American Economic Review*, *102*(3), 493–97. URL https://www.aeaweb.org/articles?id=10.1257/aer.102.3.493

Fader, P. S., & Hardie, B. G. S. (1996). Modeling consumer choice among skus. *Journal of Marketing Research*, *33*(4), 442–452. URL http://www.jstor.org/stable/3152215

Gelman, A., Roberts, G. O., Gilks, W. R., et al. (1996). Efficient Metropolis jumping rules. *Bayesian Statistics*, *5*(599-608), 42.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. URL http://www.jstor.org/stable/2246093

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood. Interface Foundation of North America. URL http://hdl.handle.net/11299/58440

Goldman, A. L., McCormick, D., Haas, J. S., & Sommers, B. D. (2018). Effects of the ACA's Health Insurance Marketplaces on the previously uninsured: A quasi-experimental analysis. *Health Affairs*, *37*(4), 591–599. PMID: 29608372. URL https://doi.org/10.1377/hlthaff.2017.1390

Greene, J., Hibbard, J. H., Sacks, R., Overton, V., & Parrotta, C. D. (2015). When patient activation levels change, health outcomes and costs change, too. *Health*

Affairs, *34*(3), 431–437.
URL https://doi.org/10.1377/hlthaff.2014.0452

Guadagni, P. M., & Little, J. D. C. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science*, *2*(3), 203–238.
URL https://doi.org/10.1287/mksc.2.3.203

Hall, A. (2007). The value of Medicare managed care plans and their prescription drug benefits. FEDS Working Paper No. 2007-19.

Hausman, J. (1996). Valuation of new goods under perfect and imperfect competition. In *The Economics of New Goods*, (pp. 207–248). National Bureau of Economic Research, Inc.
URL https://EconPapers.repec.org/RePEc:nbr:nberch:6068

HelathCare.gov (Accessed July 15, 2020). *Silver Health Plan*. https://www.healthcare.gov/glossary/silver-health-plan/.

Jiang, R., Manchanda, P., & Rossi, P. E. (2009). Bayesian analysis of random coefficient logit models using aggregate data. *Journal of Econometrics*, *149*(2), 136 – 148.
URL http://www.sciencedirect.com/science/article/pii/S0304407608002297

Keith, K. (2018). ACA enrollment continues to lag, new rule on federal user fee adjustments. DOI: 10.1377/hblog20181203.85146.
URL https://www.healthaffairs.org/do/10.1377/hblog20181203.85146/full/

Kim, Y. S. (2014). *Essays on private Medicare insurance markets*. Ph.D. thesis. https://repository.upenn.edu/dissertations/AAI3623554.

Musalem, A., Bradlow, E. T., & Raju, J. S. (2008). Who's got the coupon? estimating consumer preferences and coupon usage from aggregate information. *Journal of Marketing Research*, *45*(6), 715–730.
URL https://doi.org/10.1509/jmkr.45.6.715

Musalem, A., Bradlow, E. T., & Raju, J. S. (2009). Bayesian estimation of random-coefficients choice models using aggregate data. *Journal of Applied Econometrics*, *24*(3), 490–516.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.1060

Nevo, A. (2000a). Mergers with differentiated products: The case of the ready-to-eat cereal industry. *The RAND Journal of Economics*, *31*(3), 395–421.
URL http://www.jstor.org/stable/2600994

Nevo, A. (2000b). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy*, *9*(4), 513–548.
URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1430-9134.2000.00513.x

Nocke, V., & Schutz, N. (2018a). An aggregative games approach to merger analysis in multiproduct-firm oligopoly. Working Paper 24578, National Bureau of Economic Research.
URL http://www.nber.org/papers/w24578

Nocke, V., & Schutz, N. (2018b). Multiproduct-firm oligopoly: An aggregative games approach. *Econometrica*, *86*(2), 523–557.
URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA14720

Nosal, K. E. (2012). *Switching Costs in the Market for Medicare Advantage Plans*. Ph.D. thesis, The University of Arizona.
URL http://hdl.handle.net/10150/228491

Panhans, M. (2019). Adverse selection in aca exchange markets: Evidence from colorado. *American Economic Journal: Applied Economics*, *11*(2), 1–36.
URL https://www.aeaweb.org/articles?id=10.1257/app.20170117

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, *110*(4), 705–729.
URL https://doi.org/10.1086/340779

Petrin, A., & Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, *47*(1), 3–13.
URL https://doi.org/10.1509/jmkr.47.1.3

Polyakova, M., & Ryan, S. P. (2019). Subsidy targeting with market power. Working Paper 26367, National Bureau of Economic Research.
URL http://www.nber.org/papers/w26367

Roberts, G. O., & Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, *18*(2), 349–367.
URL https://doi.org/10.1198/jcgs.2009.06134

Rossi, P. E., Allenby, G. M., & McCulloch, R. (2005). *Bayesian Statistics and Marketing*. John Wiley & Sons, 2 ed.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., & Sobek, M. (2020). IPUMS USA: Version 10.0 [dataset]. https://doi.org/10.18128/D010.V10.0.

Sacks, D. W., Vu, K., Huang, T.-Y., & Karaca-Mandic, P. (2017). How do insurance firms respond to financial risk sharing regulations? evidence from the Affordable Care Act. Working Paper 24129, National Bureau of Economic Research.
URL http://www.nber.org/papers/w24129

Saltzman, E. (2019). Demand for health insurance: Evidence from the California and Washington ACA exchanges. *Journal of Health Economics*, *63*, 197 – 222.
URL http://www.sciencedirect.com/science/article/pii/S016762961730070X

Shaked, A., & Sutton, J. (1990). Multiproduct firms and market structure. *The RAND Journal of Economics*, *21*(1), 45–62.
URL http://www.jstor.org/stable/2555492

Small, K. A., & Rosen, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica*, *49*(1), 105–130.
URL http://www.jstor.org/stable/1911129

Smith, H. (2004). Supermarket choice and supermarket competition in market equilibrium. *The Review of Economic Studies*, *71*(1), 235–263.
URL https://doi.org/10.1111/0034-6527.00283

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(4), 795–809.
URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00265

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*(398), 528–540.
URL http://www.jstor.org/stable/2289457

Tebaldi, P. (2017). Estimating equilibrium in health insurance exchanges: Price competition and subsidy design under the ACA. Working Paper 2017-05, Becker Friedman Institute for Research in Economics Working Paper.

Tebaldi, P., Torgovitsky, A., & Yang, H. (2019). Nonparametric estimates of demand in the California Health Insurance Exchange. Working Paper 25827, National Bureau of Economic Research.
URL http://www.nber.org/papers/w25827

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, 2 ed.

WASBE (2019). Washington Health Benefit Exchange approves carrier assessment for 2020 plan year. https://www.wahbexchange.org/washington-health-benefit-exchange-approves-carrier-assessment-2020-plan-year/.

Yang, S., Chen, Y., & Allenby, G. M. (2003). Bayesian analysis of simultaneous demand and supply. *Quantitative Marketing and Economics*, *1*(3), 251–275. URL https://doi.org/10.1023/B:QMEC.0000003327.55605.26

Zhu, J., Polsky, D., & Zhang, Y. (2018). State-based Marketplaces outperform Federally-Facilitated Marketplaces. *LDI Issue Brief*, *22*(1), 1–7.

Appendix A

Post-estimation Calculation for Elasticities and Welfare Measurements

A.1 Consumer Surplus and Willness to Pay

This chapter discusses how the procedures of post-estimation are conducted in the main text for consumer surplus, willness to pay and price elasticities.

First recall the utility function specified in equation Section 4.2 is

$$u_{ijtm} = \beta_{imt} x_{ijtm} + \alpha_j + \xi_{jtm} + \epsilon_{ijtm}, \forall j \setminus \{0\}, \tag{A.1}$$

$$u_{i0} = \epsilon_{i0tm}. \tag{A.2}$$

The form of the utility function in the main text follows random utility model (RUM) and can be rationalized by a standard economic problem in which an agent maximizes her utility by purchasing a discrete amount of good j. The economic constraint is the agent's disposable income, prices of all goods, and other product characteristics or qualities. Then the welfare measurements are constructed following the classic microeconomic theory. After the agent maximizes her utility, the form of equation Section 4.2 as an indirect utility function of prices and qualities thus follows if economists substitute the solved demand function into the utility function of discrete quantities of goods. See, for example, Small and Rosen (1981) for the detailed step-by-step construction of the welfare measurements of compensation variation (CV), the expenditure required to compensate the utility-maximized agent at a fixed utility level when a particular attribute of a product changes, of discrete goods if all goods and product characteristics are observed by researchers. Small and Rosen (ibid) also show how to construct compensation variation when some product characteristics would be unobserved as in general RUM. Under mild restrictions of the RUM model in the same paper, for instance, Small and Rosen indicate CV is equivalent to consumer surplus and that has a tractable form of log-sum– a sum shown in the denominator of predicted individual probabilities of choosing good j, in a typical multinomial logit model.

As the form of the log-sum of consumer surplus is extremely simple, Train (2009, page 55-57) provides an alternatively view for the derivation of its formula. First he defines the consumer surplus of a specific economic agent as a ratio of the maximum (indirect) utility derived among all possible alternatives over the marginal utility of her income. Yet the exact quantities of indirect utilities are unobserved to researchers due to the presence of individual-product specific errors. The welfare measurement of interest is thus the expected value of the consumer surplus for that economic agent– equivalent to the expected value of the maximum order statistics of indirect utilities among all alternatives. When the error terms, ϵ_{ijtm}, are specified as draws from independent identical standard Gumbel distribution and is linearly separated from the rest of observed utilities, researchers can easily derive the distribution of indirect utility functions. That is, conditional on observed part of the indirect utilities, the indirect utility functions also follow Gumbel distributions with location parameter calculated at the observed indirect utilities.

The indirect utility functions thus are independent but not identical Gumbel distributions. And the maximum among those indirect utility functions, the maximum order statistics, of the same agent can be shown to follow a Gumbel distribution whose location parameter is the log-sum in the denominator of the agent's predicted probabilities of choosing the discrete mutually exclusive alternatives. Then the expected maximum utility is the mean of the maximum order statistics, the log sum plus a scaling constant to all economic environments.

The calculation of marginal utility of income, however, requires more assumptions. One of the frequently assumed condition, as shown by Train (ibid), is that a dollar change in one of monetary attributes among all alternatives results in a dollar increase in the consumers' income level. For instance, a dollar deduction in premium of plans can lead to marginal increase in the agent's disposable income. Hence the absolute value of the estimated preference weight of that attribute, here the premium, on the margin, is the same as the marginal utility of income.

The discussion so far can be reiterated more formally. First rewrite equation Section 4.2 as following

$$u_{ijtm} = v(i,j,t,m) + \epsilon_{ijtm}, \forall j \setminus \{0\}, \tag{A.3}$$

$$u_{i0} = \epsilon_{i0tm}. \tag{A.4}$$

and $v(i,j,t,m) \equiv \beta_{imt}x_{ijtm} + \alpha_j + \xi_{jtm}$ is the observed part of the indirect utility of good j for agent i in market (t, m). Let $J(t, m)$ be the set of products varying by markets.

Consumers make decisions by maximizing their (indirect) utilities u_{ijt}. Recall the choice set is $\{j \in J \cup \{0\}\}$. Economists observe agent i purchase product a if

$$a = argmax_j u_{ijt}.$$

This is true if and only if $\{u_{iatm} > u_{ijtm}, \forall j \neq a \}$ This statement is equivalent to $\epsilon_{iatm} - \epsilon_{ijtm} > v(i,j,t,m) - v(i,a,t,m), \forall j \neq a$.

By specifying $\epsilon_{ijtm} \sim F(\epsilon_{ijtm})$ as standard Gumbel, it can be shown that

$$\begin{aligned} Prob(\{a = argmax_j u_{ijt}\}) =& Prob(\{\epsilon_{iatm} - \epsilon_{ijtm} > v(i,j,t,m) - v(i,a,t,m), \forall j \neq a\}) \\ =& \frac{exp(v(i,a,t,m))}{\sum_{j \in J(t,m) \cup \{0\}} exp(v(i,j,t,m))} \\ =& \begin{cases} \frac{exp(v(i,a,t,m))}{1+\sum_{j \in J(t,m)} exp(v(i,j,t,m))}, & \text{for } a \neq \{0\} \\ \frac{1}{1+\sum_{j \in J(t,m)} exp(v(i,j,t,m))}, & \text{for } a = \{0\}. \end{cases} \end{aligned} \tag{A.5}$$

See Train (2009, p.74-p.75) for the details for the derivation of the last equality.

Then by the discussion of consumer surplus, and use the coefficient of premium, β_{itm}, to replace the marginal utility of income, I have CS for the particular agent as

$$\begin{aligned} CS_{itm} =& \frac{1}{\beta_{prem,itm}} E(max_{j \in (t,m)} u_{ijtm}), \\ =& \frac{1}{\beta_{prem,itm}} (log(1 + \sum_{j \in J(t,m)} exp(v(i,j,t,m)) + \text{constant})). \end{aligned}$$

The last equality implies the change of consumer surplus is independent of the scaling constant for the same individual. Also it can be seen that the denominator of equation (A.5) appears in the log sum of CS. Now suppose research conduct two policy scenarios 0 and 1, the change of CS from 0 to 1 is

$$\begin{aligned} \Delta CS_{itm} =& \frac{1}{\beta_{prem,itm}} (log(1 + \sum_{j \in j \in (t,m)} exp(v^1(i,j,t,m))) - \\ & log(1 + \sum_{j \in j \in (t,m)} exp(v^0(i,j,t,m)))). \end{aligned} \tag{A.6}$$

Then the aggregated consumer surplus is the sum of CS_{itm} over relevant individuals over markets and periods.

Lastly, the same assumption on the marginal utility of income applies to calculations of willness to pay (WTP). Specifically, WTP is defined by a unit change in an product attribute, say one unit increase in a plan's benefit in terms of actuarially fair value, the amount of a personal income required to reduce such that the indirect utility level of that agent is not changed. In the case of linear utility of premiums and actuarially fair value, WTP is the absolute value of the ratio of $\beta_{AV}/\beta_{itm,prem}$.

A.2 Substituional Patterns

Very often after demand function is estimated, price coefficients are summarized by price elasticities of the products to evaluate market structures or for counterfactual economic policies. The calculation of elasticities, intuitively, is the same as that in random coefficient mixed logit model at individual level, even though the observed shares are aggregated by firms. The main insight, and as many other approaches using aggregated data, is based on the the same micro-foundation of individuals making choices. And parameters are identified by sufficient data variations.

In particular, in this thesis the estimates of parameters of interests are based on augmented individuals. This implies that once MCMC converges, the estimates joint posterior distributions of the parameters can be plugged into the calculation of elasticities by the the same technique as that conducted in standard econometric literature. I follow Nevo (2000b) and reiterate the formula by first conditional on an individual i and hence D_i, the predicted choice probability for good j is

$$s_{ijtm} \equiv \int_{(b_{itm},v_j)} Pr(z_{ijtm} = 1 | b_{itm}, x, v, \phi, \delta) dPr(b_{itm} | \mu_\beta, \Sigma_\beta) dPr(v_{jtm} | 0, \sigma_v^2).$$

The intergral in the equation above is over the first stage priors of unobserved part of the preference weights b and exogenous demand shocks v defined in Appendix B. In the Bayesian approach of this thesis, those quantities are indeed augmented in estimation. In next subsection I will discuss the computational issue given the augmented quantities. Here I use a conventional approach– the unobserved quantities are integrated out by drawing those variables through Halton sequences. Thus s_{ijtm} is a simulated predicted probability of choosing good j for individual i. The market share of the good in the particular market is the average probability of s_{ijtm} over the population of interest.

Let $s_{jtm} \equiv \sum_i s_{ijtm}/R$ be the predicted share of j over R individuals. A semi-elasticity of good j with respect to one dollar change of the price in good l is $\mathscr{S}_{jl} \equiv \frac{\partial s_{jtm}/s_{jtm}}{\partial p_{ltm}}$. And

$$\frac{\partial s_{jtm}}{\partial p_{ltm}} \equiv \begin{cases} \frac{1}{R}\sum_i \int_{(b_{itm},v_j)} \beta_{i,price} Pr(z_{ijtm}=1)(1-Pr(z_{ijtm}=1))\times \\ dPr(b_{itm}, v_{jtm}|\mu_\beta, \Sigma_\beta, \sigma_v^2) \text{ if l = j,} \\ \frac{-1}{R}\sum_i \int_{(b_{itm},v_j)} \beta_{i,price} Pr(z_{ijtm}=1)Pr(z_{iltm}=1)\times \\ dPr(b_{itm}, v_{jtm}|\mu_\beta, \Sigma_\beta, \sigma_v^2) \text{ if l} \neq \text{j.} \end{cases} \tag{A.7}$$

In practice, the integral of (b, v) is replaced by a set of fixed draws from the Halton sequences and averaged out for each draw, holding the rest of parameters at draws from MCMC at an iteration (k). Then the calculation using the same set of Halton sequence but the rest parameters are evaluated at another iteration.

