December 2022

Essay in Bayesian Econometrics

Loren P. Wagner
University of Wisconsin-Milwaukee

ESSAYS IN BAYESIAN ECONOMETRICS

by

Loren Wagner

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Economics

at

The University of Wisconsin-Milwaukee

December 2022

ABSTRACT

ESSAYS IN BAYESIAN ECONOMETRICS

by

Loren Wagner

The University of Wisconsin-Milwaukee, 2022
Under the Supervision of Professor Scott Adams

Bayesian Econometrics is not as popular as the more common frequentists methods. However, there are problems that Bayesian methods can handle that are not amenable to frequentists methods. The essays to follow show two such applications where Bayesian methods offer solutions to problems posed in widely different areas of economics. In the first essay, the predictive posterior distribution is used to calculate a number of economic quantities that often require extensive work arounds using frequentist methods. In the second essay, it is shown that Bayesian estimation techniques can handle situations wherein an independent variable exhibits no variability, a situation that would lead frequentists methods to fail.

In the first essay, A new model is proposed for estimating demand systems based on the Dirichlet and Beta Regression work that has been developed. This model can overcome some of the criticisms aimed at other models used to estimate demand system, in particular the heterogeneity of preferences and the relationship of those preferences to other predictors. By modeling the preferences explicitly, the model can be given a causal interpretation of consumers decisions. Finally, it will be shown that how to aggregate the model up to the market demands.

In the second essay, Nurse lead primary care is a relatively rare model for primary care, but given nurses more holistic approach to medicine can prove effective in both reducing costs and disparities in access and outcomes. To compare the cost adjusted quality of care between the traditional physician and a nurse lead primary care model the outcomes from a population of patients that visited each are compared with respect to three complications of hypertension (stroke, heart attack, coronary heart disease). It is found that the population

that visited the nurse lead primary care facility had better cost adjusted outcomes than the population that visited the physician lead facility.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Thanks go to Drs. Scott Adams and Scott Drewianka for tolerating and guiding my eccentric explorations in a number of areas that culminated in this work. Dr. Bev Zabler, who sadly passed away in January of 2022 before this thesis could be completed. Without her direction and support, it is likely this work would never have been completed.

In addition, my mother, who always thought I was going to change the world, not quite mother, never gave up on this project of mine, despite the years stretching far longer than I had hoped. And my late father, who in the acknowledgments of his own dissertation wrote "I hope [my two boys] have also learned something about the value of working toward achieving long-term goals, and they will not categorically rule our the option of graduate education for themselves." I am not sure I have satisfied the first of those, but I can say dad "I did not categorically rule out graduate education". Lord knows, though, that sometimes during this whole process, I wish I had.

Chapter 1

Demand Estimation: A Dirichlet/Beta Regression Approach

1.1 Introduction

Browning and Carro, 2007 convinced econometricians in the demand estimation literature that heterogeneity in consumer preferences had to be taken seriously in order to accurately and precisely estimate models of consumer demand, whether in systems or individual goods. Since then, much work has been done constructing nonparametric models to satisfy this need (S. T. Berry and Haile, 2014, Blundell et al., 2017). Relatively less work has been done finding parametric models to capture this heterogeneity, for understandable reasons. As Browning and Carro, 2007 argue the kind of general heterogeneity that exists everywhere seems to recommend models that capture heterogeneity in a non-separable way. In the discrete choice literature, the multinomial logit model of McFadden, 1974 and the developments it inspired have found success in dealing with general heterogeneity, even if some still think there is work to be done capturing this general heterogeneity, Allenby et al., 1998.

The model developed in this paper follows in the spirit of the developments of the discrete choice literature, but applies to the wider context of continuous demand situations

and demand systems. To do this, an older model (Woodland, 1979, J. M. Fry et al., 2000, J. M. Fry et al., 1996 and T. Fry, 2011) is dusted off and given new life. The model, inspired by the Generalized Linear Model type procedure used in the discrete choice literature, captures general heterogeneity in budget shares, from which demand curves can be derived. It makes advances on the work of Woodland, 1979, J. M. Fry et al., 2000, J. M. Fry et al., 1996 and T. Fry, 2011 by connecting the Dirichlet/Beta distributions to utility maximization explicitly. It also has some advantageous features, including being explicitly consistent with adding up conditions, parameters have economically meaningful interpretations, unconditional and conditional distributions of demand can be estimated and derived relatively simply with a paucity of parameters, and a means of deriving market demand in a way consistent with microfoundations, without requiring aggregate demand exhibit the same properties and functional forms as individual demands.

The rest of this paper will proceed as follows. The next section will develop the foundations and econometric details of the model. The following section will explain the advantages of the model and respond to possible objections in light of the details presented. Two empirical exercise will follow in the penultimate sections. The first empirical exercise will use data from the Consumer Expenditure Survey to compare the results of the beta regression model to the standard AIDS and QUAIDS models (Deaton and Muellbauer, 1980a, Banks et al., 1997). The second will use data from the National Household Transport survey to compare the beta regression to a nonparametric model developed by Blundell et al., 2017. Finally, some concluding remarks will be made.

1.2 The Foundations and Econometric Specification of the Model

The model starts from a familiar starting point, the individual consumers maximization problem where preferences take the simple Cobb-Douglas form[1].

$$\begin{aligned} &\max_{\mathbf{c}} \Pi_i^d c_i^{\alpha_i} \\ &subject\ to \\ &\langle \mathbf{p}, \mathbf{c} \rangle \leq m. \end{aligned} \tag{1.1}$$

Where $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_i x_i y_i$ is the Euclidean inner product. d is the number of goods or composite goods the consumer can choose from; this value plays little role in the discussion to follow, so will be omitted unless expressly needed. Normally, this would be an empirically bad specification to work with for many reasons: unity or zero elasticities, constant Engel curves, etc.[2] However, Many of these disadvantages can be removed with the following assumptions:

$$\begin{aligned} \sum_i \alpha_i = \langle \boldsymbol{\alpha}, \mathbf{1} \rangle = 1 \\ \boldsymbol{\alpha} \sim Dirichlet(\mathbf{h}) \\ \mathbf{h} = \mathbf{h}(\mathbf{X}) \end{aligned} \tag{1.2}$$

where $\mathbf{1}$ is a vector of d ones and $\mathbf{X}$ is a vector of predictors that can include anything that the researcher thinks has an effect on consumers preferences. $Dirichlet(\mathbf{h})$ is the Dirichlet distribution defined on the unit simplex, $S^{d-1} = \{x : \sum_i^d x_i = 1,\ x_i \geq 0\ for\ all\ i\}$ with density

$$f(x) = \frac{\Pi_i x_i^{h_i}}{Beta(\mathbf{h})} \mathbf{1}(x \in S^{d-1}) \tag{1.3}$$

[1] In what follows, I follow the convention of putting vectors in bold type face and scalars in normal type face.

[2] see the math appendix for comparisons of these

Where the denominator is the d-dimensional beta function defined as: $Beta(\mathbf{h}) = \frac{\Pi_i \Gamma(h_i)}{\Gamma(\sum_i h_i)}$ and $\Gamma(h) = \int_0^\infty t^{h-1} e^{-t} dt$ is the gamma function. $\mathbf{1}(x \in S^{d-1})$ is the indicator function of the unit simplex.

The first condition of (1.2) is standard fare, capturing explicitly the adding up condition of demand systems. The second condition allows for general heterogeneity to be modeled explicitly by allowing every consumer to have a different vector of preference parameters while still maintaining some constraints on those preferences. This is intuitively plausible. Heterogeneity of preferences is easily experienced (offer passers-by free candy, but let them have many choices. It is expected that not everyone will choose the same candy). Yet, these differences in our preferences are constrained by human evolved biology and psychology and the various cultures that we share.

It is the last condition of (1.2), though, which greatly expands the power of the Cobb-Douglas utility and allows for the simple estimation procedure that will be developed in the following. The interpretation of the combined distributional assumptions on the preference parameters is that each individual/household has a different set of preference parameters but the distribution of those preferences is systematically affected by a set of predictors, $\mathbf{X}$. When total expenditure/income/wealth are part of $\mathbf{X}$ this entails that there are systematic differences in the distribution of preferences across the expenditure/income/wealth distribution; in other words, Engel curves will not necessarily be linear. This Dirichlet distributional assumption requires the Cobb-Douglas preference specification. One can make similar distributional assumptions using other preference specifications, but the parameters of many other common utility functions do not exist in as nice a space as the unit simplex. The simplex is nice in that it is a compact space. Thus, there exists a minimally informative distribution, in the sense of Shannon information, that does not have moment conditions, namely the uniform distribution, $Dirichlet(\mathbf{1})$. In a Bayesian setting, this allows for an easy choice of uninformative priors.

Under this specification of the consumers preferences, the demand functions take the

form:

$$c_i = \frac{m}{p_i}\alpha_i \tag{1.4}$$

or in their budget share form:

$$s_i = \frac{c_i p_i}{m} = \alpha_i \tag{1.5}$$

Even though, a consumer's preference parameters are random, if it is assumed a consumer's order in the distribution is invariant to changes, we can recover a single consumer's demand and utility function by looking at quantiles of the preference distribution. In subsequent sections we shall focus on $\mathbf{X} = (\mathbf{p}, m)$, where $\mathbf{p}$ is a vector of prices and m is the consumers expenditure/income/wealth, but I shall just call it expenditure from here on. The predictor vector can contain any other observable consumer trait that a researcher thinks may systematically change the distribution of preferences. For example, if a researcher believes that the consumption patterns of consumers/households that have two adults will be different from those with one adult, a term in $\mathbf{X}$ can be included to represent number of adults in the household.

With the assumption that $\boldsymbol{\alpha} \sim Dirichlet(\mathbf{h})$ and $\mathbf{h} = \mathbf{h}(\mathbf{X})$, the conditional distribution of the consumption vectors can be derived. Conditional on the predictors $\mathbf{X}$, the distribution of the consumption vectors will follow what I will call the scaled Dirichlet distribution on a weighted simplex,

$$\partial B = \{\mathbf{c} : \forall i c_i \geq 0, \langle \mathbf{p}, \mathbf{c} \rangle = m\} \tag{1.6}$$

The conditional density will be

$$f_{\mathbf{c}|\mathbf{X}} = \frac{\Gamma(h_+)}{\Pi_i^d \Gamma(h_i)} \frac{\Pi_i^d p_i^{h_i} c_i^{h_i - 1}}{m^{h_+}} \mathbf{1}(c \in \partial B). \tag{1.7}$$

where $h_+ = \sum_i h_i$. The marginal conditional distributions follow a four parameter Beta with

densities[3]:

$$f_{c_i|\mathbf{x}} = \frac{1}{B(h_i, h_+ - h_i)} c_i^{h_i - 1} \left(\frac{m}{p_i} - c_i \right)^{h_+ - h_i - 1} \left(\frac{m}{p_i} \right)^{-(h_+ - 1)} \mathbf{1}(c \in \partial B) \tag{1.8}$$

We can also get the conditional marginal distribution of consumption expenditure:

$$f_{p_i c_i|\mathbf{x}} = \frac{1}{B(h_i, h_+ - h_i)} (p_i c_i)^{h_i - 1} (m - p_i c_i)^{h_+ - h_i - 1} m^{-(h_+ - 1)} \mathbf{1}(p_i c_i \leq m) \tag{1.9}$$

Everything that follows can be worked out in either the demand system case or single good case. For the demand system, the Dirichlet distribution will be used. For a single good or composite good, the beta marginals will be used. Given the nice properties of the Dirichlet, the derivations going from one to the other are fairly simple. In what follows, for simplicity of exposition, I shall focus on the beta regression.

To estimate the model, we have to put a bit more structure to the $\mathbf{h}(\mathbf{X})$ functions. To fit into a Generalized Linear Model (GLM)[4] type estimation scheme, the Beta/Dirichlet will be reparametrized in terms of the mean vector, $\boldsymbol{\mu}$, and a precision parameter, ϕ. On the standard parametrization of the Dirichlet distribution, $X \sim Dirichlet(\boldsymbol{\alpha})$, the expected value of individual components are

$$\mathbf{E}[X_i] = \frac{\alpha_i}{\alpha_+} \tag{1.10}$$

Let $\mu_i = \mathbf{E}[X_i]$ and $\alpha_+ = \sum_i \alpha_i = \phi$. The density under this parameter set is

$$f(x) = \frac{\Pi_i x_i^{\phi \mu_i - 1}}{B(\phi \boldsymbol{\mu})} \tag{1.11}$$

Then, following the GLM procedure, We model the expected value of the components via a link function taking a linear function of the predictors of interest as input. After several

[3]See the mathematical appendix for derivations of these claims

[4]But, strictly not a GLM model in the guise of McCullagh and Nelder, 1989 on account of having more than one parameter and not having a canonical link function

trials, the log link function offered the best fit to the empirical distributions:

$$log(\mu_i) = \langle \boldsymbol{\beta}, \mathbf{X}_i \rangle \tag{1.12}$$

Where $\mathbf{X}_i$ is the vector of predictors that are thought to influence the expected value of the components and $\boldsymbol{\beta}$ are their effect sizes. This method allows for each component's mean to be modeled by a different, possibly overlapping, set of predictors. However, there is nothing that demands use of this link function. One could use any function for the demand shares, even the AIDS or QUAIDS functions that have been developed previously.