The formula in equation (A.7) holds similarly for other cases like aggregated firm market shares or market shares by metal groups of health plans with respect to a price variation of a particular product. However, in Chapter 5 I approximate the elasticities of metal plans by replacing the individual probabilities of products with individual probabilities of aggregated metal level in equation (A.7). Thus the elasticities in Chapter 5 would be interpreted as the change of market shares of a representative metal plan with respect to the change of prices in any of the four representative typical metal plans and the outside option. Further research should focus on comparing those two distinct approaches of calculation.

A.2.1 Calculations by Posterior Distributions of Parameters

The quantities of interest in the previous two subsections are derived from economic theory and based on a particular economic agent. The micro-foundation could have made Bayesian statistics with augmented individuals more than appealing, as the estimation procedure depends on representative but heterogeneous individuals. And the joint posterior distribution sampled through each iteration in MCMC provides an intuitive plug-in approach for the calculations of distributions of CS, WTP or semi-elasticities.

However, it should be noticed that, while it's possible to keep augmented heterogeneous preference weights from each MCMC iteration, it would be potentially memory demanding for a local machine. One reason is the space needed grows with the augmented individuals, market size, as well as the number of iterations of MCMC. Alternatively, as the unobserved part of heterogeneous preference weights is specified as a joint normal distribution whose variance-covariance is also estimated, I adopt a hybrid approach of Bayesian and simulated maximum likelihood as the following. First, I keep the joint distribution of the key parameters of interests and exclude the heterogeneous preferences of augmented individuals or the random effects of products.

The exclusion of random parameters is not ignored, instead, they will be can be drawn from their first stage priors using Halton sequences. Then given an joint distribution of MCMC at iteration k, the draws of random coefficients are updated through their first stage prior given the same set of Halton draws. Then the policy variables like willness to pay, consumer surplus, or elasticities, are calculated through each of MCMC iteration – hence we have the distributions of policy variables.

Lastly, in many empirical application, the draws of demand shocks, v, are often set at a value of zero. This would potentially simplify the time to calculate the simulated probabilities. In my empirical exercise, I adopt the same assumption in Chapter 6 for the interest of time, while I drew the demand shocks in the calculation of elasticities in Chapter 5.

Appendix B

Control Function, Endogenous Price and Supply Side

B.1 Endogenous Prices from Bertrand Pricing Game

This chapter discusses how prices become endogenous in a market equilibrium. And how I address the endogenous problem empirically. Following the model for monopolistic competition of differentiated product as described in Berry et al. (1995), I consider the case when firms at time t in a particular market m offers multiple products $j \in J_f$. The monopolistic competition gives market power to firms. They will set prices above marginal costs. Let marginal costs be mc_j. Berry et al. (ibid.) assume a log-linear form of marginal costs $log(mc_j) = z_j\gamma + w_j$, where z_j is a vector of instruments consisting of firms' characteristics or product-characteristics. The log-linear specification allows the cost to be non-negative. They also allow the supply shocks of marginal costs, w_j, be endogenous to prices. Yet for simplicity, and to motive the need to instrument prices, I assume w_j is iid white noise.

The firms compete in the market by playing a price-bidding Bertrand game. Firms offer prices for each of their products to gain market shares and earn profits. Assume in the model firms know the characteristics of consumers and can compute product-level market shares for a market is

$$s_{jt}(x_t, p_t, \xi_t) = \int Pr(y_{ijt} = 1|\beta_i, x_t, \xi_t) dP(\beta_i|\mu_\beta, \Sigma_\beta) \tag{B.1}$$

$$= \int \frac{exp(\xi_{jt} + \beta_i x_{jt})}{\sum_j exp(\xi_{jt} + \beta_i x_{jt})} dP(\beta_i|\mu_\beta, \Sigma_\beta). \tag{B.2}$$

For simplicity, x_t, the collection of product-level attributes is given. This could be a reasonable assumption since ACA requires every qualified health insurance plan to offer essential health benefits in individual markets. Also given is the vector of demand shocks ξ_t. They are not observed by researchers but available to firms when companies set up prices or premiums. Then given the total number of potential consumers M, the demand of product j is $Ms_{jt}(x_t, p_t, \xi_t)$. Firms set up prices $\{p_j\}$ and maximize their net profits as the sum of each product in their product portfolios $\{j \in J_f\}$.

$$\Pi_f = \sum_{j \in J_f} (p_j - mc_j) M s_{jt}(x_t, p_t, \xi_t). \tag{B.3}$$

I assume interior solutions of $\{p_j\}$. The first order condition above is

$$s_{jt}(x_t, p_t, \xi_t) + \sum_{l \in J_f} (p_l - mc_l) \frac{\partial s_{lt}(x_t, p_t, \xi_t)}{\partial p_j} = 0, \forall j. \tag{B.4}$$

We can collect all j equations in (B.4) and write prices in vectors

$$p_t = mc_t + \Delta(x_t, p_t, \xi_t)^{-1} s_t(x_t, p_t, \xi_t), \tag{B.5}$$

where $\Delta(x_t, p_t, \xi_t)$ is an "ownership" matrix consisting of changes in product-level market shares with respect to changes in prices

$$(\Delta(x_t, p_t, \xi_t))_{lk} = \begin{cases} \frac{\partial s_{lt}(x_t, p_t, \xi_t)}{\partial p_k}, & \text{if } l \text{ and } k \in J_f \\ 0, & \text{if } l \text{ or } k \notin J_f. \end{cases} \tag{B.6}$$

The second term of the pricing equation (B.5) suggests the simultaneity of both prices and unobserved demand shocks. As long as the demand shocks are unobserved to researchers, prices on the demand side estimation become endogenous by the structure of this model and economic equilibrium.

When researchers are interested in estimating the parameters on the supply side for marginal costs, the pricing equation can be re-written as such

$$mc_t = p_t - \Delta(x_t, p_t, \xi_t)^{-1} s_t(x_t, p_t, \xi_t). \tag{B.7}$$

It should be noticed that the ownership matrix and the product-level shares can be calculated once the parameters on the demand side are estimated. These quantities are imputed through drawing a hypothetical population for equation (B.1). Given a draw of R population, the simulated product-level shares and the simulated ownership matrix have the below formula

$$\hat{s}_{jt}(x_t, p_t, \xi_t) = \frac{1}{R} \sum_i^R Pr(y_{ijt} = 1 | \beta_i, x_t, \xi_t), \tag{B.8}$$

$$\frac{\partial \hat{s}_{jt}(x_t, p_t, \xi_t)}{\partial p_j} = \frac{1}{R} \sum_i^R Pr(y_{ijt} = 1 | \beta_i, x_t, \xi_t) * (1 - Pr(y_{ijt} = 1 | \beta_i, x_t, \xi_t))\beta_{i,price}, \tag{B.9}$$

$$\frac{\partial \hat{s}_{jt}(x_t, p_t, \xi_t)}{\partial p_k} = \frac{1}{R} \sum_i^R -Pr(y_{ijt} = 1 | \beta_i, x_t, \xi_t) * Pr(y_{ikt} = 1 | \beta_i, x_t, \xi_t)\beta_{i,price}. \tag{B.10}$$

In Berry et al. (1995), these quantities require simulations after parameters on demand side are solved by moment conditions. In Bayesian BLP with either product-level or firm-level data, the probabilities of individual choices and (β_i, ξ_t) are indeed treated as data and calculated in each MCMC iteration. Therefore, in Bayesian BLP, one can estimate the full economic models in two steps. The first step requires estimating demand side by a reduced form pricing equation with instruments till the MCMC chain converges (Yang et al., 2003; Musalem et al., 2009). The second step is to use the quantities of (β_i, ξ_t) along with product characteristics x_t in each converged MCMC run to recover parameters on the supply side.

Alternatively, Yang et al. (2003) shows how to estimate the model with endogenous prices by fully specifying both demand and supply shocks in the likelihood function. The trade-off between using fully specified likelihood of pricing equations and reduced form pricing equations would be statistical efficiency versus estimation convenience.

B.2 Control Function

The control function approach in this thesis follows Train (2009, p.337) and assume a set of observed valid instrument, Z, is readily available. A reduced form pricing equation is correlated with unobserved product specific constant in the following way

$$\begin{pmatrix} \xi_j \\ \eta_j \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\xi^2 & cov(\xi, \eta) \\ cov(\xi, \eta) & \sigma_\eta^2 \end{pmatrix} \right], \tag{B.11}$$

$$x_j = Z\Gamma + \eta_j. \tag{B.12}$$

The strategy illustrated is to assume a bivariate normal distribution of unobserved errors in both demand and supply as the source of endogeneity of prices. Then from the conditional property of bivariate normal distributions, I could write ξ_j conditional on η_j as

$$\xi_j = \rho_{\xi,\eta}\frac{\sigma_\xi}{\sigma_\eta}\eta_j + \sqrt{(1-\rho^2_{\xi,\eta})\sigma^2_\xi}\tilde{v}_j \tag{B.13}$$

$$= \frac{cov(\xi,\eta)}{\sigma^2_\eta}\eta_j + \sqrt{(1-\frac{cov^2(\xi,\eta)}{\sigma^2_\eta\sigma^2_\xi})\sigma^2_\xi}\tilde{v}_j \tag{B.14}$$

$$\rho_{\xi,\eta} \equiv \frac{cov(\xi,\eta)}{\sigma_\xi\sigma_\eta}, \sigma_{\xi|\eta} \equiv \sqrt{\sigma^2_\xi(1-\frac{cov^2(\xi,\eta)}{\sigma^2_\eta\sigma^2_\xi})}. \tag{B.15}$$

So x_j is endogenous with ξ_j because of it is correlated with η_j; and $\tilde{v}_j$ is a standard normal distribution by construction. Or one can write $v_j \equiv \tilde{v}_j\sigma_{\xi|\eta} \sim N(0,\sigma^2_v)$.

This control function will substitute equation (B.14) into the utility function and requires estimating only one variables, $\phi \equiv \frac{cov(\xi,\eta)}{\sigma^2_\eta}$. The strategy requires estimating η_j like a first stage regression in two-stage least squares. It's possible, however, in Bayesian analysis researchers could simply estimate η_j once and treats the estimated residuals as data. And iterations through MCMC necessarily results in the posterior distribution of ϕ. The strategy in Bayesian thus differs slightly from the classic method of control function that requires standard error adjustments.

In addition, I report the comparison of including updates of η_j in the following Table B.1. When the updates of η_j is included, the estimation procedure is similar to conduct standard error adjustments of ϕ in classic statistics. Updating η_j along with demand estimates is conceptually similar to a simultaneous equation of Yang et al. (2003) for joint demand and supply, but with a reduced form estimation of the pricing equation.

In Table B.1 I consider the effects of including the first stage equation and update η in each MCMC iteration by both product-level or firm-level BLP. The number of simulated individual in each market is 150, and the setting is similar to that in Section 4.4 but here I exclude fixed effects for simplicity. It can be seen that for product-level BLP, whether updating η through MCMC iteration has little effects on the standard deviations or biases of the draws from posterior distributions. The difference is up-to second digits. In comparison, the difference between updating η or not in MCMC for firm-level BLP is also quite small. But in this experiment, updating η improve the performance of estimating the variance-covariance matrix by doubling the precision of the estimates. The bias of σ^2_{prem}, for example, drops to 0.06 if η is updated through MCMC iteration versus 0.12 if η is treated as fixed data. See the bottom panel of Table B.1 for details.

Table B.1: Comparison of Posterior Draws by Including Updates of η

Bayesian BLP	Update η?		prem	MOOP	ϕ	σ^2_{prem}	$cov(prem, moop)$	σ^2_{moop}	σ^2_v
product-level	No	s.d.	0.05	0.04	0.07	0.12	0.09	0.14	0.01
		Bias	0.00	0.00	0.00	0.01	0.00	0.03	0.01
	Yes	s.d.	0.04	0.04	0.07	0.11	0.09	0.13	0.01
		Bias	0.00	0.00	0.00	0.01	0.00	0.03	0.01
firm-level	No	sd	0.10	0.08	0.16	0.37	0.26	0.39	0.07
		Bias	0.01	0.01	0.02	0.12	0.04	0.20	0.01
	Yes	s.d.	0.09	0.08	0.16	0.33	0.26	0.36	0.06
		Bias	0.01	0.01	0.01	0.06	0.02	0.13	0.01
		True Value	-1.50	-0.50	-1.33	1.11	1.06	1.04	0.25

Note: Number of sampled or augmented individuals $R = 150$. Total MCMC iteration is 200,000 and convergence is observed after 100,000th iteration, after which every 500th draw is saved for the summary table.

Appendix C

MCMC Implemetation

C.1 Priors and Likelihoods for MCMC

Bayesian estimation requires specifying a joint prior distribution of all of parameters $(\mu_\beta, \Sigma_\beta, \sigma_v^2, \phi, \delta, \alpha)$[1] and a set of hyper-parameters θ_h for the prior distributions. Then the posterior distribution as a function of data (D_i, x, z_f, η) is proportional to $LL * Prior$ and can be derived by Bayes' theorem. Let the joint prior distribution be $G(\mu_\beta, \Sigma_\beta, \sigma_v^2, \phi, \delta, \alpha|\theta_h)$. The posterior distribution of equation (C.1) can be derived below

$$
\begin{aligned}
g((\mu_\beta, \Sigma_\beta, \sigma_v^2, \phi, \delta, \alpha|x, z, D, \eta, \theta_h) =& \frac{\Pi_{tm} L_{tm} G(\mu_\beta, \Sigma_\beta, \sigma_v^2, \phi, \delta, \alpha|\theta_h)}{\int_\Phi \Pi_{tm} L_{tm} G(\mu_\beta, \Sigma_\beta, \sigma_v^2, \phi, \delta, \alpha|\theta_h)} \\
&\propto \Pi_{tm} L_{tm}(\mu_\beta, \Sigma_\beta, \sigma_v^2, \phi, \delta, \alpha|x, D, z, \eta) G(\Phi|\theta_h), \\
\Phi \equiv& (\mu_\beta, \Sigma_\beta, \sigma_v^2, \phi, \delta, \alpha). \qquad \text{(C.1)}
\end{aligned}
$$

The exact formula of the posterior above is usually hard to compute due to the normalizing constant in the denominator. However, knowing the posterior up to a normalizing constant allows me to use MCMC for simulating from the joint posterior.

Specifically, I follow Train (2009, page 302) and assume a normal prior on μ_β, inverse Wishart (IW) priors on the variance of Σ_β, and an inverse Gamma (IG) distribution on σ_v^2, and uniform prior on $(\alpha, \delta, \phi))^2$. Since the parameters are assumed to be drawn from independent distributions, I have

[1]Or $(\mu_\beta, \Sigma_\beta, \sigma_v^2, \phi, \delta, \alpha, \Gamma, \sigma_\eta)$ when updating η_j is included. See Appendix C.2 for details. The following discussion assumes η_j is estimated and treated as data.

[2]Assuming uniform prior on those parameters corresponds to estimate fixed effects in classic approach. Alternatively, one can assume those parameters follow normal distributions and thus the estimates are similar to random effect models in frequentist school.

$$\mu_\beta \sim N(\beta_h, H_\beta), \tag{C.2}$$
$$\Sigma_\beta \sim IW(\nu_\beta, S_\beta), \tag{C.3}$$
$$\sigma_v^2 \sim IG(\nu_v, S_v), \tag{C.4}$$
$$(\phi, \delta, \alpha) \sim_{iid} uniform[-M, M], \tag{C.5}$$
$$G(\mu_\beta, \Sigma_\beta, \sigma_v^2, \phi, \delta, \alpha|\theta_h) = \frac{N(\beta_h, H_\beta) IW(\nu_\beta, S_\beta) IG(\nu_v, S_v)}{M^{length(\phi,\delta,\alpha)}}, \tag{C.6}$$
$$0 < M < \inf. \tag{C.7}$$

The function $length()$ imputes the number of parameters in the arguments. The benefit of these specifications is each of $(\mu_\beta, \Sigma_\beta, \sigma_v^2)$ has a marginal posterior distributions in conjugated formula[3], conditional on data and other parameters. Thus estimating those parameters can be proceeded either by calling well-tested routines and reduce significant amount of time for testing and computation via Gibbs sampling.

So far the model specified is enough for estimating parameters, given that η is treated as data and predicted by valid instruments from the reduced form pricing equation in Appendix B. Alternatively, the predicted value of η can be included in MCMC by a single-equation Bayesian linear regression with known conditional variance of $\sigma_{\eta|\xi}$ and conditional mean of $\eta|\xi$. Parameters in Figure 4.2 can thus be drawn using Metropolis within Gibbs given the extra data of $\eta|\xi$. Let (k) denote the posterior draws of parameters in the k-th iteration. I have the draws of $(\{z_{iftm}^{(k)}\}, \alpha^{(k)}, \{\beta_{itm}^{(k)}\}, \mu_\beta^{(k)}, \Sigma_\beta^{(k)}, \sigma_v^{2(k)}, \delta^{(k)}, \phi^{(k)}, \{v_{jt}^{(k)}\})$ from their joint posterior distributions conditional on the set of data $\{(D, X, \eta)\}$ and portfolio market shares. The MCMC algorithm is based on Gibbs sampling where a subset of parameters is sampled from the parameters' full conditional posterior distribution conditional on the rest of parameters and data. The procedure is further detailed as following steps[45].

C.2 Steps of Gibbs Sampling

I iterate my MCMC estimation in the following order:

1. Given draws of $(\{\beta_{itm}^{(k)} - \delta^{(k)} D_{itm}\}, \Sigma_\beta^{(k)})$, draw $\mu_\beta^{(k+1)}$ from its normal posterior distribution with its conjugate prior $N(\beta_h, H_\beta)$. This step is the same as

[3]For example, if the prior distribution of μ_β is normal, then its posterior distribution is also normal.

[4]This algorithm describes the steps when endogeneity is accounted by the control function approach. If no endogeneity is assumed in the model, then v_{jtm} is equivalent to the demand shock ξ_{jmt}, and $\phi \equiv 0$.

[5]The data generating process is depicted in Figure 4.2 by a uni-directional acyclic graph. Thus in many steps the full conditional posterior distributions are simplified to well-known analytic forms.

multivariate Bayesian linear regressions with known variance covariance $\Sigma_\beta^{(k)}$.

2. Given $\{(e_{itm}^{(k)} \equiv \beta_{itm}^{(k)} - \delta^{(k)} D_{itm} - \mu_\beta^{(k+1)})\}$, draw $\Sigma_\beta^{(k+1)}$ from its inverse Wishart posterior distribution with its conjugate prior $IW(\nu_\beta, S_\beta)$.

3. Given $(\{z_{iftm}^{(k)}\}, \alpha^{(k)}, \delta^{(k)} D_{itm}, \mu_\beta^{(k+1)}, \Sigma_\beta^{(k+1)}, \phi^{(k)}\eta_j^{(k)}, \{v_{jtm}^{(k)}\})$, update the mean of individual specific preference weights. Let $b_{itm}^{(k)} \equiv (\beta_{itm}^{(k)} - \delta^{(k)} D_{itm})$ and draw $b^{(k+1)}$ from its full conditional posterior by MH algorithm. The proposal distribution is $N((\beta_{itm}^{(k)} - \delta^{(k)} D_{itm}), \rho_1^2 \Sigma_\beta^{(k+1)})$, random walk. Notice that when $\delta^{(k)} D_{itm}$, the slope terms or observed heterogeneity, are deducted from $\beta_{itm}^{(k)}$, the distribution of $\{b_{itm}\}$ follows $N(\mu_\beta, \Sigma_\beta)$. Thus this step of MH draw is the same as that without observed demographics and can be proceeded as suggested by (Train, 2009, Chapter 12). The target density for each b_{itm} is proportional to $P(z_{ift} = 1) N(b_{itm} | \mu_\beta, \Sigma_\beta)$, the product of a current choice probability of firms at individual level and the density of the preference weights without slope terms. Lastly, ρ_1 is a tuning parameter that will be explained later.

4. When D_i, demographic attributes are included in the model, and η_j is included, do the following. Otherwise, let $\beta_{itm}^{(k+1)} \equiv b_{itm}^{(k+1)}$ and update only η_j if necessary. If η_j is not included, skip this step. Given $(\{z_{iftm}^{(k)}\}, \{b_{itm}^{(k+1)}\}, \{v_{jt}^{(k+1)}\}, \alpha^{(k)}, \eta_j^{(k)})$, draw $(\delta^{(k+1)}, \phi^{(k+1)})$ jointly from the full conditional posterior distribution using MH algorithm where the proposal is $N((\delta^{(k)}, \phi^{(k)}), \rho_2^2 I)$. The prior is the improper uniform distribution so that the full conditional posterior is proportional to $\Pi_{tm}(L_{tm})$, the likelihood function[6]. Also update $\beta_{itm}^{(k+1)} \equiv b_{itm}^{(k+1)} + \delta^{(k+1)} D_{itm}$[7].

5. Given $(\{z_{iftm}^{(k)}\}, \{\beta_{itm}^{(k+1)}\}, \phi^{(k+1)}\eta_j^{(k)}, \{v_{jtm}^{(k)}\})$, update $\alpha^{(k+1)}$ from its full conditional posterior distribution using MH algorithm where the proposal is $N(\alpha^{(k)}, \rho_3^2 I)$. The prior is the improper uniform distribution so that the posterior is proportional to $\Pi_{tm}(L_{tm})$, the likelihood function. This step is similar to the previous one and is used when any fixed effects independent of product attributes are included, while the data matrix for α could be a high dimension sparse matrix of zero and ones.

---

[6] I combine drawing (ϕ, δ) in a single step as those parameters are interacted with product attributes and constant among individuals.

[7] An advantage of updating the individual heterogeneous by unobserved part, b_{itm}, and observed part δD_{itm} is a flexibility to specify D_{itm}. For example, one can allow two demographic characteristics to explain observed heterogeneity in prices while one demographic attribute to explain product designs. An alternative specification is to use the same set of demographics for both prices and product designs. Then the system of equation follows Bayesian seeming unrelated equations and can be easily estimated using existing packages. See Rossi et al. (2005) for details.

6. Given $(\alpha^{(k+1)}, \{\beta_{itm}^{(k+1)}\}, \phi^{k+1}\eta_j^{(k)}, \{z_{iftm}^{(k)}\})$, update $\{v_{jtm}^{(k+1)}\}$ jointly from its full conditional posterior by MH algorithm in which the proposal distribution is $N(v_{jtm}^{(k)}, \rho_4^2\sigma_v^{2,(k)})$. In this step, the full conditional posterior is proportional to $\Pi_{ijtm}Pr(z_{iftm} = 1)N(\{v_{jtm}\}|0, \sigma_v^2)$[8].

7. Given updated shocks of $\{v_{jtm}^{(k+1)}\}$, draw $\{\sigma_v^{2,(k+1)}\}$ from its conjugated posterior distribution of inverse Gamma distribution using Gibbs sampling. Recall that $v \sim N(0, \sigma_v^2)$ and σ_v^2 has a prior inverse Gamma distribution $IG(\nu_v, S_v)$.

8. Given $(\{\beta_{itm}^{(k+1)}\}, \alpha^{(k)}, \phi^{k+1}\eta_j^{(k)}, \{v_{jtm}^{(k)}\})$, draw $\{z_{iftm}^{(k+1)}\}$ by the modified Gibbs sampling algorithm of Musalem et al. (2009). Notice that in this step only the choice probabilities of selected firms, $P(z_{ift}^{(k)} = 1)$, for pair individuals are calculated[9].

9. If the draw of η_j is required to account for sampling variations in the first stage of the control function approach, then update $\Gamma^{(k+1)}$ from its full conditional posterior distribution proportional to $\Pi_{tm}(L_{tm}) \times N(x_j - Z\Gamma, \sigma_\eta^{2,(k)})$ using MH algorithm. The proposal is $N(\Gamma^{(k)}, \rho_5 I)$ while the prior of Γ follows the improper uniform distribution. This step implies $\eta_j^{(k+1)} \equiv x_j - Z^{(k+1)}\Gamma$ is also updated if a proposed vector of Γ is accepted.

10. If the previous step updates any elements of $\eta_j^{(k+1)}$, then $\sigma_\eta^{2,(k+1)}$ is updated using Gibbs sampling from its inverse Gamma posterior distribution with its conjugate prior $IG(\nu_\eta, S_\eta)$.

11. Repeat the steps from (1) to (10) for the $(k+1)$th run. Stop the runs of MCMC until it is desired.

The random-walk Metropolis-Hasting (MH) algorithm from step 3 to step 9 requires a tuning or jumping parameter $\rho = (\rho_1, \rho_2, \rho_3, \rho_4, \rho_5)$. The proposed kernel in each step is a multivariate normal distribution centered at the draw of current iteration while its variance is multiplied with one of the tuning parameters. Intuitively, when the tuning parameter is higher, MH tends to accept too few new draws as proposed new parameters move too much. If the jump is lower, then in each MCMC run too many draws would be accepted when the proposed draws center more closely around the current parameters.

[8] However, as the product shocks in market m_1 and m_2 are assumed to be independent, the draws of v_{jtm_1}, v_{jtm_2} can be proceeded independently and simultaneously by market and time combinations. For example, conditional on (t_1, m_1), the full conditional posterior in this pair of market is proportional to $\Pi_{ij\in(t_1,m_1)t_1m_1}P(z_{ift_1m_1} = 1)N(\{v_{jt_1m_1}\}|0, \sigma_v^2)$. This implies that in some occasions a market-wise update can be computed using parallel algorithm when computation speed is a concern.

[9] This step is skipped if D is a zero vector or no demographic attributes are included.

Theoretically the tuning parameter can be adjusted such that each of MH algorithm meets an optimal acceptance rate as suggested by Gelman et al. (1996) so that the random walks can cover well the true posterior distributions. For example, the optimal accept rate for drawing one parameter is about 44% and decreases with the number of parameters. When there are more than 10 parameters in one step of MH, the optimal acceptance rate is about 26.7%. In practice, I adjust the vector of ρ for every iteration such that the vector of acceptance rates in each iteration is close to the optimal acceptance rates depending on the number of parameters.

Next I detail how MH algorithms are conducted in each of steps above using pseudo-codes. The algorithm is implemented in R and is accessible as a package, *MultiProdBLP*.

C.3 Pseudocodes

My MH for step (3) is drawn by the following algorithm for each market-period (t, m) and notice that individuals are conditionally independent in the market period. The full conditional posterior of each individual is

$$\begin{aligned}&f(b_{itm}|(z_{iftm}, \mu_\beta, X_{tm}, \Sigma_\beta, \alpha, \phi\eta, \delta D_{itm}, v_{tm}))\\ &\propto Pr(z_{iftm} = 1|b_{itm}, X_{tm}, \alpha, \phi\eta, \delta D_{itm}, v_{tm})N(b_{itm}|\mu_\beta, \Sigma_\beta).\end{aligned}$$

The update is based on the un-normalized density as the product of an individual choice probability and the first stage prior. The detail is shown below

---

Algorithm 1: Update b_{itm}

Input: $b_{itm}^{(k)}$
Output: $b_{itm}^{(k+1)}$
Data: $(z_{iftm}^{(k)}, X_{tm}, \alpha^{(k)}, \phi^{(k)}\eta^{(k)}, \delta^{(k)}D_{itm}, v_{tm}^{(k)}, \mu_{\beta}^{(k+1)}, \Sigma_{\beta}^{(k+1)})$

1 Sample $u \sim uniform(0,1)$.
2 Draw $e_{itm} \sim N(0_K, \rho_1^2 \Sigma_{\beta}^{(k+1)})$
3 Propose $b_{itm}^{(k+1)} = b_{itm}^{(k)} + e_{itm}, \forall i$.
4 **while** $(i,t,m) <= (R,T,M)$ **do**
5   Calculate the ratio
$$MH = \frac{Pr(z_{iftm}^{(k)}=1|b_{itm}^{(k+1)}, X_{tm}, \alpha^{(k)}, \phi^{(k)}\eta^{(k)}, \delta^{(k)}D_{itm}, v_{tm}^{(k)})N(b_{itm}^{(k+1)}|\mu_{\beta}^{(k+1)}, \Sigma_{\beta}^{(k+1)})}{Pr(z_{iftm}^{(k)}=1|b_{itm}^{(k)}, X_{tm}, \alpha^{(k)}, \phi^{(k)}\eta^{(k)}, \delta^{(k)}D_{itm}, v_{tm}^{(k)})N(b_{itm}^{(k)}|\mu_{\beta}^{(k+1)}, \Sigma_{\beta}^{(k+1)})}.$$
6   **if** $u <= MH$ **then**
7   Return $b_{itm}^{(k+1)}$ `/* Accept the proposal. */`
8   **else**
9   Return $b^{(k+1)} = b^{(k)}$. `/* Reject the proposal. */`
10 **if** *the average acceptance rate is too high* **then**
11   raise ρ_1^2
12 **else**
13   lower ρ_1^2.
14 Go to step 4.

The step (4) requires calculating the full conditional posterior that is proportional to joint likelihoods

$$f((\delta,\phi)|(z_{iftm}, \mu_{\beta}, X_{tm}, \Sigma_{\beta}, \alpha, \eta, D_{itm}, v_{tm})) \propto \Pi_{itm} Pr(z_{iftm} = 1|b_{itm}, X_{tm}, \alpha, \phi\eta, \delta D_{itm}, v_{tm}).$$

The MH algorithm at this step is

Algorithm 2: Update (δ, ϕ)

Input: $(\delta^{(k)}, \phi^{(k)})$
Output: $(\delta^{(k+1)}, \phi^{(k+1)})$
Data: $(z_{iftm}^{(k)}, \{b_{itm}^{(k+1)}\}, X_{tm}, \alpha^{(k)}, \eta^{(k)}, D_{itm}, v_{tm}^{(k)}, \mu_{\beta}^{(k+1)}, \Sigma_{\beta}^{(k+1)})$

1 Sample $u \sim uniform(0,1)$.
2 Draw $r \sim N(0, \rho_2^2)$
`/* The dimension of` r `is the total dimension of` (δ, ϕ) `*/`
3 Propose $(\delta^{(k+1)}, \phi^{(k+1)}) = (\delta^{(k+1)}, \phi^{(k+1)}) + r$.
`/* Prior distribution is uniform. So the ratio of joint density of` (δ, ϕ) `is 1. */`
4 Calculate the ratio $MH = \frac{\Pi_{itm} Pr(z_{iftm}^{(k)}=1|b_{itm}^{(k+1)}, X_{tm}, \alpha^{(k)}, \phi^{(k+1)}\eta^{(k)}, \delta^{(k+1)} D_{itm}, v_{tm}^{(k)})}{\Pi_{itm} Pr(z_{iftm}^{(k)}=1|b_{itm}^{(k+1)}, X_{tm}, \alpha^{(k)}, \phi^{(k)}\eta^{(k)}, \delta^{(k)} D_{itm}, v_{tm}^{(k)})}$.
`/* Now update the slope terms of observed demographics and the coefficients for control function */`
5 **if** $u <= MH$ **then**
6   Return $(\delta^{(k+1)}, \phi^{(k+1)})$ `/* Accept the proposal. */`
7 **else**
8   Return $(\delta^{(k+1)}, \phi^{(k+1)}) = (\delta^{(k)}, \phi^{(k)})$. `/* Reject the proposal. */`
9 **if** *the average acceptance rate is too high* **then**
10   raise ρ_2^2
11 **else**
12   lower ρ_2^2.
13 Go to step 5.

The step (5) also uses a MH that is similar to step (4) and its full conditional posterior is also proportional to joint likelihoods

$$f(\alpha|(z_{iftm}, \mu_{\beta}, X_{tm}, \Sigma_{\beta}, \phi\eta, \delta D_{itm}, v_{tm}))$$
$$\propto \Pi_{itm} Pr(z_{iftm} = 1|b_{itm}, X_{tm}, \alpha, \phi\eta, \delta D_{itm}, v_{tm}).$$

The MH algorithm at this step is

Algorithm 3: Update (α): Parameters of Fixed Effects

Input: $\alpha^{(k)}$
Output: $\alpha^{(k+1)}$
Data: $(z_{iftm}^{(k)}, \{b_{itm}^{(k+1)}\}, X_{tm}, \alpha^{(k)}, \phi^{(k+1)}\eta^{(k)}, \delta\phi^{(k+1)}D_{itm}, v_{tm}^{(k)})$

1 Sample $u \sim uniform(0,1)$.
2 Draw $r \sim N(0, \rho_3^2)$
/* The dimension of r is the total dimension of α. */
3 Propose $\alpha^{(k+1)} = \alpha^{(k+1)} + r$.
/* Prior distribution is uniform. So the ratio joint density of α is 1. */
4 Calculate the ratio $MH = \frac{\Pi_{itm} Pr(z_{iftm}^{(k)}=1|b_{itm}^{(k+1)}, X_{tm}, \alpha^{(k+1)}, \phi^{(k+1)}\eta^{(k)}, \delta^{(k+1)}D_{itm}, v_{tm}^{(k)})}{\Pi_{itm} Pr(z_{iftm}^{(k)}=1|b_{itm}^{(k+1)}, X_{tm}, \alpha^{(k)}, \phi^{(k+1)}\eta^{(k)}, \delta^{(k+1)}D_{itm}, v_{tm}^{(k)})}$.
/* Now update the slope terms of observed demographics and the coefficients for control function */
5 **if** $u <= MH$ **then**
6   Return $\alpha^{(k+1)}$ /* Accept the proposal. */
7 **else**
8   Return $\alpha^{(k+1)} = \alpha^{(k)}$. /* Reject the proposal. */
9 **if** *the average acceptance rate is too high* **then**
10   raise ρ_3^2
11 **else**
12   lower ρ_3^2.
13 Go to step 6.