Using the log-link function, the specification of $\mathbf{h}(\mathbf{X})$ from the previous sections is the exponential function,

$$h_i(\mathbf{X}) = \phi\mu_i = \phi e^{\langle \boldsymbol{\beta}, \mathbf{X}_i \rangle} \tag{1.13}$$

$$h_+(\mathbf{X}) = \sum_i h_i(\mathbf{X}) = \phi \tag{1.14}$$

Estimation of the resulting model will be either by Maximum likelihood or Bayesian posterior simulation. The likelihood function, assuming a random sample of expenditure share data and predictors will be

$$L = \Pi_i^N \frac{1}{B(\phi\boldsymbol{\mu_i})} \Pi_j^d s_{ij}^{\phi\mu_{ij}-1} \tag{1.15}$$

The corresponding log-likelihood being,

$$logL = \sum_i^N -log\left(B(\phi\boldsymbol{\mu_i})\right) + \sum_i^N \sum_j^d (\phi\mu_{ij} - 1) log(s_{ij}) \tag{1.16}$$

where s_{ij} are observed budget shares; these observed budget shares are not required to be total budget shares. Because of the subcompositional invariance of the Dirichlet distribution, it does not matter if we take budget shares relative to total expenditure or budget shares relative to some subset of total expenditure. Maier, 2014 reports that this log-likelihood

function is concave[5]. So, the optimization in the ML procedure is particularly well-behaved, if computationally cumbersome. The Bayesian procedure will also exhibit nice properties in simulating from the posterior.

In this parametrization, we face a similar problem to the standard system of demand estimation techniques. As written above, the model is over-specified since

$$\sum_j \mu_j = 1 \tag{1.17}$$

To handle this, we will have to effectively leave one μ_i out, by leaving its β parameters unestimated, $\beta_i = \mathbf{0}$. It has been shown that the estimation is invariant to which component's parameters are unestimated (Maier, 2014), much like estimates in the standard literature are invariant to which of the equations are left out.

For the empirical exercise to follow, A Bayesian estimation procedure will be used via Stan's Hamiltonian Monte Carlo posterior sampling algorithm. The Bayesian model requires specification of priors for all estimated parameters. Minimally informative priors will be used for all parameters. Thus,

$$\boldsymbol{\beta} \sim MVNormal(\mathbf{0}, \mathbf{I}\sigma) \tag{1.18}$$

An initial run of the HMC algorithm set $\sigma = 1000$, entailing a very diffuse prior for all parameters. Convergence results for this specification were very poor. This was due to the sampling procedure. If the initial values, which were first assigned randomly, happened to be far away from the interval $[-1, 1]$, wherein most of the posterior distributions mass was located, the sampling took a long time (in terms of sample draws from the posterior) to get there. Thus, the means were highly influenced by these initial draws and not reliable. However, if the warm-up phase was extended or initial values were set inside the interval, the convergence problems disappeared and the posterior sample had no draws outside the interval. So, the results below will reflect a specification with $\sigma = 1$ for all parameters.

[5]See the appendix for a discussion of this point

Marginal posteriors given this prior specification corroborate central limit theorem expectations, having approximately normal appearance.

For ϕ, the minimally informative prior is an exponential distribution. We default to an exponential with mean 1, with sensitivity analyses corroborating that the mean of this prior does not significantly effect any of the parameter estimates.

1.3 Advantages of the Model

1.3.1 Handling Heterogeneity

Browning and Carro, 2007 claims that it is hard to allow for heterogeneity in a general way. They further claim that most schemes are "chosen more for their statistical convenience than for their fit to the data or congruence with economic theory." It is unlikely that any econometric model will be able to completely dispense with this need to make simplifying assumptions for the sake of computational tractability, the Dirichlet/Beta regression will be no different[6]. However, the Dirichlet/Beta regression attempts to connect some of these tractability assumptions to plausible regularities in the data-generating processes as well as economic theory. As is discussed in the empirical exercises, these assumptions do not come at the cost of empirical fit. This section will discuss how the model handles the general heterogeneity that Browning and Carro, 2007 claim is everywhere.

Browning and Carro, 2007 object to how demand estimation has handled heterogeneity in past work because it has either not been as general as we know heterogeneity to be or by handing it with an additive term. The simplest work in the AIDS family saw heterogeneity handled by a demographic correction term (Deaton and Muellbauer, 1980a)

$$s_i = \alpha_i + \sum \gamma_{ij} log p_j + \beta_i log(\frac{m}{kP}) \tag{1.19}$$

[6]For example, leaving one set of parameters un-estimated; assuming that a consumers order in the distribution is invariant to changes in their $\mathbf{X}$ vector; others will be seen in succeeding sections

Where k is a measure of household size and P is a price index. This is both additive and not general; the heterogeneity only exists at the level of household size. With the addition of a normally distributed error term, this model exhibits all the problems Browning and Carro, 2007 point-out with regard to how heterogeneity is handled.

Banks et al., 1997 expanded the AIDS model to handle a quadratic income/total expenditure term and offered more generality to the heterogeneity:

$$s_i = \alpha_i + \sum \gamma_{ij} log p_j + (\beta_i + \eta z) log\left(\frac{m}{m(z)a(p)}\right) + \frac{\lambda_i}{b(p)c(p,z)}\left[log\left(\frac{m}{m(z)a(p)}\right)\right]^2 \quad (1.20)$$

Where $m(z)$ is a linear function of demographic predictors, z, and $c(p,z)$ adjust the price index $b(p)$ to the demographic groups. This treatment of heterogeneity appears not to be additive, though, aside from the price index, $b(p)c(p,z)$, this appearance is a tad deceptive. It still assumes that heterogeneity only exists at the demographic group level and within group heterogeneity is handled by an additive error term.

Blundell et al., 2017 developed a nonparametric quantile regression[7] for demand:

$$s_i = G_\alpha(p,m) + V_\alpha, \;\; P(V_\alpha \leq 0) = \alpha \quad (1.21)$$

This model seems to satisfy the demand of Browning and Carro, 2007. Each consumer is assumed to be captured by a certain quantile in the resulting distribution of budget shares for each price and total expenditure pair. However, this model is constrained to require more an more parameters in order to estimate demand across the distributions. Thus, its applicability to answering wider questions is limited.

The model developed in the previous section handles heterogeneity in a general non-additive way. By assuming that preferences for an individual are random draws from a distribution, each individual (household or individual) will have a different set of prefer-

[7]The details will be discussed in more depth in the empirical section

ence parameters. This assumption also captures the fact that human preferences cannot be drastically different, considering the shared biology and psychology. It is hard to imagine a more general way of handling heterogeneity. In addition, the Cobb-Douglas utility specification entails that the heterogeneity is not just added on as an error term. Thus, the model developed satisfies the demands of Browning and Carro, 2007.