The step (6) updates the draws of $\{v_{jtm}\}$ for each pair of t and m. In a pair of (t, m), the full conditional posterior is also proportional to joint likelihoods and the prior as

$$
\begin{aligned}
& f(\{v_{jtm}\}|(z_{iftm}, X_{tm}, \sigma_v^2, \phi\eta, \delta D_{itm}, v_{tm})) \\
\propto & \Pi_i Pr(z_{iftm} = 1|b_{itm}, X_{tm}, \alpha, \phi\eta, \delta D_{itm}, v_{tm}) \Pi_j N(v_{jtm}|0, \sigma_v^2).
\end{aligned}
$$

Algorithm 4: Update v_{jtm}, Demand Shocks

Input: $v_{jtm}^{(k)}$

Output: $v_{jtm}^{(k+1)}$

Data: $(z_{iftm}^{(k)}, X_{tm}, \alpha^{(k+1)}, \phi^{(k+1)}\eta^{(k)}, \delta^{(k+1)}D_{itm}, \sigma_v^{2,(k)})$

1 Sample $u \sim uniform(0,1)$.

/* Proposal distribution is random walk with variance $\rho_4^2\sigma_v^{2,(k)}$. */

2 Draw $e_{jtm} \sim N(0, \rho_4^2\sigma_v^{2,(k)})$

3 Propose $v_{jtm}^{(k+1)} = v_{jtm}^{(k)} + e_{jtm}, \forall j$ in (t,m).

/* Prior distribution is $N(0,\sigma_v^{2,(k)})$. */

4 **while** $(t,m) <= (T,M)$ **do**

5 Calculate the ratio

$$MH = \frac{\Pi_i Pr(z_{iftm}^{(k)}=1|b_{itm}^{(k+1)}, X_{tm}, \alpha^{(k+1)}, \phi^{(k+1)}\eta^{(k)}, \delta^{(k+1)}D_{itm}, v_{tm}^{(k+1)})\Pi_j N(v_{jtm}^{(k+1)}|0,\sigma_v^{2,(k)})}{\Pi_i Pr(z_{iftm}^{(k)}=1|b_{itm}^{(k+1)}, X_{tm}, \alpha^{(k+1)}, \phi^{(k+1)}\eta^{(k)}, \delta^{(k+1)}D_{itm}, v_{tm}^{(k)})\Pi_j N(v_{jtm}^{(k)}|0,\sigma_v^{2,(k)})}.$$

6 **if** $u <= MH$ **then**

7 Return $v_{jtm}^{(k+1)}$ /* Accept the proposal. */

8 **else**

9 Return $v^{(k+1)} = v^{(k)}$. /* Reject the proposal. */

10 **if** *the average acceptance rate is too high* **then**

11 raise ρ_4^2

12 **else**

13 lower ρ_4^2.

14 Go to step 7.

Step (8) updates augmented choices using Gibbs sampling. Recall that the data augmentation approach requires researchers to specify the number of augmented individuals in each market and time period. Let the number be R and the the vector of initial augmented individual choice indexes of firms be $(c_{1tm}^{(0)}, c_{2tm}^{(0)}, ..., c_{Rtm}^{(0)})$. And define $c_{itm} = f$ if and only if $z_{iftm} = 1$. That is, the vector of c denotes a vector of categorical index while z is a vector of binary variables. By drawing $c_{itm}^{(0)}$ we have $\frac{\sum_i z_{iftm}^{(0)}=1}{R} = s_{ftm}$. The relative frequency of augmented individual choices is consistent with observed portfolio market shares[10].

To fix the intuition of the approach, suppose there is a market-time with three firms (f_0, f_1, f_2), and the augmented choice sequence of four individual at initial draw is $(c_{1tm}^{(0)} = f_0, c_{2tm}^{(0)} = f_0, c_{3tm}^{(0)} = f_1, c_{4tm}^{(0)} = f_2)$. The Gibbs sampling uses the same

[10] Initial the vector of choices can be either drawn from a multinomial distribution; or I follow Musalem et al. (2009) and draw the indexes by cumulative density function of portfolio market shares.

vector of augmented draws of $C \equiv (f_0, f_0, f_1, f_2)$ and proposes to reassign the sequence *pairwisely* by randomly selecting two individuals to update their choice indexes[11].

In the example of four augmented individuals, the full conditional posterior of $(z^{(0)}_{1ftm}, z^{(0)}_{2ftm}, z^{(0)}_{3ftm}, z^{(0)}_{4ftm})$ conditional on $(C, \{\beta^{(1)}_{itm}\}, \alpha^{(1)}, \phi^1\eta^{(0)}_j, \{v^{(1)}_{jtm}\})$ and all other augmented parameters is

$$\begin{aligned}
&Pr((z^{(0)}_{1ftm}, z^{(0)}_{2ftm}, z^{(0)}_{3ftm}, z^{(0)}_{4ftm}|C, \{\beta^{(1)}_{itm}\}, \alpha^{(1)}, \phi^1\eta^{(0)}_j, \{v^{(1)}_{jtm}\}) \\
=&const \times Pr(z^{(0)}_{1ftm}, z^{(0)}_{2ftm}, z^{(0)}_{3ftm}, z^{(0)}_{4ftm}|C, MNL(i)) \\
=&const \times \Pi^4_{i=1} MNL(i) \\
MNL(i) =&\Pi_f(\sum_{j\in f} \frac{exp(\beta^{(1)}_{itm}X_{jtm} + \alpha^{(1)} + \phi^{(1)}\eta^{(0)}_j + v^{(1)}_j)}{\sum_{j\in\cup f} exp(\beta^{(1)}_{itm}X_{jtm} + \alpha^{(1)} + \phi\eta^{(0)}_j + v^{(1)}_j)})^{z_{itm}}.
\end{aligned}$$

The second equality follows from the fact that once we have augmented individuals the full conditional joint distributions of choices can be calculated from an individual-level multinomial logit distribution. The third equality is from the fact that join distribution of $\{z_{iftm}\}$ is independent by individuals. The last equality is the definition of predicted probabilities for the choices of a firm– it is the summation of the choice probabilities over the product portfolio of the multi-product firm. The probability distributions of choices are multinomial distributions where the choice probabilities vary by individuals. The market share constraint C is satisfied by the permutation of c_{itm}. For an event $\{c_{itm} = f\}$ such that $z_{iftm}(c_{itm} = f) = 1$ and $z_{iftm}(c_{itm} \neq f) = 0$, its probability is

$$Pr(z_{iftm} = 1) = (\sum_{j\in f} \frac{exp(\beta^{(1)}_{itm}X_{jtm} + \alpha^{(1)} + \phi^{(1)}\eta^{(0)}_j + v^{(1)}_j)}{\sum_{j\in\cup f} exp(\beta^{(1)}_{itm}X_{jtm} + \alpha^{(1)} + \phi\eta^{(0)}_j + v^{(1)}_j)}), \forall f.$$

By random pairing, there are three possible pairs of individuals using Gibbs sampling $((1,2),(3,4))$,$((1,3),(2,4))$,$((1,4),(2,3))$, Suppose $((1,3),(2,4))$ is the random pair. Then the full conditional posterior of $(z^{(0)}_{1tm}, z^{(0)}_{3tm})$ conditional on $(z^{(0)}_{4tm}, z^{(0)}_{4tm})$ and other augmented parameters $(\{\beta^{(1)}_{itm}\}, \alpha^{(1)}, \phi^1\eta^{(0)}_j, \{v^{(1)}_{jtm}\})$ is

$$\begin{aligned}
&Pr((z^{(0)}_{1ftm}, z^{(0)}_{3ftm})|(c^{(0)}_{2tm} = f_0, c^{(0)}_{4tm} = f_2), \{\beta^{(1)}_{itm}\}, \alpha^{(1)}, \phi^1\eta^{(0)}_j, \{v^{(1)}_{jtm}\}) \\
=&\frac{const \times Pr(z^{(0)}_{1ftm}, z^{(0)}_{2ftm}, z^{(0)}_{3ftm}, z^{(0)}_{4ftm}|C, MNL(i))}{\sum_{z_{1ftm}, z_{2ftm}} const \times Pr(z^{(0)}_{1ftm}, z^{(0)}_{2ftm}, z^{(0)}_{3ftm}, z^{(0)}_{4ftm}|C, MNL(i))}.
\end{aligned} \tag{C.8}$$

[11]The random pairing process allows each individual has a positive probability of selecting any of the firms in the market-period. Without the positive probability of random pairs of individuals in the same market-period, the necessary condition for the convergence of MCMC will fail.

The denominator is the joint distribution of (z_{3ftm}, z_{4ftm}) by summing out all possible pairs of (z_{1ftm}, z_{2ftm}). It can be further verified that, conditional on $(c^{(0)}_{2tm} = f_0, c^{(0)}_{4tm} = f_2)$, either the event of updated choices $\{c^{(1)}_{1tm} = f_1, c^{(1)}_{3tm} = f_0\})$ or the event $\{c^{(1)}_{1tm} = f_0, c^{(1)}_{3tm} = f_1\}$ satisfies the constraint C that augmented choices need to be consistent with observed market shares. Thus the full conditional posterior distribution of (z_{1ftm}, z_{3ftm}) follows a Bernoulli distribution– either the pair of augmented choices of firms remain the same or they are swapped for next iteration. Let the event of "success" be $\{c^{(1)}_{1tm} = f_1, c^{(1)}_{3tm} = f_0\}$, or equivalently,$\{z^{(1)}_{1f_1tm} = 1, z^{(1)}_{3f_0tm} = 1\}$. That is, the paired individuals swap their choice indexes. Then the probability of success from equation (C.8) is, after applying the conditional independence of choice probabilities among individuals,

$$\begin{aligned}
&Pr(\{z^{(1)}_{1f_1tm}, z^{(1)}_{3f_0tm}\}|(c^{(0)}_{2tm} = f_0, c^{(0)}_{4tm} = f_2)) \\
&= \frac{Pr(z^{(0)}_{1f_1tm})Pr(z^{(0)}_{3f_0tm})}{Pr(z^{(0)}_{1f_0tm})Pr(z^{(0)}_{3f_1tm}) + Pr(z^{(0)}_{1f_1tm})Pr(z^{(0)}_{3f_0tm})}.
\end{aligned}$$

Therefore, the paired individuals of $(1, 3)$ have their choices swapped if a successful event is drawn with the above probability. Otherwise the individuals keep the same choice indexes for next iteration. Similarly the calculation holds for the pair individuals of $(2, 4)$.

However, if $((1, 2), (3, 4))$ is the random pairing at iteration $k = 0$, then conditional on $(c^{(0)}_{3tm} = f_1, c^{(0)}_{4tm} = f_2)$, the event that the updated choices of $(c^{(1)}_{1tm} = f_0, c^{(1)}_{2tm} = f_0)$ is the only pair that satisfies market share constraints. In other words, $(c^{(1)}_{1tm} = c^{(0)}_{1tm}, c^{(1)}_{2tm} = c^{(0)}_{2tm})$ and the choice sets for individuals $(1, 2)$ remain the same for next iteration. While individuals $(3, 4)$ will have a chance to swap their choices conditional on $(c^{(1)}_{1tm}, c^{(1)}_{2tm})$.

The pseudo-codes for this Gibbs sampling is

Algorithm 5: Update $\{c_{itm}\}$ and $\{z_{iftm}\}$

Input: Paired individuals $\{(a, b)\}$ and their choices $(z^{(k)}_{af_atm}, z^{(k)}_{bf_btm})$

Output: Updated Choices $(z^{(k+1)}_{af_atm}, z^{(k+1)}_{bf_btm})$

Data: $(\beta^{(k+1)}_{itm} X_{tm}, \alpha^{(k+1)}, v^{(k+1)}_{tm})$

1 **while** $(t, m) <= (T, M)$ **do**

2   Randomly pair any two individuals in (t, m) without replacements. Call the paired individuals (a, b). Denote their augmented choice indexes $(c^{(k)}_{atm}, c^{(k)}_{btm})$, and their corresponding binary variables $(z^{(k)}_{af_atm}, z^{(k)}_{bf_btm})$.

3   **if** $(c^{(k)}_{atm} \neq c^{(k)}_{btm})$ **then**

4   Let $p \equiv \frac{Pr(z^{(k)}_{af_btm})Pr(z^{(k)}_{bf_atm})}{Pr(z^{(k)}_{af_atm})Pr(z^{(k)}_{bf_btm})+Pr(z^{(k)}_{af_btm})Pr(z^{(k)}_{bf_atm})}$. Sample $u \sim Bernoulli(p)$.

5   **if** $u = 1$ **then**

6   Return $(z^{(k+1)}_{af_atm} = z^{(k)}_{af_btm}, z^{(k+1)}_{bf_btm} = z^{(k)}_{af_atm})$. `/* Accept the draw and swap choices. */`

7   **else**

8   Return $(z^{(k+1)}_{af_atm} = z^{(k)}_{af_atm}, z^{(k+1)}_{bf_btm} = z^{(k)}_{bf_btm})$.

9   **else**

10   Return $(z^{(k+1)}_{af_atm} = z^{(k)}_{af_atm}, z^{(k+1)}_{bf_btm} = z^{(k)}_{bf_btm})$.

11   Go to next pair.

12 Go to step 9.

The step (9) updates the draws of Γ by Metropolis Hasting algorithm,

---

Algorithm 6: Update $\Gamma, \{\eta_{jtm}\}$ in Control Function

---

Input: $\Gamma^{(k)}, \{\eta_{jtm}^{(k)}\}$

Output: $\Gamma^{(k+1)}, \{\eta_{jtm}^{(k+1)}\}$

Data: $(z_{iftm}^{(k+1)}, \{b_{itm}^{(k+1)}\}, X_{tm}, \alpha^{(k+1)}, \phi^{(k+1)}, \delta^{(k+1)} D_{itm}, v_{tm}^{(k+1)}, Z_{tm}, \sigma_\eta^{2,(k)})$

Sample $u \sim uniform(0, 1)$.

Draw $r \sim N(0, \rho_5^2)$

/* The dimension of r is the total dimension of Γ. */

Propose $\Gamma^{(k+1)} = \Gamma^{(k+1)} + r$.

Update proposed $\eta_{jtm}^{(k+1)} = price_{jtm} - \Gamma^{(k+1)} Z_{jtm}$.

/* Prior distribution on Γ is uniform. So the ratio joint density of Γ is 1. */

Calculate the ratio

$$MH = \frac{\Pi_{itm} Pr(z_{iftm}^{(k+1)}=1 | b_{itm}^{(k+1)}, X_{tm}, \alpha^{(k+1)}, \phi^{(k+1)} \eta^{(k+1)}, \delta^{(k+1)} D_{itm}, v_{tm}^{(k+1)}) \Pi_j N(\eta_{jtm}^{(k+1)} | 0, \sigma_\eta^{2,(k)})}{\Pi_{itm} Pr(z_{iftm}^{(k+1)}=1 | b_{itm}^{(k+1)}, X_{tm}, \alpha^{(k+1)}, \phi^{(k+1)} \eta^{(k)}, \delta^{(k+1)} D_{itm}, v_{tm}^{(k+1)}) \Pi_j N(\eta_{jtm}^{(k)} | 0, \sigma_\eta^{2,(k)})}.$$

if $u <= MH$ **then**

  Return $\Gamma^{(k+1)}$ and $\{\eta_{jtm}^{(k+1)}\}$ /* Accept the proposal. */

else

  Return $\Gamma^{(k+1)} = \Gamma^{(k)}$ and $\{\eta_{jtm}^{(k+1)} = \eta_{jtm}^{(k)}\}$. /* Reject the proposal. */

if *the average acceptance rate is too high* **then**

  raise ρ_5^2

else

  lower ρ_5^2.

Go to next step.

C.4 Adaptive Metropolis Hasting for Multi-modalities

Multi-modalities of the posterior distribution in portfolio demands have resulted in extra difficulties in estimation using MCMC. One particular issue is the algorithm would have a tendency to stick to one of the local extreme values in many runs while not exploring the whole possible ranges of the posteriors. Two possible solutions, as Geyer (1991) suggests, are either to allow Metropolis algorithms to propose a new state with sizable movements, or to use Metropolis-coupled Markov chain Monte Carlo (MCMCMC) that runs parallel chains at once to explore different modals. Both approaches will attempt to propose a new draw of parameters such that the draw moves faster from one modal to another of the posterior.

The first approach is the adaptive MCMC described in the earlier section. The adaptive algorithm is a tuning parameter ρ updated through pre-specified iteration to reach the optimal acceptance rates for all MH steps but the draws of heterogeneous preference weights. Indeed, to draw b_i, I follow Roberts & Rosenthal (2009) and draw $e_{itm} \sim 0.95N(0_K, 2.38/\sqrt{(k)}\Sigma_\beta^{(k+1)}) + .05N(0_K, 0.1/\sqrt{(k)}I)$ as the dimension or unobserved preference weights grow by the size of simulated individuals and markets.

If the adaptive algorithm within a single chain of MCMC fails, researchers can run multiple parallel MCMC chains by a cluster of computers. The approach of MCMCMC (Altekar et al., 2004) runs multiple chains at different values at once. Specifically, let Ξ be a vector of parameters in interest, $\psi(\Xi|y)$ the posterior distribution given the vector of data y. The MCMCMC algorithm let k chains of MCMC running simultaneously and each chain is raised to the power of "heated value" $0 < \alpha_1 < \alpha_2 < ...\alpha_k <= 1$. The heated value flattens a posterior distribution and thus allows a chain of MCMC draws moves rapidly within the heated chain, while the chain with $\alpha_k == 1$, or $\psi(\Xi|y)^{\alpha_k}$, is called "cold chain" and only draws from the cold chain will be used for analyses. Then after certain runs, two MCMC chains are randomly selected and swap their heated states by another metropolis algorithm. After two chains have swapped or stayed the same, the state from the cold chain is saved. Then all MCMC chains move to next iteration till desired runs for convergence.

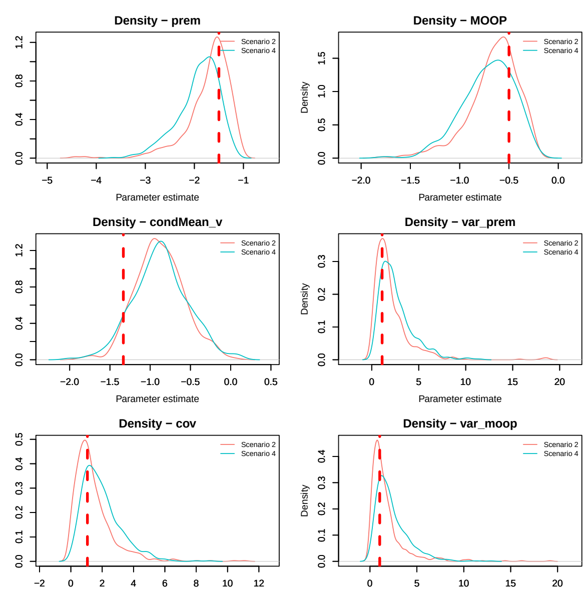
The original algorithm of Altekar et al. (ibid.) uses a cluster of computers and assigns each MCMC chain to one node. Then the swapping of two different chains occur only if each node returns the current state of draws at the same run. The process of synchronization as well as receiving and distributing draws of MCMC from multiple nodes could result in extra computation time if the time from overhead far exceeds the gain from paralleling. Instead, given today's multiple cores of personal computers the algorithm can be adapted as the running multiple chains on a local machine at once. Then after the burn-in periods, the chains can stop and swap states on the same local machine.

Appendix D

Additional Plots in Simulation

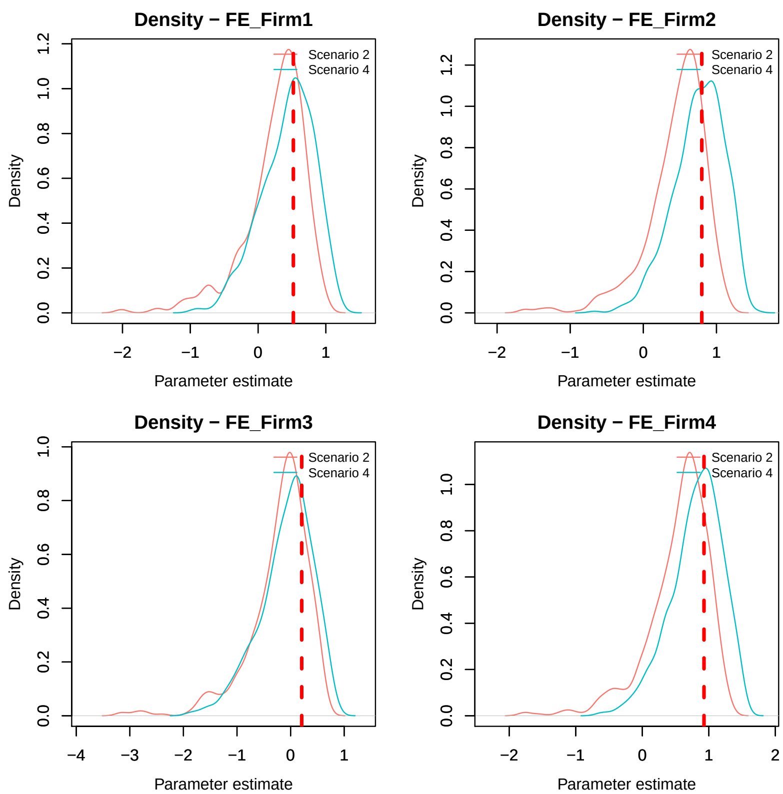
D.1 Additional Figures by Numbers of Simulated Individuals

Figure D.1: Posterior Distributions for Brand Choices, High Unobserved Heterogeneity, $\delta = 0$, $R = 200$



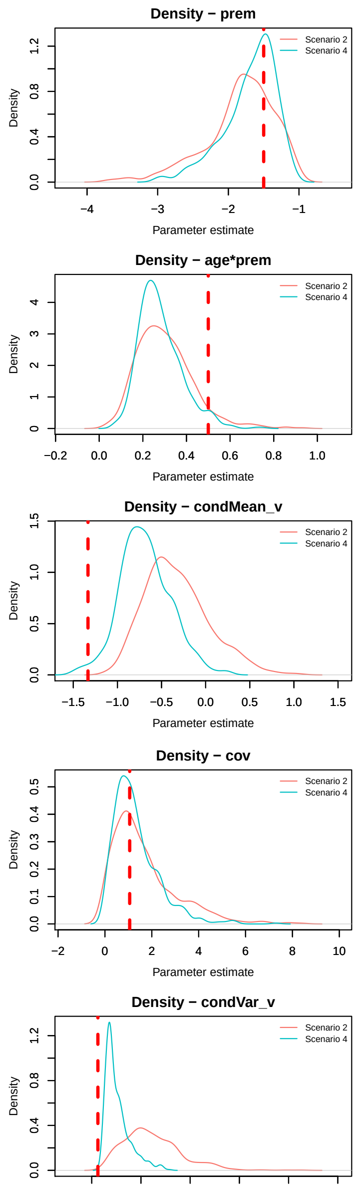
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $R = 200$.

Figure D.2: Posterior Distributions of Firm Fixed Effects, High Unobserved Heterogeneity, $\delta = 0$, $R = 200$



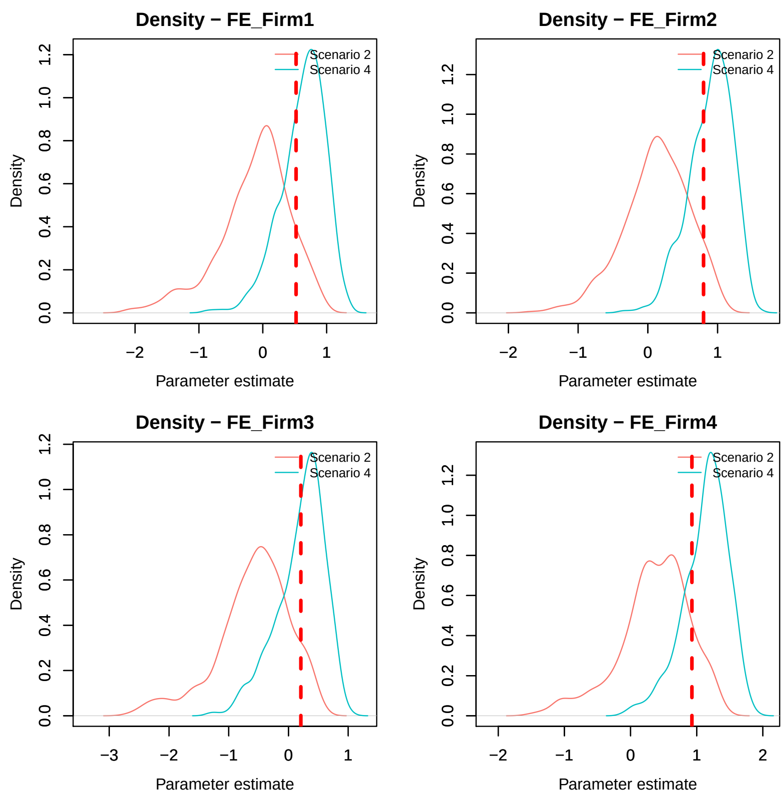
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $R = 200$.

Figure D.3: Posterior Distributions, High Unobserved Heterogeneity, $\delta \neq 0$, and $R = 200$



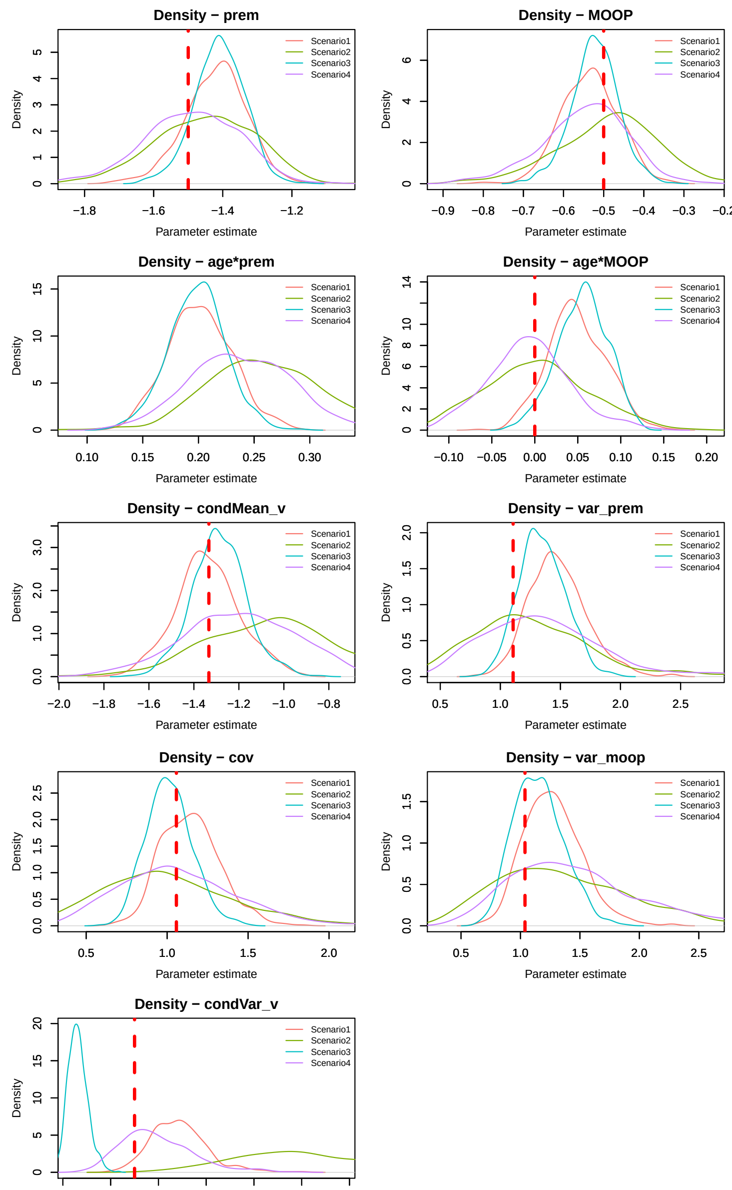
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $R = 200$.

Figure D.4: Posterior Distributions of Firm Fixed Effects, High Unobserved Heterogeneity, $\delta \neq 0$, $R = 200$



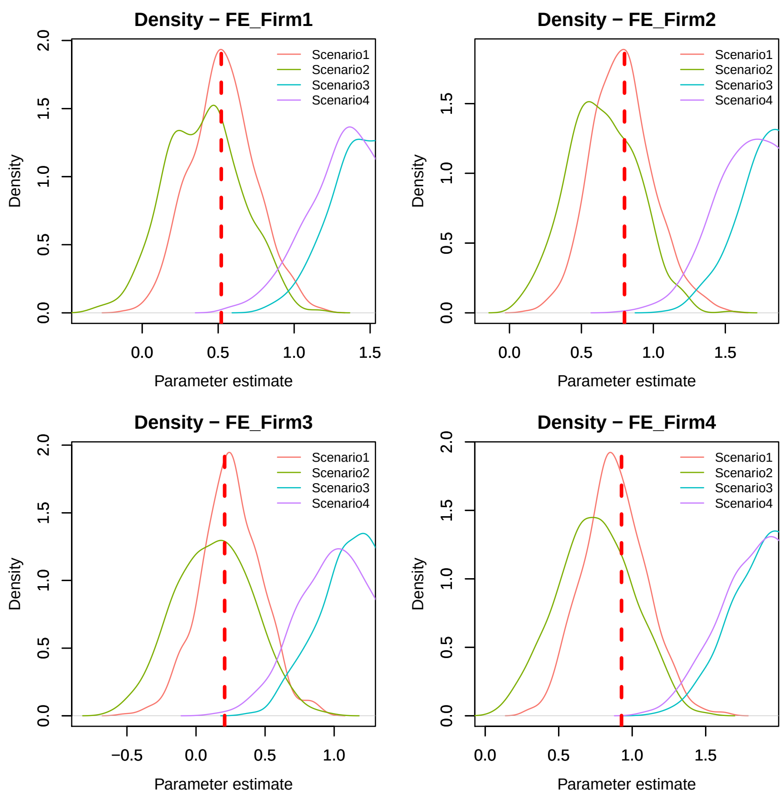
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $R = 200$.

Figure D.5: Posterior Distributions, High Unobserved Heterogeneity, $\delta \neq 0$,$R = 50$



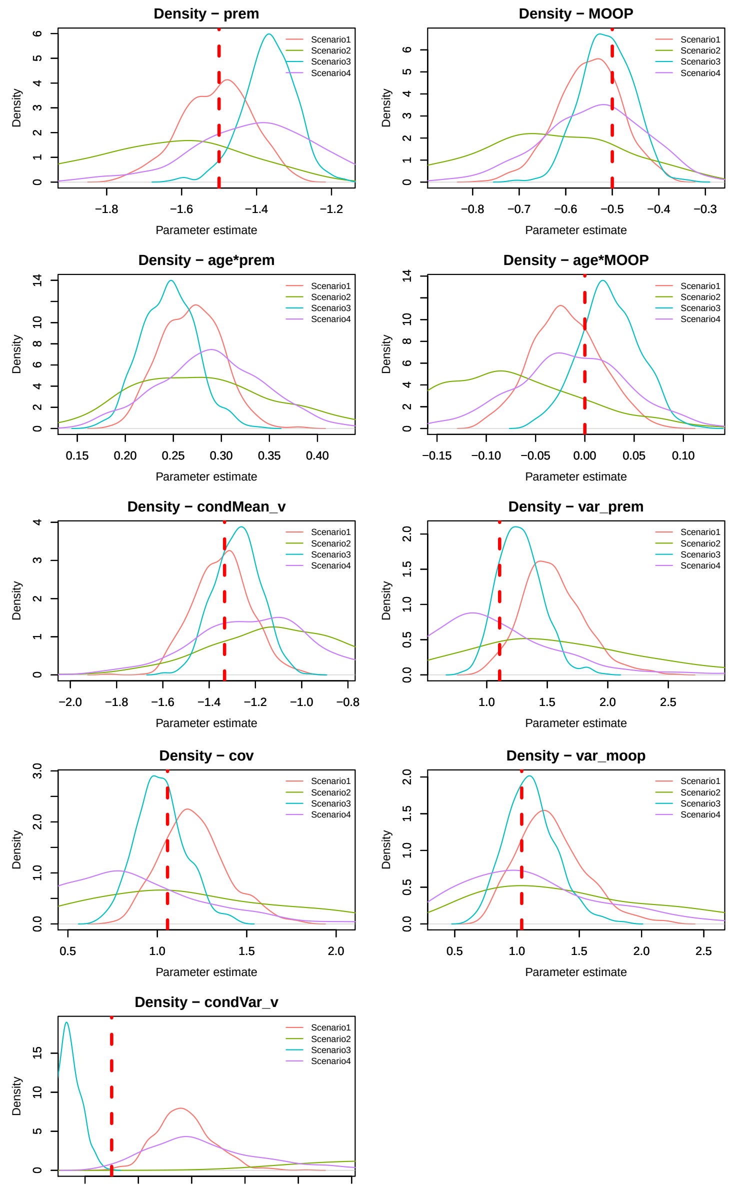
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $R = 50$.