The two assumptions (Dirichlet distribution and Cobb-Douglas preferences) this model rests on may seem overly restrictive. However, this is only appearances. The Dirichlet distribution on the Cobb-Douglas is a natural one. It is by no means the only one. In fact, using the isometric logratio (ilr)[8] transform, one can transform any distribution on $\mathbf{R}^{d+1}$ to a distribution on $\mathbf{S}^d$ (the unit simplex in $\mathbf{R}^{d+1}$). Thus, future work might explore if other distributional assumptions on the Cobb-Douglas preferences parameters offer better fits to the data. The Dirichlet, and its Beta marginals, do recommend themselves for statistical convenience.

The Cobb-Douglas preference specification may strike some readers as quite odd. In its pure form it is a very bad fit to almost any data set. As summarized in the mathematical appendix, elasticities are necessarily unity and zero for the pure Cobb-Douglas. However, if we are willing to dispense with the dubious psychological assumptions that our preferences are not affected by prices or incomes or wealth, then every demand curve can be derived from a Cobb-Douglas preference specification.

[8]The isometric logratio transform is an isomorphism and isometry between the simplex, S^d, and the Euclidean real space, $\mathbf{R}^{d+1}$. If $\mathbf{e}$ is an orthonormal basis for S^d, then the ilr transform of a vector $x \in S^d$ takes the Aitchinson inner product of x with the elements of the orthonormal basis, $\mathbf{e}$,

$$ilr(x) = \Big(\langle x, e_1\rangle, ..., \langle x, e_{d+1}\rangle\Big) \tag{1.22}$$

The Aitchinson inner product is part of the structure that gives the simplex the form of a real vector space, a Hilbert space. The Aitchinson inner product is defined as

$$\langle x, y\rangle = \frac{1}{2d}\sum_i \sum_j log\frac{x_i}{x_j} log\frac{y_i}{y_j} \tag{1.23}$$

1.3.2 On Compositional Data Analysis

In any system of demand or single good demand, there is always an adding up condition that must be satisfied. This follows from constrained utility maximization. The traditional ways to handle this in econometric works is to impose it though linear constraints on any parameters estimated, or as a condition on non-parametric estimators. In parametric systems of demand, this is not a satisfactory solution, especially if any additive errors are treated as normal. Aitchison, 1986 and Pearson, 1897 show why this solution cannot be satisfying. Data that satisfies an adding up condition cannot be reliably treated with standard multivariate statistical methods.

This is not new to the demand system literature. Woodland, 1979, J. M. Fry et al., 2000, J. M. Fry et al., 1996 and T. Fry, 2011 explore demand systems in light of work done in compositional data analysis. Woodland, 1979 is to be preferred, and can be considered inspiration for the model developed in the previous section, an updated version, perhaps. J. M. Fry et al., 1996 suffers from not having a full grasp of the geometry of the simplex. It would be a few years after J. M. Fry et al., 1996 before an isomorphic transform from the simplex to the Euclidean plane was found. With such a transform (the ilr transform) the method of J. M. Fry et al., 1996 could be properly attempted. The additive log-ratio (alr) transform that J. M. Fry et al., 1996 uses is not an isomorphism between the simplex and the Euclidean plane. Thus, the alr transform need not preserve distances when transforming between the two vector spaces. This entails that an MLE procedure may identify a maximum in the Euclidean plane that is not a maximum when transformed back into the simplex[9].

There are a few advantages to working directly in the simplex. The first is that adding up conditions will be satisfied implicitly. There will be no need to impose extraneous linear conditions on the parameters of the model. The second is that we will not allow for

[9]It is common knowledge that maximums are invariant to monotonic transformations. Monotonic transformations are a special kind of isomorphism on the Real line. Isomorphisms are one-to-one transformations that preserve binary relations, for example the "$\leq$" relation. Thus, the maximum of a $log(G(l(\theta; x)))$ where G is not an isomorphic transform of the underlying density, f with loglikelihood l, cannot be assured to preserve the ordering of the parameter space, Θ. Thus, we cannot be sure that it preserves the maximum.

budget shares or expenditures that contradict budget constraints, which, despite the linear constraints on the parameters, was always possible because of the unrestricted error terms used. Finally, working directly in the simplex ensures that we are not working with methods that are demonstrably meaningless in the simplex (Aitchison, 1986).

Pearson, 1897 showed that adding up conditions play havoc on standard methods of multivariate data analysis. In particular, the covariance is not meaningful when applied to compositional data. Pearson showed, and it is easily repeated, that the covariance can change drastically when we divide by a third random variable, i.e.

$$Cov(X, Y) \neq Cov(\frac{X}{Z}, \frac{Y}{Z}) \tag{1.24}$$

In fact, it is easy to see the sign change when dividing by a third random variable, as when we divide consumption expenditure by total expenditure to get budget shares. Least squares methods will run afoul of this since they attempt to minimize variances. Maximum likelihood methods will suffer when they try to transform densities from $\mathbf{R}^{d+1}$ to $\mathbf{S}^d$, without properly accounting for the special geometry of the simplex.

The method developed in the previous section explicitly models the adding up condition in the distributional assumption on preferences. It is also respectful of the geometry of the simplex. So, there is little question concerning its reliability in this regard.

1.3.3 Market Demand

An issue that always motivates those developing models of demand is the aggregation problem. For a time, finding micro-models that can be aggregated without changing the form of the demand specification was a priority. It has been accepted that only under very unrealistic restrictions will the micro-demand curves and the market demand curves have the same form (Hartley, 1997). However, it would still be valuable to have a model that can be aggregated into something manageable. The Dirichlet/Beta regression model satisfies this

weaker condition.

Reminding ourselves, the Beta regression (for one good) entails that the conditional demand of budget shares follows a Beta distribution:

$$S|X_i \sim Beta(\mu_i, \phi) \tag{1.25}$$

$$\mu_i = e^{X_i\beta} \tag{1.26}$$

This entails that expenditure on this good follows a four parameter Beta distribution:

$$pc|X_i \sim Beta(\mu_i, \phi, 0, m_i) \tag{1.27}$$

Where m is total expenditure. If this model captures how demands of different consumers are interrelated, then the market can be simulated by N i.i.d. draws from this distribution. This entails that market demand, $p\mathbf{C}$, will follow an n-fold convolution of the above expenditure distribution:

$$p\mathbf{C} \sim *^n\Big(f_{pc|\mathbf{X}}(pc)f_X(\mathbf{X})\Big) \tag{1.28}$$

If the variances of the predictor distributions, $f_X(\mathbf{X})$, are all finite, then a central limit theorem applies. We can then calculate the expected value of the market demand distribution as

$$\mathbf{E}(p\mathbf{C}) = N\int_{pc}\int_{\mathbf{X}} pc f_{pc|\mathbf{X}} dF(\mathbf{X}) \tag{1.29}$$