Figure D.6: Posterior Distributions of Firm Fixed Effects, High Unobserved Heterogeneity, $\delta \neq 0$, $R = 50$



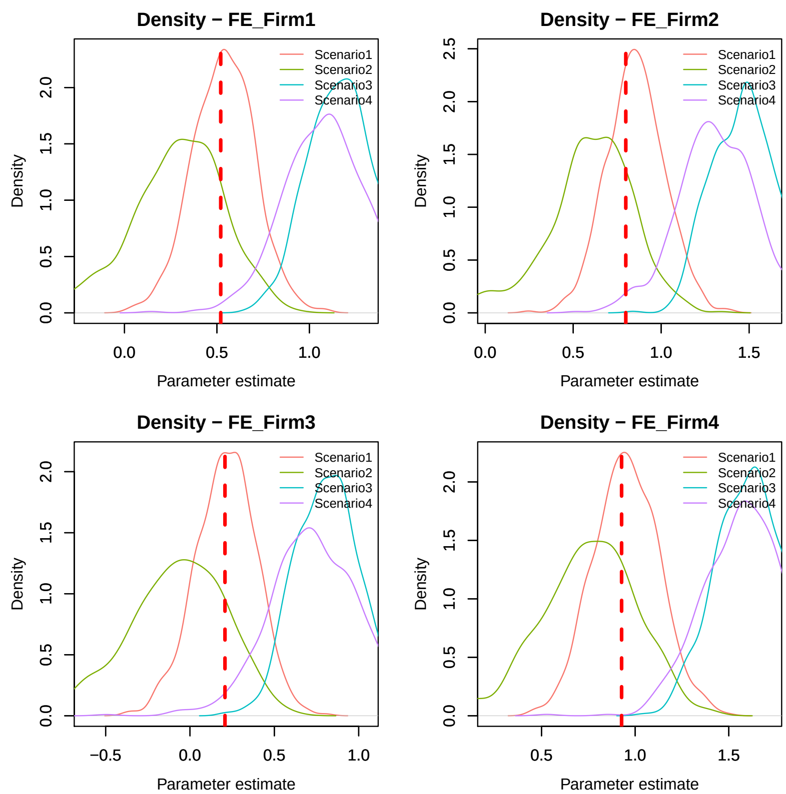
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $R = 50$.

Figure D.7: Posterior Distributions, High Unobserved Heterogeneity, $\delta \neq 0$, and $R = 100$



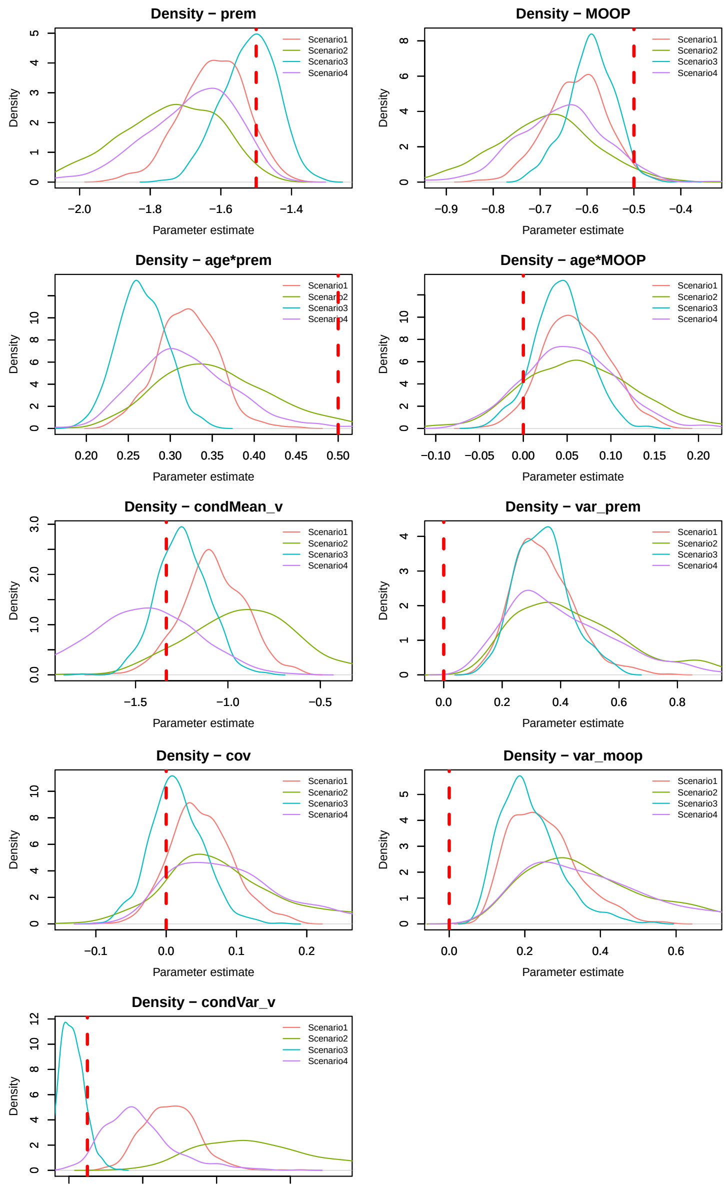
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $R = 100$.

Figure D.8: Posterior Distributions of Firm Fixed Effects, High Unobserved Heterogeneity, $\delta \neq 0$, $R = 100$



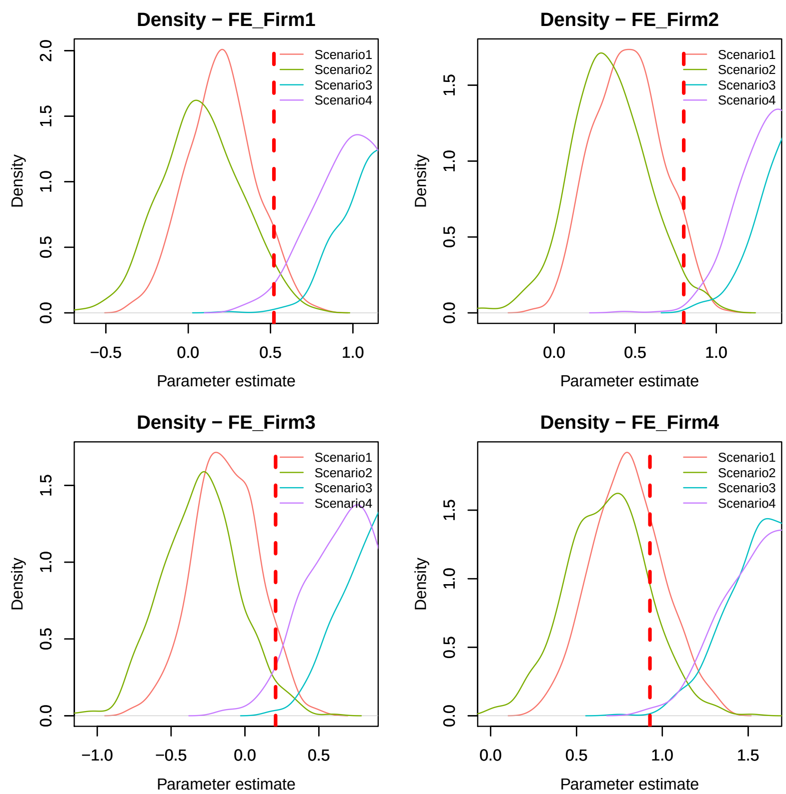
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $R = 100$.

Figure D.9: Posterior Distributions, Low Unobserved Heterogeneity, $\delta \neq 0$, $R = 50$



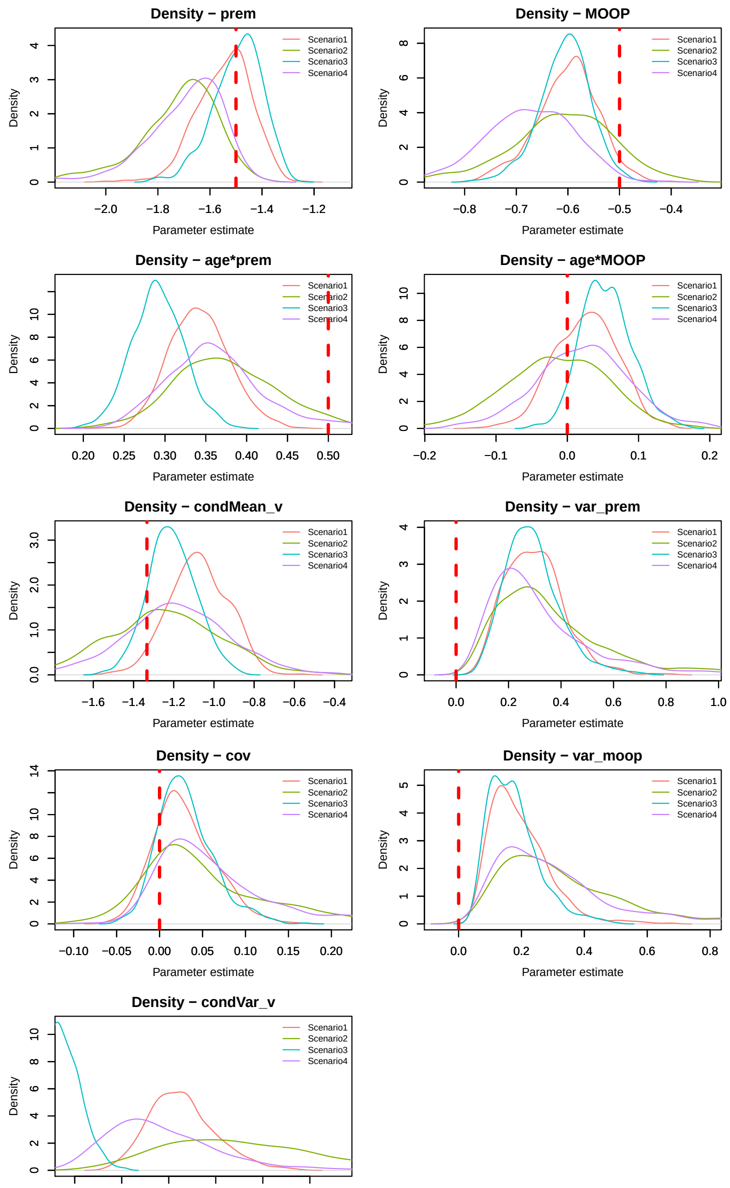
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $R = 50$.

Figure D.10: Posterior Distributions of Firm Fixed Effects, Low Unobserved Heterogeneity, $\delta \neq 0$, $R = 50$



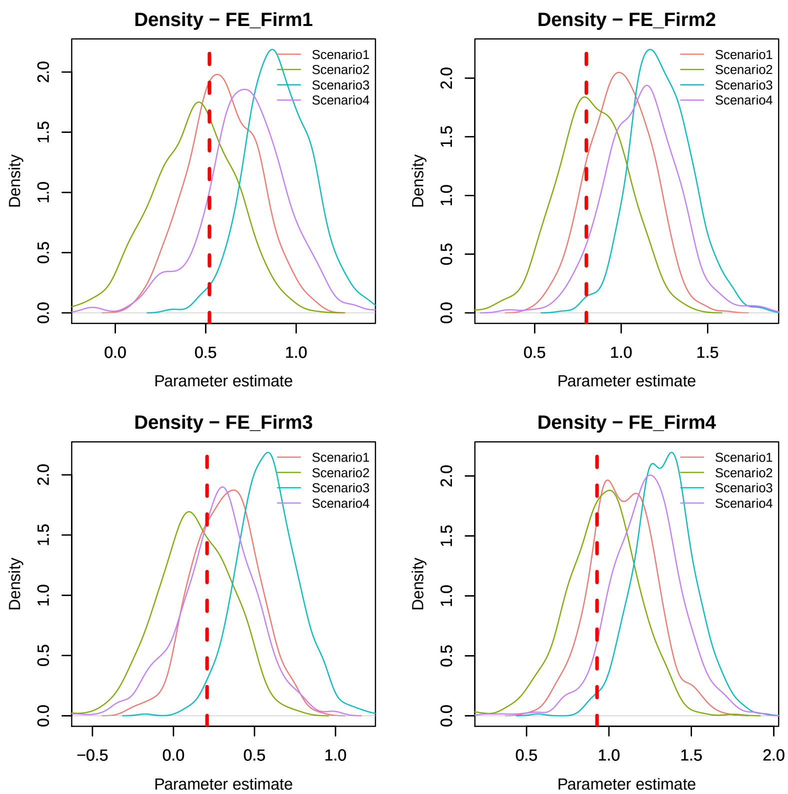
Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $R = 50$.

Figure D.11: Posterior Distributions, Low Unobserved Heterogeneity, $\delta \neq 0$, $R = 100$



Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $R = 100$.

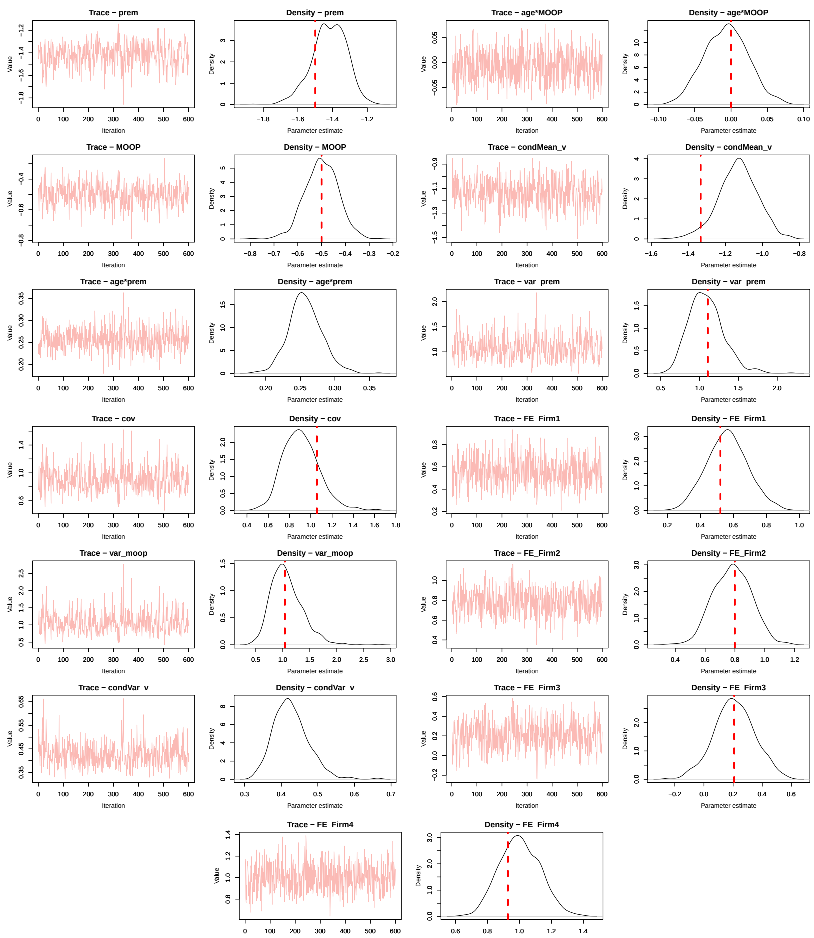
Figure D.12: Posterior Distributions of Firm Fixed Effects, Low Unobserved Heterogeneity, $\delta \neq 0$, $R = 100$



Note: Scenario three replicates the data augmentation algorithm as in Musalem et al. (2009) using product-level market shares; scenario four shows the posterior distributions when the dependent variable is aggregated by multi-product firms. Case is for high unobserved heterogeneity of preference weights. Number of simulated individuals per market $R = 100$.

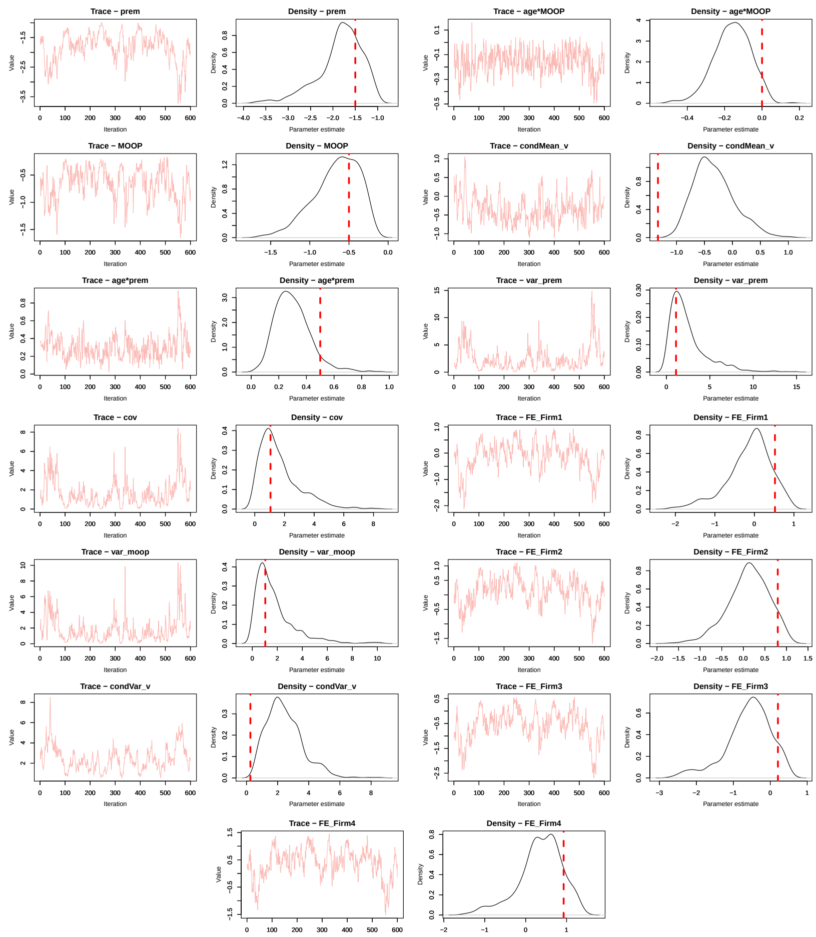
D.2 Diagnoses of Convergence: Trace Plots

Figure D.13: Trace Plot, High Unobserved Heterogeneity, Scenario 1, $\delta \neq 0$, $R = 200$



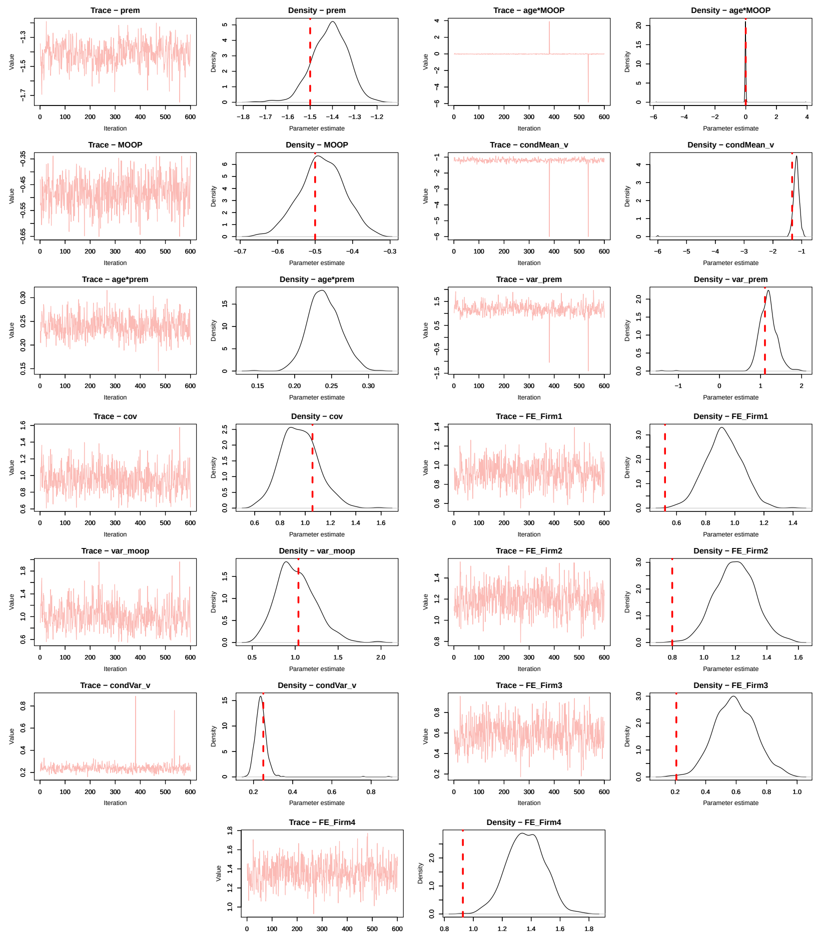
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.14: Trace Plot, High Unobserved Heterogeneity, Scenario 2, $\delta \neq 0$, $R = 200$



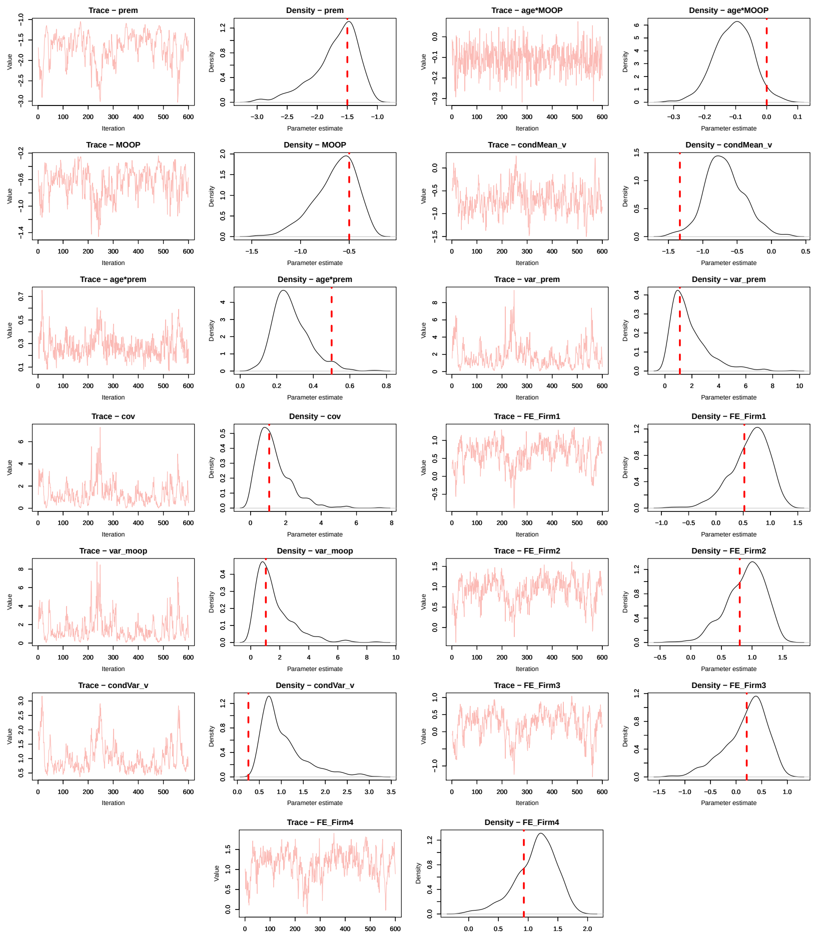
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.15: Trace Plot, High Unobserved Heterogeneity, Scenario 3, $\delta \neq 0$, $R = 200$



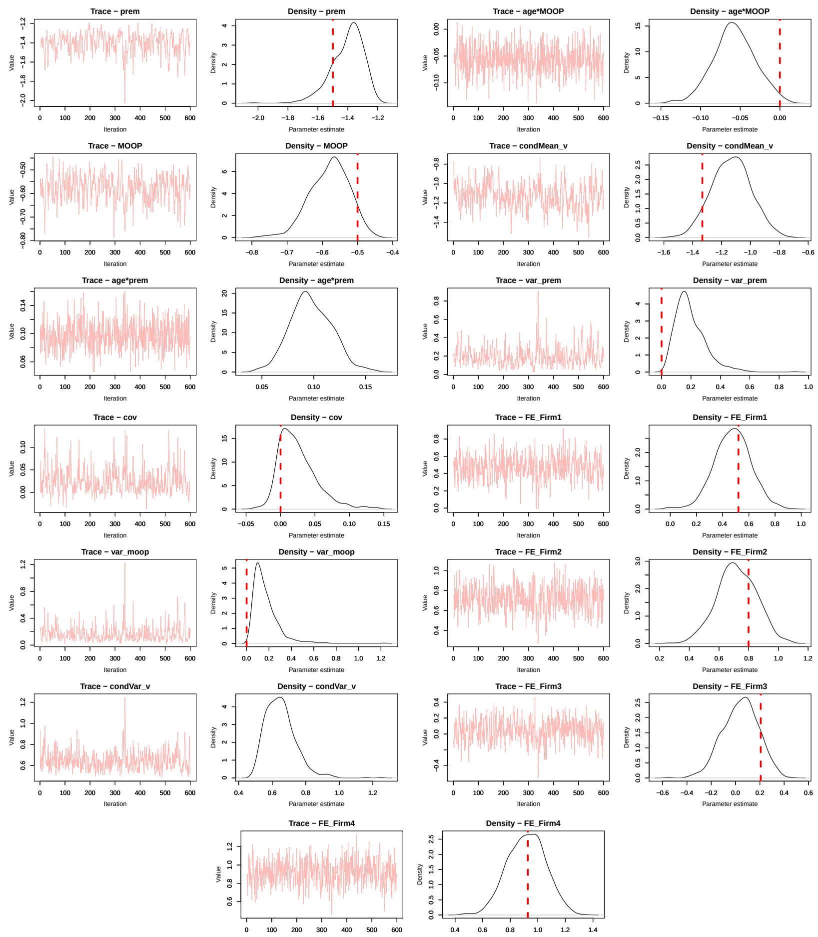
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.16: Trace Plot, High Unobserved Heterogeneity, Scenario 4, $\delta \neq 0$, $R = 200$



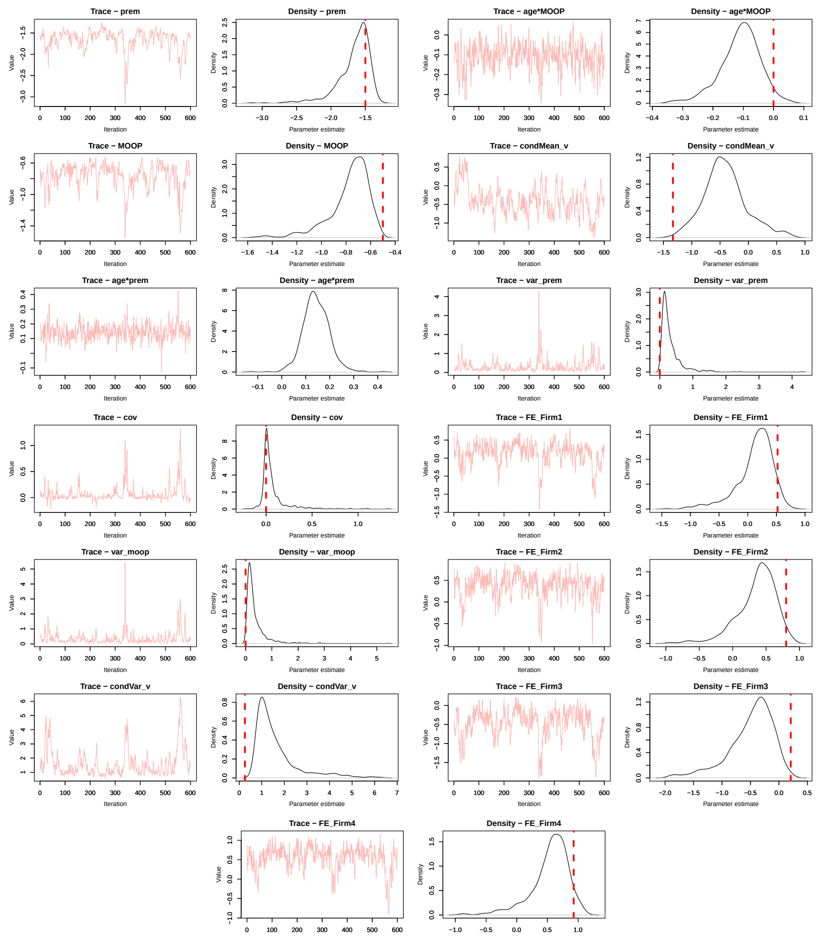
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.17: Trace Plot, Low Unobserved Heterogeneity, Scenario 1, $\delta \neq 0$, $R = 200$



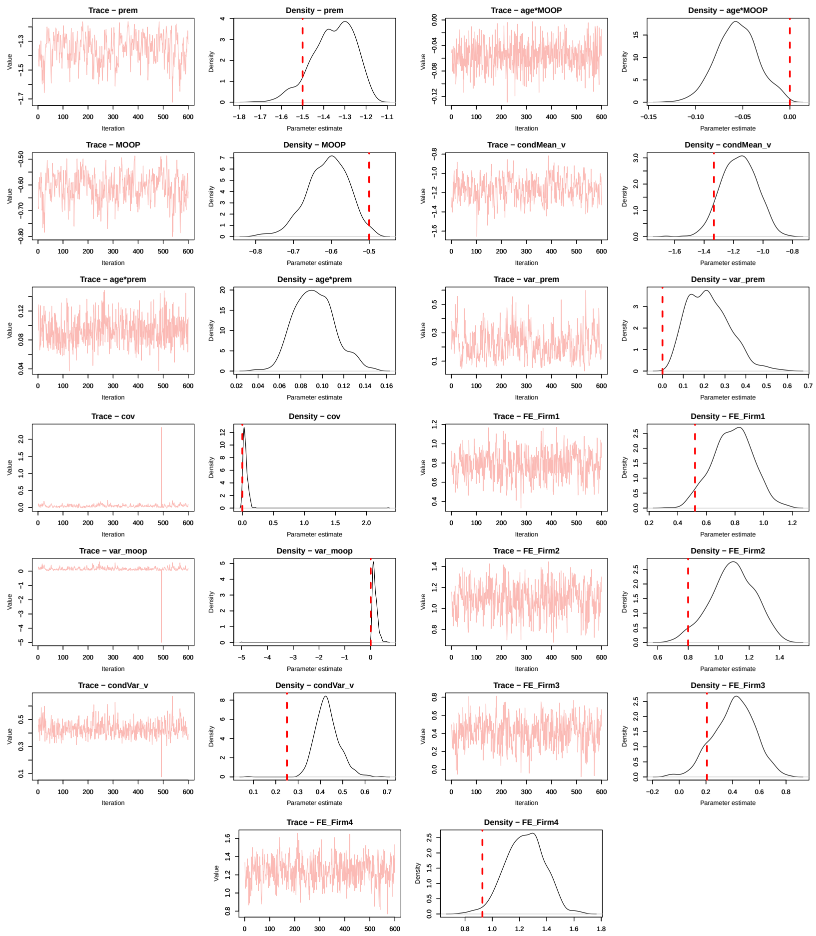
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.18: Trace Plot, Low Unobserved Heterogeneity, Scenario 2, $\delta \neq 0$, $R = 200$



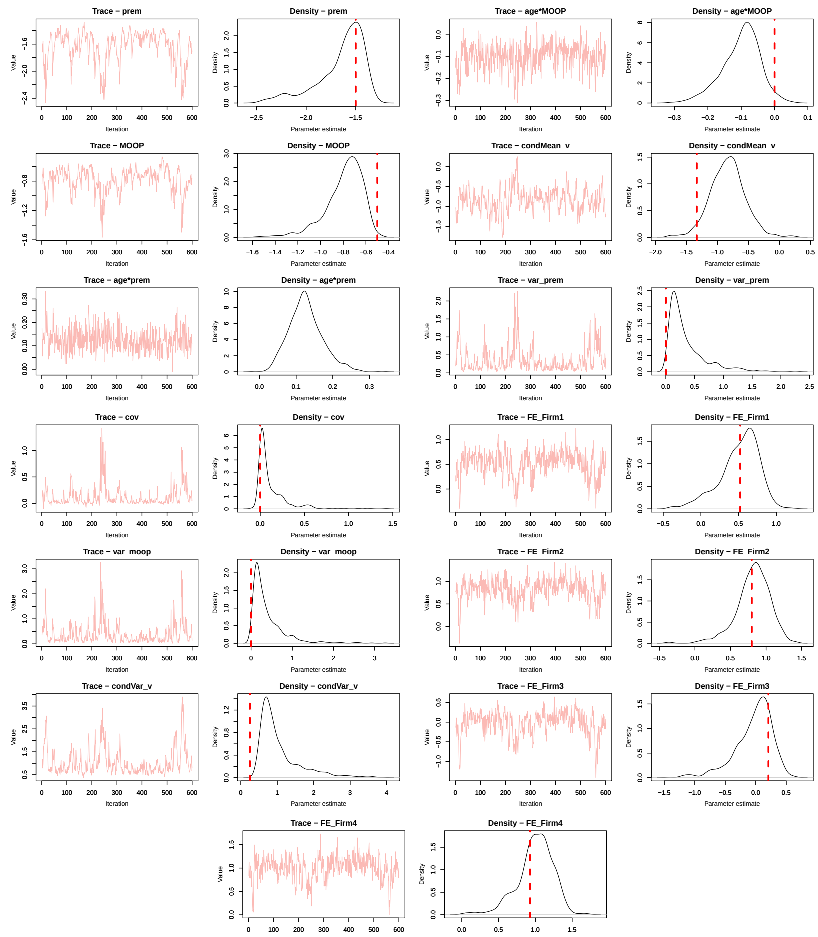
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.19: Trace Plot, Low Unobserved Heterogeneity, Scenario 3, $\delta \neq 0$, $R = 200$



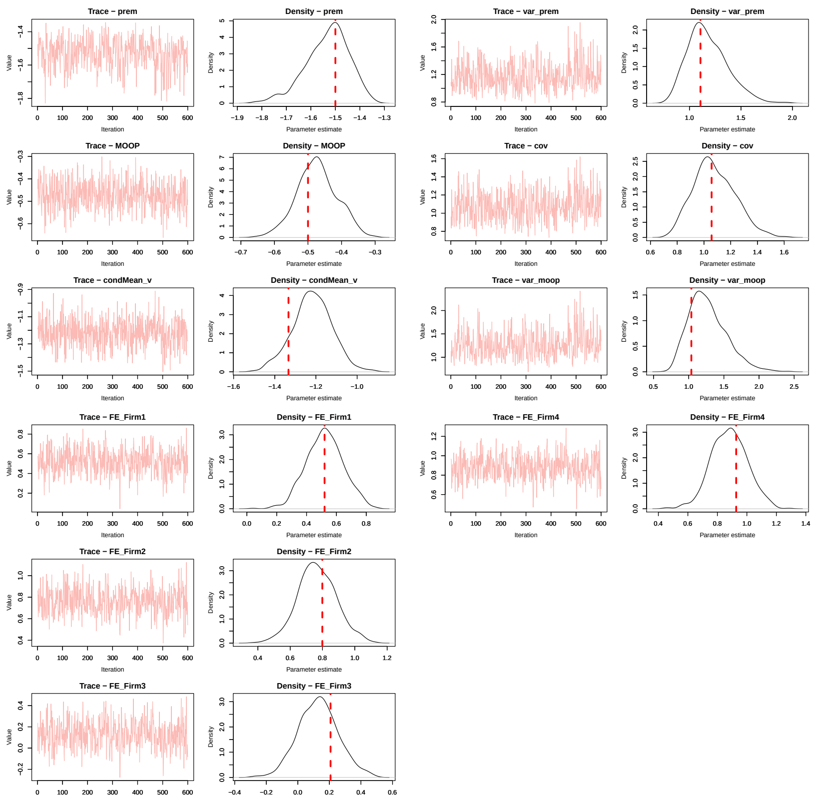
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.20: Trace Plot, Low Unobserved Heterogeneity, Scenario 4, $\delta \neq 0$, $R = 200$