The finite variances assumption, entails finite expectations, which then entails[10] we can switch the order of the integration. Using the fact the conditional expectation is known, and modeled, this can be simplified to

$$\mathbf{E}(p\mathbf{C}) = N\int_{\mathbf{X}} m\mu(\mathbf{X})dF(\mathbf{X}) = N\int_{\mathbf{X}} me^{\mathbf{X}\beta}dF(\mathbf{X}) \tag{1.30}$$

[10]Finite expectations entails the integrand is integrable, which allows the use of Fubini and Tonelli's theorems to switch the order of integration

The second equality follows from the econometric specification of the conditional expectation that was assumed in the previous section. Under this specification, the terms inside the integral is the derivative of the conditional expectation with respect to β_m. Under integrability conditions, the derivative and the integral can be re-ordered, resulting in:

$$\mathbf{E}(p\mathbf{C}) = N\frac{\partial}{\partial t_m}\mathbf{E}_{\mathbf{X}}(e^{\mathbf{X}t})|_{t=\beta} \tag{1.31}$$

This is the derivative with respect to the coefficient of m, total expenditure, of the moment generating function of $\mathbf{X}$.

Feuerverger, 1989 proves that the empirical MGF and its derivatives are unbiased estimators for the theoretical MGF and its derivatives. So, in the above we can replace expectations with sample averages to get an unbiased estimator of market demand,

$$\widehat{\frac{p\mathbf{C}}{N}} = \frac{\partial}{\partial t_m}M_{N,\mathbf{X}}(\boldsymbol{\beta}) \tag{1.32}$$

where

$$M_{N,\mathbf{X}}(\boldsymbol{t}) = \frac{1}{N}\sum_{j}^{N} e^{\langle \boldsymbol{t},\mathbf{X}_j\rangle} \tag{1.33}$$

If, however, as is common in applied economic literature, we use the log values of our predictors, that is

$$\mathbf{X} = log(\mathbf{x}) \tag{1.34}$$

Then our market demand can be simplified even more.

$$\mathbf{E}\left[\frac{p\mathbf{C}}{N}\right] = \int_{\mathbf{x}} me^{\langle \boldsymbol{\beta}, log(\mathbf{x})\rangle}dF(\mathbf{x}). \tag{1.35}$$

This simplifies to a Cobb-Douglas type function in our predictors,

$$= \int_{\mathbf{x}} m^{1+\beta_m}\mathbf{x}^{\boldsymbol{\beta}}dF(\mathbf{x}) \tag{1.36}$$

This last can be estimated using a bootstrapped sample from the empirical distribution of the predictors and Monte Carlo integration.

1.3.4 Interpretation of Parameters

There is always interest to connect the parameters of an econometric model to interpretable economic quantities. Elasticities of relevant responses with respect to relevant predictors are of primary interest. In the case of the Beta regression, it would be ideal to understand how the parameters connect to the elasticies of expenditure or market demand. There is not just one elasticity that might be of interest, though. Ignoring the differences between conditional and unconditional expectations for the time being, there seems to be four elasticites the econometrician might be interested in. Letting Y be market expenditure, y be individual expenditure, and x the exogenous variable of interest[11]:

1) Elasticity of average (individuals) expenditure:

$$\epsilon_1 = \frac{\partial log \mathbf{E}(y)}{\partial log x} = \beta_x \tag{1.37}$$

2) Expected elasticity of individual expenditure[12]:

$$\epsilon_2 = \mathbf{E}\left(\frac{\partial log y}{\partial log x}\right) \tag{1.38}$$

$$= \mathbf{E}\left(\delta_m(x) + \beta_x e^{\beta \mathbf{X}} \psi^{(1)}(e^{\beta \mathbf{X}})\right) \tag{1.39}$$

3) Elasticity of expected market expenditure:

$$\epsilon_3 = \frac{\partial log \mathbf{E}(Y)}{\partial log x} = \beta_x \tag{1.40}$$

[11] derivations can be found in the mathematical appendix

[12] $\psi^{(1)}(x) = \frac{\partial^2 log \Gamma(x)}{\partial x^2}$ is the trigamma function. $\delta_m(x)$ is Kronecker's delta, which equals 1 when $x = m$. The final expectation is taken with respect to the predictors $\mathbf{X}$

4) Expected elasticity of market expenditure:

$$\epsilon_4 = \mathbf{E}\left(\frac{\partial logY}{\partial logx}\right) \tag{1.41}$$

It is important to ask if an econometric model allows the calculation of all of the elasticities above. Any estimation of an expected value will not allow the calculation of the expected value of an elasticity. Least squares methods will not be able to calculate all of the above, in particular ϵ_2 and ϵ_4, save in an ad hoc manner. Maximum likelihood methods are in a better state than least squares methods, but still have an ad hoc aspect to the calculations. Bayesian methods, on the other hand, through the predictive posterior allows for the derivation of the distribution of functions of the underlying outcome variables; that is predictive posteriors of functions of the outcome. It is not unreasonable to suspect that a model that can estimate all four of the above elasticities is better than one that can only calculate some of them. Granted that, it is still an open question of which one might be most desirable.This question likely does not have one answer, but will depend on the researchers question. My intuition is that ϵ_4 is the most desirable to know, in general. However, it also seems to be the hardest to calculate.

The first three have relatively simple analytic relations to the parameters of the beta regression, even if ϵ_2 does not have an altogether intuitive relation. Qualitative assessments about substitutes and compliments relations, and luxury and normal goods all depend on the sign of the parameters of interest. However, ϵ_4 is not so simple. Even in light of the discussion in a previous section concerning the derivation of market expenditure from the individual expenditures, $Y|X \sim N(\mu_{Y|X}, \sigma^2_{Y|X})$ if the market is large enough, there is no analytic representation for ϵ_4. The reason for this is that the *log* of a normal random variable does not have a distribution that admits of analytic representations for the mean and variance. Thus, this must be calculated using Monte Carlo integration and simulation from the market expenditures asymptotic distribution. Assuming we can calculate $\mu_{Y|X}$

and $\sigma^2_{Y|X}$, we can simulate from the market expenditures asymptotic distribution for specific X patterns. With these we can calculate the expected value of the log market expenditure, $\mathbf{E}(logY|X)$. A dominated convergence argument[13] allows us to switch the order of integration and derivation. If we Monte Carlo integrate $\mathbf{E}(logY|X)$ for a grid of X patterns, the point derivatives can be approximated using finite difference methods.

1.3.5 Welfare Analysis

One of the many desideratum of a demand estimation model is welfare analysis. This allows the economists to evaluate the effects of policies on the welfare of consumers. However, the requirements for the welfare analysis are that demand functions can be used to derive the utility functions that represent preference orderings.

Since at least Deaton and Muellbauer, 1980b, it has been thought that demand functions have to satisfy a trio of conditions that have come to be called the integrability conditions in order to be derivable from the maximization of a utility function that represents a consistent set of preferences:

1. homogeneous of degree 0
2. adding up
3. symmetry and negative semi-definite Slutsky matrix

However, the demand functions that were derived by constrained maximization of the generalized Cobb-Douglas utility function in previous sections only satisfy the adding up condition, in general. They satisfy homogeneity and Slutsky symmetry and negative semi-definiteness if and only if the distribution of the preference parameters are invariant across income and prices. This is something that can be tested with the econometric specification of the model, but will not, in general be true. Yet, the demand functions are derived from constrained maximization of a utility function.

[13] See the mathematics appendix

The objection might be raised that the utility function is not actually representing a consistent preference ordering, one that satisfies transitivity, completeness, and irrelevance of irrelevant alternatives. This is false, though. For a single consumer that is facing constant prices and income, which can not but be true when a consumer goes to make market purchases, the utility function is just a standard Cobb-Douglas utility function. Any binary relation induced by a Cobb-Douglas utility function is a consistent preference ordering[14] The demand functions derived in previous sections can be used for welfare analysis. Following Hausman, 1981, measures of welfare for consumers are derived from the Hicksian demand function, in particular, the compensated variation (CV) for a price change:

$$CV = \int_{p_0}^{p_1} h(p, \bar{u})dp \tag{1.42}$$

where $h(p, \bar{u})$ is the Hicksian demand function for a set of prices and constant level of utility, $\bar{u}$. Roy's identity allows us to rewrite this in terms of the expenditure function, $e(p_0, \bar{u})$, which tells us the expenditure needed to achieve utility $\bar{u}$ given prices p_0:

$$CV = e(p_1, v(p_0, m)) - e(p_0, v(p_0, m)) = e(p_1, v(p_0, m)) - m \tag{1.43}$$

Where $v(p_0, m)$ is the indirect utility function, the utility from the utility maximizing bundle under prices p_0 and expenditure m. For the Cobb-Douglas all of these functions are known:

$$e(p, u) = u \prod_i \left(\frac{\alpha_i}{p_i}\right)^{\alpha_i} \tag{1.44}$$

$$v(p, m) = m \prod_i \left(\frac{p_i}{\alpha_i}\right)^{\alpha_i} \tag{1.45}$$

Therefore, the compensated variation for a price change given our specification is

[14]This follows because the Cobb-Douglas utility is a function defined on all of $\mathbf{R}^n$ and $\leq$ defines a strict order on $\mathbf{R}$. From these two it follows that a binary relation, R, defined by xRy if and only if $U(x) \leq U(y)$ is transitive, complete, and satisfies irrelevance of irrelevant alternatives. Therefore, R is a preference ordering.

$$CV = m\left(\Pi_i \left(\frac{\alpha_{i,1}}{p_{i,1}}\right)^{\alpha_{i,1}} \Pi_i \left(\frac{i,0}{\alpha_{i,0}}\right)^{\alpha_{i,0}} - 1\right) \tag{1.46}$$

Given the specification of the utility function defined in previous sections, the preference parameters change as the prices change. Hence, the CV formula gets a tad complicated and cannot be simplified more, as would be the case in a standard Cobb-Douglas utility function. Hausman, 1981 then defines the dead weight loss (DWL) as the CV less the extra expenditure due to the price rise:

$$DWL = CV - (p_1 - p_0)x(p_1, m) \tag{1.47}$$

Where $x(p_1, m)$ is the Marshallian demand curve.

To actually evaluate the CV for a single consumer, the α terms have to be drawn from the relevant population distributions, i.e. as prices change the distributions change. To ensure that the draws from the different distributions correspond to the same consumer, we assume that consumers order in the distribution is invariant to changes in the distribution. This assumption and its opposite have nothing empirical to recommend themselves. In a probabilistic sense, it is far more likely that consumers order in the distribution is likely to change (there is only one way for their order not to change, but an uncountably infinite number of ways for their order to change). However, very little rests on this assumption, and it allows us to make welfare calculations for individual consumers. Thus, I follow Blundell et al., 2017 in maintaining this assumption.

$$\alpha_j = F^{-1}(u_j|\mathbf{X}_j; \hat{\beta}, \hat{\phi}) \tag{1.48}$$

Where $u_j \in (0, 1)$ is the quantile position for consumer j and F^{-1} is the inverse cumulative distribution function of the Dirichlet/Beta distribution. With estimates of β and ϕ, the DWL can be calculated for any individual consumer. The distribution of the dead weight loss can also be evaluated by letting u range over the entire unit interval and the (p, m) range of their

empirical distribution.

1.4 Empirical Exercises

1.4.1 Consumption Expenditure Survey Data

To present an example of this model in action, it has been fit to data from the Consumer Expenditure Survey (CES)(U.S. Census Bureau, n.d.). This data set is collected by the U.S. Census Bureau (CB) for the Bureau of Labor Statistics (BLS). The BLS uses this data in its calculations of the consumer price indices. They also make the data available to the public in both aggregate and microdata forms. The model will be fit to data pulled from the microdata releases.

The CES has two parts: the interview and diary surveys. The diary survey contains data on expenditures on "small, frequently purchased items, including most food and clothing." The CB surveys 12,000 households, of which roughly 6,900 return usable diaries. These diaries record information on all expenditure surveys for a number of everyday categories, like food purchased for consumption at home, food purchased away from the home, housekeeping supplies, etc... In what follows, we shall estimate our model using the food at home and food away from home categories, with an other category containing everything else reported, when needed. We use data from from years 2012 through 2017.

Our model does not attempt to solve the problem of zeros in demand data. So, for each category we estimate, we will eliminate observations that report less than or equal to zero expenditure on a certain category. We will calculate total expenditure by summing all the reported categories in the diary survey. This is justified due to the sub-compositional consistency of the Dirichlet distribution. Thus, each model is fit to $46,906$ observations.

We also use price data in the form of the BLS's CPI measures. In particular, we use the CPI measures for food away from home, food at home, non-food and energy, and the total CPI. An attempt is made to match each respondent in the diary survey to the CPI measures

in their region. Thus, for Midwest respondents to the CES, the Midwestern CPI measures were used. Since the diary survey presents data in quarters, the quarterly averages from the CPI data is used.

In the next section we compare the results of fitting the Consumer expenditure survey data with the beta regressions and the standard AIDS and QUAIDS models (Deaton and Muellbauer, 1980a, Banks et al., 1997). We compare them with respect to the estimated coefficients, a numerical goodness of fit measure (ELPD), the estimated elasticities, and graphical measures of fit.