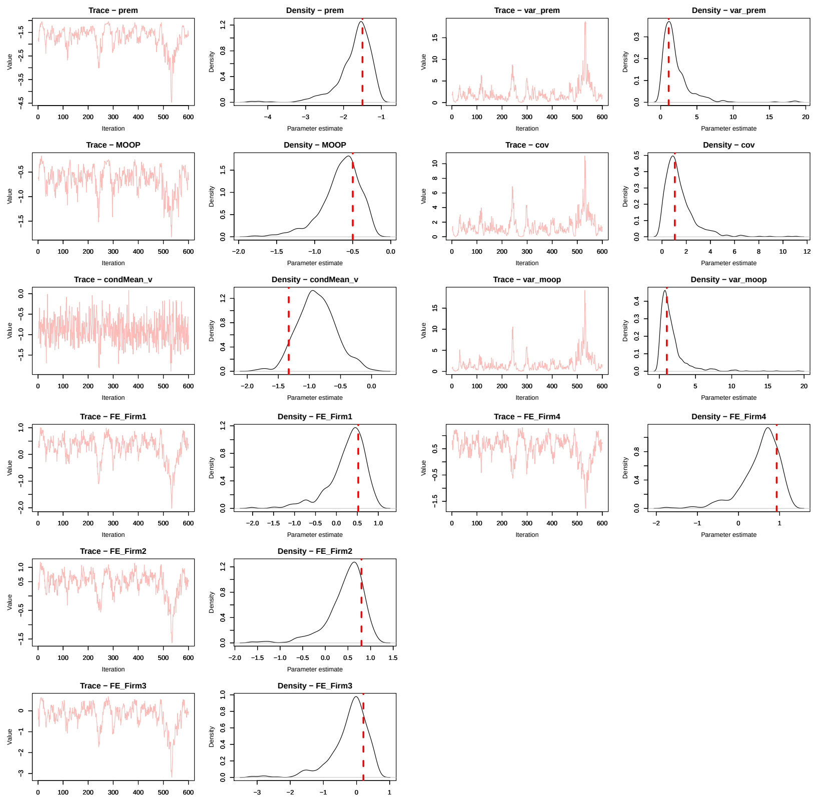
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.21: Trace Plot, High Unobserved Heterogeneity, Scenario 1, $\delta = 0$, $R = 200$



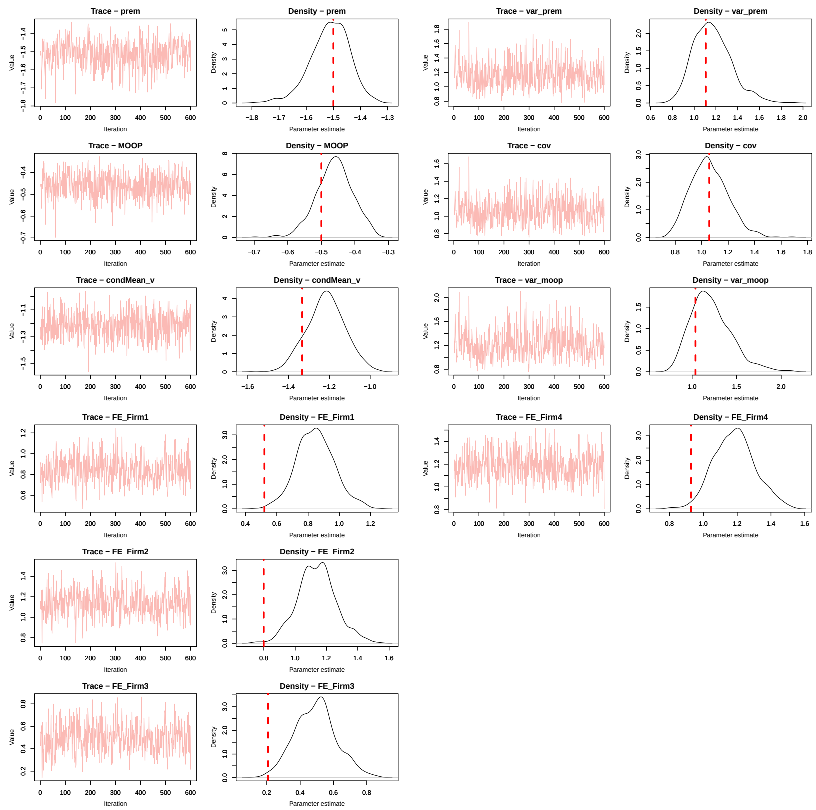
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.22: Trace Plot, High Unobserved Heterogeneity, Scenario 2, $\delta = 0$, $R = 200$



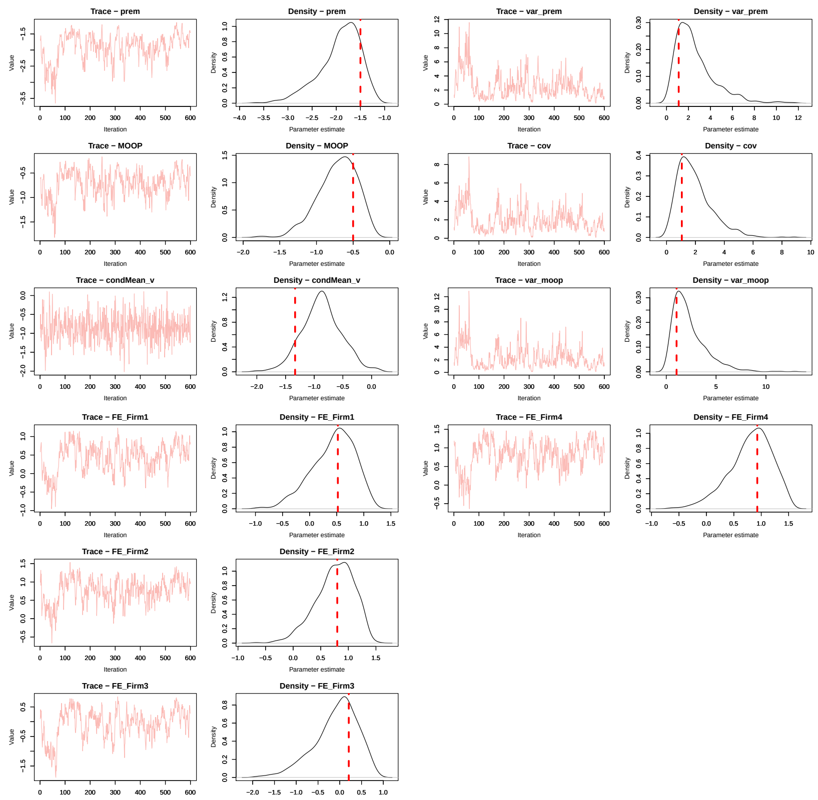
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.23: Trace Plot, High Unobserved Heterogeneity, Scenario 3, $\delta = 0$, $R = 200$



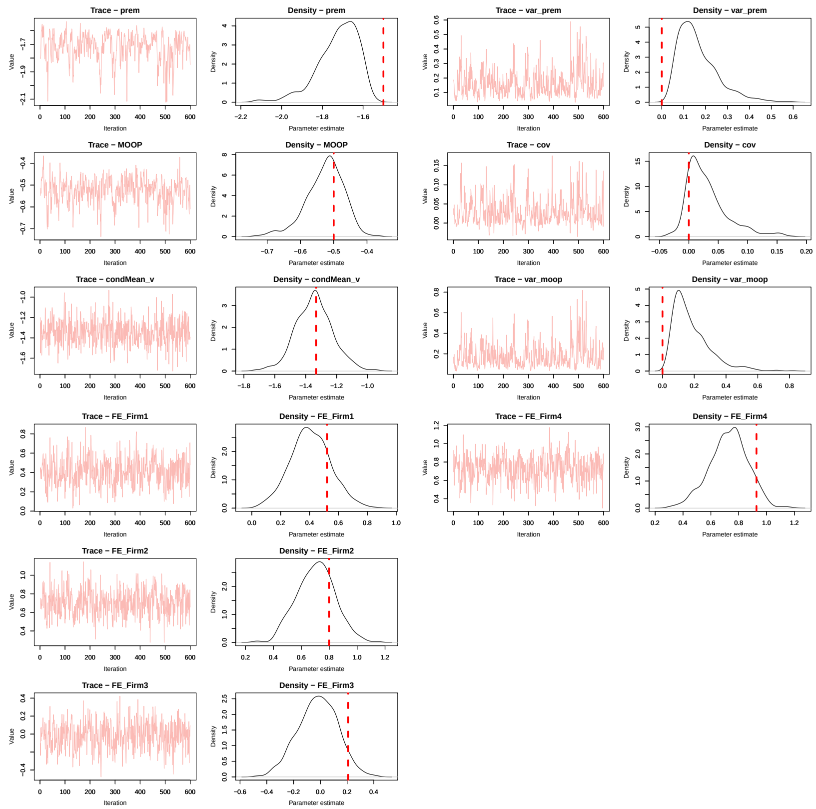
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.24: Trace Plot, High Unobserved Heterogeneity, Scenario 4, $\delta = 0$, $R = 200$



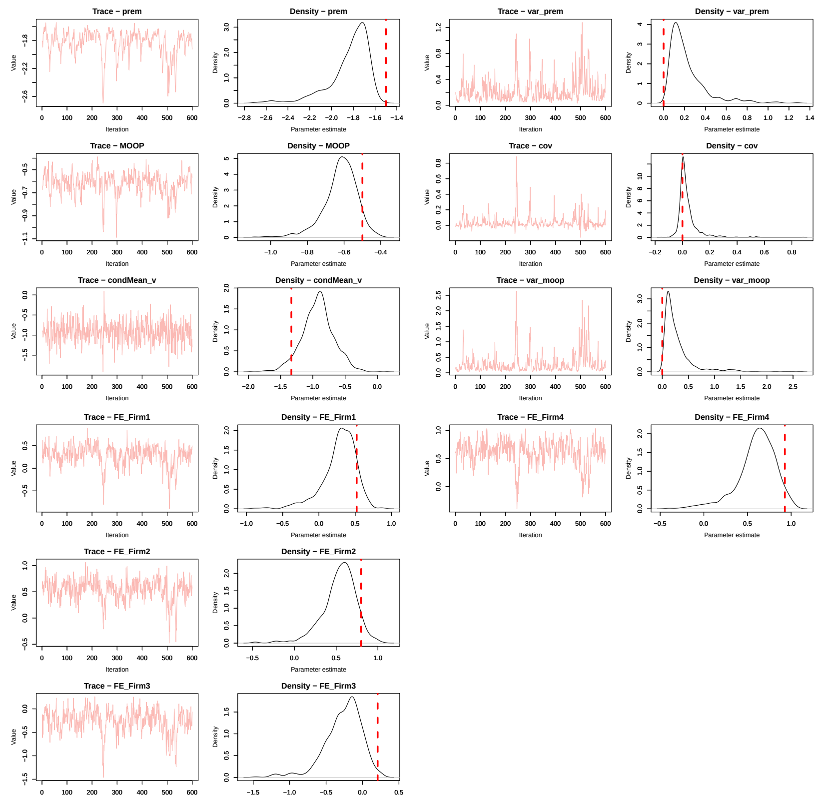
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.25: Trace Plot, Low Unobserved Heterogeneity, Scenario 1, $\delta = 0$, $R = 200$



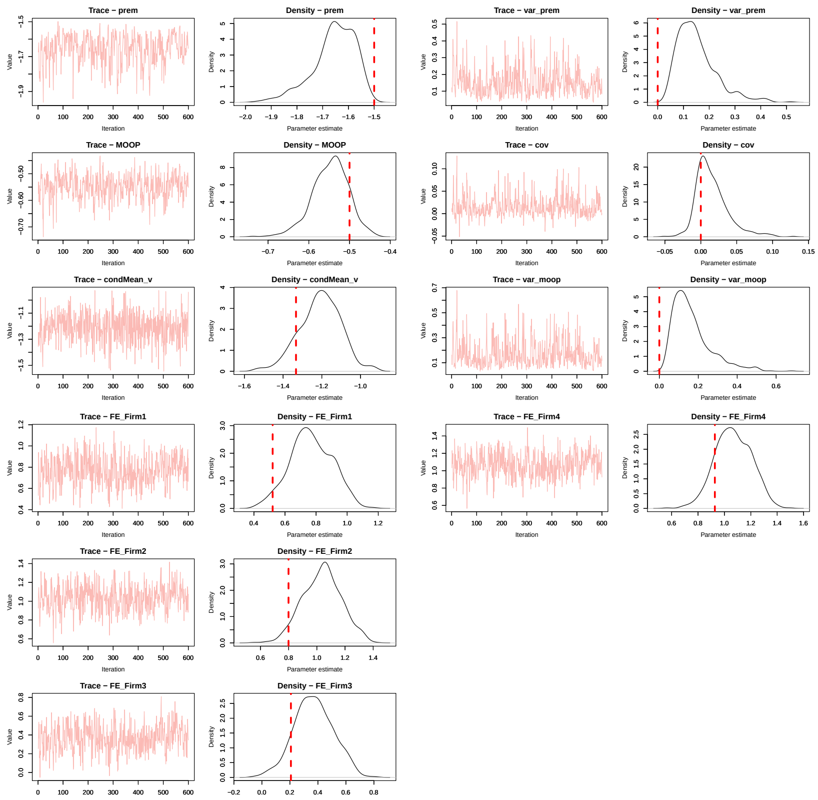
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.26: Trace Plot, Low Unobserved Heterogeneity, Scenario 2, $\delta = 0$, $R = 200$



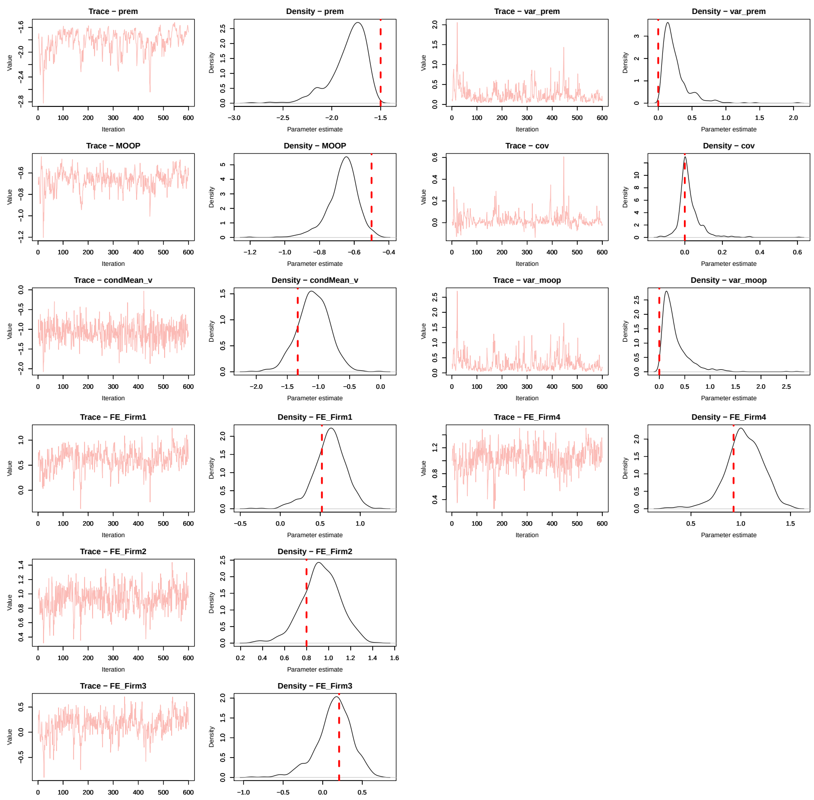
Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.27: Trace Plot, Low Unobserved Heterogeneity, Scenario 3, $\delta = 0$, $R = 200$



Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.

Figure D.28: Trace Plot, Low Unobserved Heterogeneity, Scenario 4, $\delta = 0$, $R = 200$



Note: Trace plots and density plots are drawn again thinned posterior draws. In density plots, true parameters are shown in red vertical lines.