1.4.2 Consumption Expenditure Survey Results

Tables 1 and 2 present the model coefficients estimates and their 96% highest density intervals (HDI) as well as the leave-one-out cross validated (LOO-CV) log score (ELPD), which, following Vehtari et al., 2017, is a goodness of fit value that can be used to compare models fit to the same data. The only thing in these tables that can be compared between the two families of models is the ELPD and the signs of the coefficient estimates. The ELPD values are strongly in favor of the Beta regressions (larger ELPD is better). These numbers are corroborated by the posterior predictive checks, especially the posterior density checks, that can be seen in the first column of figures 1 and 2. The second two columns of figures 1 and 2 present functionals (mean, median, standard deviation, and Interquartile range) of the predictive posteriors. The functionals of the predictive posteriors offer similar corroboration of the ELPD, though not as compelling as the predictive densities. We see that all the models do a fairly good job of capturing the mean, save the linear beta regression (Beta 1) for food away. The AIDS models do a better job of capturing the standard deviation of the data for food at home than either beta regression; however, the quadratic income beta regression (Beta 2) captures the standard deviation better than any of the other models for food away. For the Median and IQR, the story is consistent between the two outcomes, the beta regressions do a significantly better job than the AIDS family models. The explanation

for these performances fitting functionals of the data is that the beta regressions fit the data whereas the AIDS family models fit the expected values.

The coefficient estimates, in that they are coming from different models, will have different meanings. Thus, there is little value in comparing the magnitudes across the different families. However, the signs are still meaningful. For the most part, the signs are the same between the two families. There is one exception. The own-price coefficient in the food at home models has a negative sign in the Beta regressions but a positive sign in the AIDS models. However, we notice that the HDI of the own-price coefficient in all the models spans 0 and thus overlap. So, we don't have strong evidence that the signs are actually different.

Tables 3 and 4 present ϵ_2, expected elasticity of individual expenditure, for prices and income/total expenditure and 96% HDI. One of the advantages of the Bayesian estimation paradigm is that we are able to calculate the expected value of elasticities, rather than the elasticities of the expected values, which is the standard in frequentist estimation[15]. Using the predictive posterior, we calculate an individual's elasticities conditional on the observed predictors. Then, by means of a Monte Carlo integration, we can take the expected value of these with respect to the observed predictor distributions, i.e. against the empirical CDF of the predictors.

The first pattern to notice in table 3 and 4 is that the estimates of the price elasticities within each family are statistically indistinguishable. Not surprisingly, the quadratic income terms do not alter the estimates of the price elasticities. However, the two families of models disagree significantly between each other. The magnitude of the cross-price elasticities are larger in the beta regressions than the AIDS family models, without exception. The magnitudes of the own-price elasticities don't share this pattern; for food at home, the beta regression estimates larger own-price elasticity magnitude, whereas for food away, the AIDS family estimates a larger magnitude.

On the other hand, the estimated income elasticities are all very similar to each other.

[15]The issue stems from the elasticity function being non-linear. So, the expectation operator and the log cannot be interchanged.

For both outcomes, the linear income models (Beta 1 and AIDS) estimate statistically indistinguishable income elasticites (the HDI overlap). The quadratic models (Beta 2 and QUAIDS) have statistically indistinguishable estimates for income elasticity for food away, but a slight difference for food at home (Beta 2's HDI is $(0.854, 0.874)$ compared to QUAIDS' HDI of $(0.898, 0.911)$).

Figures 3 and 4 present predictions of 2018 posterior consumption expenditure shares using the models estimated with 2012-17 data. These are not pure predictions. The price levels and total expenditure from 2018 were used as input to the model to calculate predictions for 2018 expenditure share. Visual inspection of the densities show the same pattern as the 2017 posterior densities did. The second beta regression has the closest fit to the empirical kernel density. We see the same kind of pattern in the plots of functionals. The beta regressions capture the summary statistics of the observed sample as well or better than the AIDS family models do, without exception.

Table 1.1: Food at Home

	Beta 1	Beta 2	AIDS	QUAIDS
Constant	0.354	0.196	0.234	0.221
	[−0.488, 1.176]	[−0.671, 0.978]	[0.014, 0.521]	[0.018, 0.497]
Food at Home Price	−0.181	−0.160	0.073	0.074
	[−0.415, 0.056]	[−0.391, 0.062]	[−0.019, 0.161]	[−0.016, 0.158]
Food Away Price	−0.353	−0.337	−0.118	−0.114
	[−0.459, −0.240]	[−0.446, −0.231]	[−0.163, −0.073]	[−0.160, −0.068]
Other Price	0.325	0.320	0.085	0.083
	[0.169, 0.476]	[0.174, 0.468]	[0.018, 0.152]	[0.018, 0.147]
Expenditure/CPI	−0.062	−0.101	−0.032	−0.048
	[−0.068, −0.057]	[−0.108, −0.093]	[−0.035, −0.029]	[−0.052, −0.045]
Expenditure/CPI squared		−0.037		−0.017
		[−0.041, −0.032]		[−0.019, −0.015]
Num.Obs.	46 906	46 906	46 906	46 906
ELPD	5858.7	5975.7	3869.4	3999.7
ELPD s.e.	89.8	91.1	122.5	123.3

ELPD is the expected log score evaluated using a LOO-PIS approximation. Higher numbers indicate better predictive performance

Table 1.2: Food Away

	Beta 1	Beta 2	AIDS	QUAIDS
Constant	−4.955	−4.872	0.026	0.026
	[−6.072, −3.968]	[−5.973, −3.859]	[0.001, 0.095]	[0.001, 0.098]
Food at Home Price	0.787	0.776	0.051	0.053
	[0.504, 1.079]	[0.489, 1.063]	[−0.016, 0.122]	[−0.018, 0.125]
Food Away Price	0.264	0.245	0.028	0.024
	[0.139, 0.411]	[0.113, 0.384]	[−0.013, 0.068]	[−0.017, 0.065]
Other Price	−0.365	−0.353	−0.029	−0.028
	[−0.552, −0.176]	[−0.541, −0.164]	[−0.088, 0.029]	[−0.086, 0.032]
Expenditure/CPI	−0.087	−0.048	−0.032	−0.018
	[−0.094, −0.080]	[−0.057, −0.040]	[−0.034, −0.029]	[−0.021, −0.015]
Expenditure/CPI squared		0.034		0.014
		[0.030, 0.039]		[0.012, 0.016]
Num.Obs.	46 906	46 906	46 906	46 906
ELPD	13 243.9	13 329.1	7518.0	7623.8
ELPD s.e.	127.7	128.4	149.6	150.2

ELPD is the expected log score evaluated using a LOO-PIS approximation. Higher numbers indicate better predictive performance

Table 1.3: Elasticities of Food Away with respect to the prices of Food home, Food Away, and Other, and total expendiutre by model

	Beta 1[b]	Beta 2[b]	AIDS[a]	QUAIDS[a]
Foodhome	1.166	1.148	0.211	0.207
	[0.734, 1.592]	[0.72, 1.572]	[-0.008, 0.443]	[-0.021, 0.443]
Foodaway	-0.61	-0.637	-0.88	-0.898
	[-0.814, -0.412]	[-0.839, -0.439]	[-1.013, -0.75]	[-1.032, -0.769]
Other	-0.54	-0.522	–0.093	-0.092
	[-0.82, -0.261]	[-0.803, -0.243]	[-0.288, 0.095]	[-0.281, 0.102]
Income	0.871	0.928	0.897	0.924
	[0.86, 0.881]	[0.915, 0.941]	[0.89, 0.905]	[0.916, 0.932]

(a) For the AIDS and QUAIDS models, the elasticities were calculated following the equations in Poi, 2012.

(b) For the Beta models, these are unconditional average elasticities as derived in previous sections

Table 1.4: Elasticities of Food Home with respect to the prices of Food home, Food Away, and Other, and total expendiutre by model

	Beta 1[b]	Beta 2[b]	AIDS[a]	QUAIDS[a]
Foodhome	-1.245	-1.217	-0.804	-0.792
	[-1.565, -0.928]	[-1.521, -0.909]	[-1.01, -0.602]	[-0.996, -0.598]
Foodaway	-0.476	-0.454	-0.245	-0.23
	[-0.625, -0.325]	[-0.6, -0.308]	[-0.351, -0.142]	[-0.338, -0.125]
Other	0.438	0.431	0.191	0.188
	[0.232, 0.646]	[0.235, 0.632]	[0.04, 0.343]	[0.041, 0.334]
Income	0.916	0.864	0.928	0.905
	[0.908, 0.923]	[0.854, 0.874]	[0.922, 0.933]	[0.898, 0.911]

[(a)] For the AIDS and QUAIDS models, the elasticities were calculated following the equations in Poi, 2012.

[(b)] For the Beta models, these are unconditional average elasticities as derived in previous sections

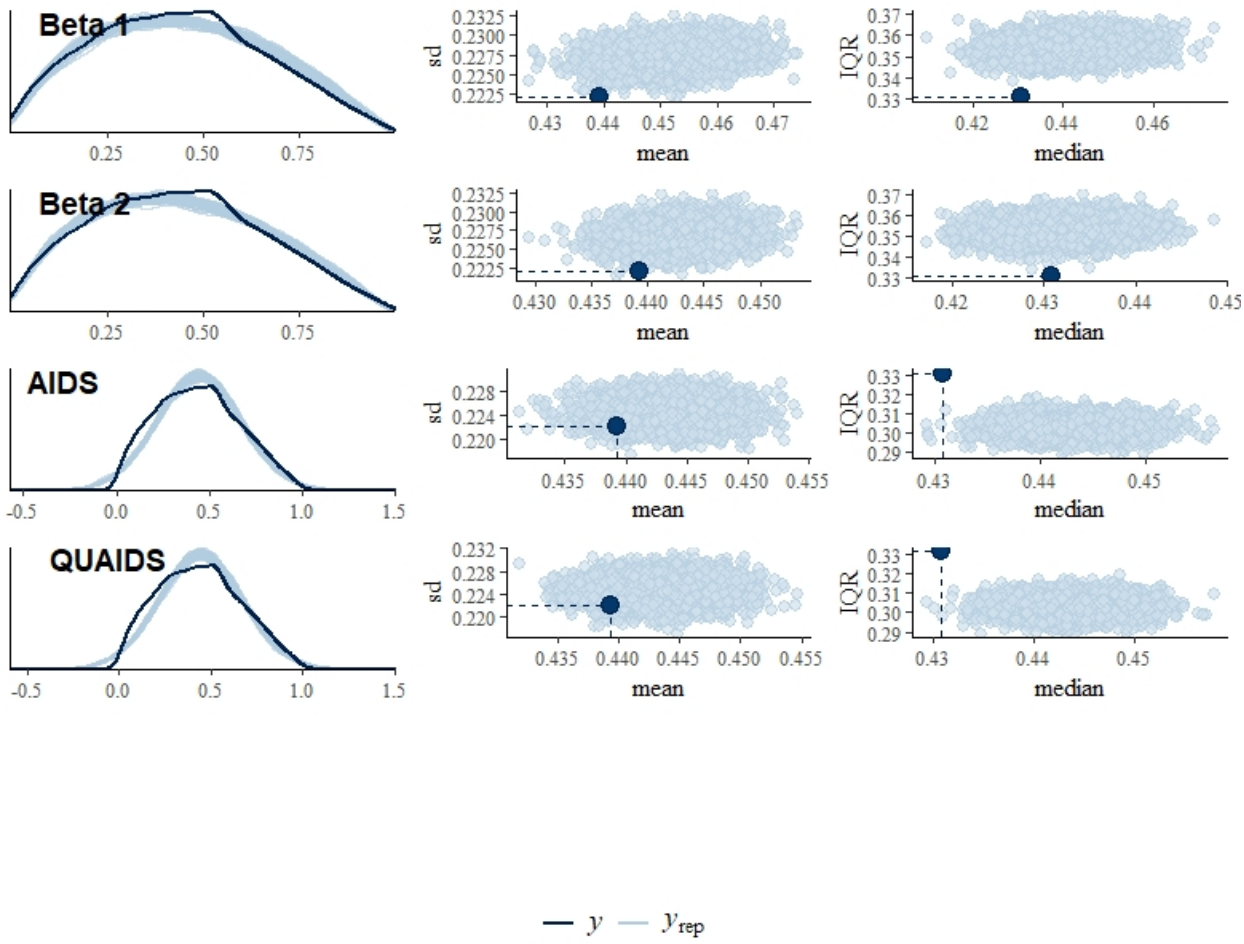

Figure 1.1: Predictive Posterior Checks for Food at Home 2017

1.4.3 National Household Travel Survey Data

Blundell et al., 2017 developed a nonparametric quantile model of demand that they used to fit gasoline demand data from the 2001 National Household Travel Survey (NHTS)(U.S. Department of Transportation, Federal Highway Administration, n.d.). This is a household level survey conducted by telephone and complemented by travel diaries and odometer readings. Blundell et al., 2017 "selected the sample to minimize the heterogeneity." Their sample was restricted to white households with two or more adults with at least one child under 16. They dropped rural and Hawaiian respondents. They restricted the sample to gasoline-based vehicles and households that had non-zero gasoline budget shares (and removed one household that had a budget share for gasoline above 1). They restricted their sample in this way

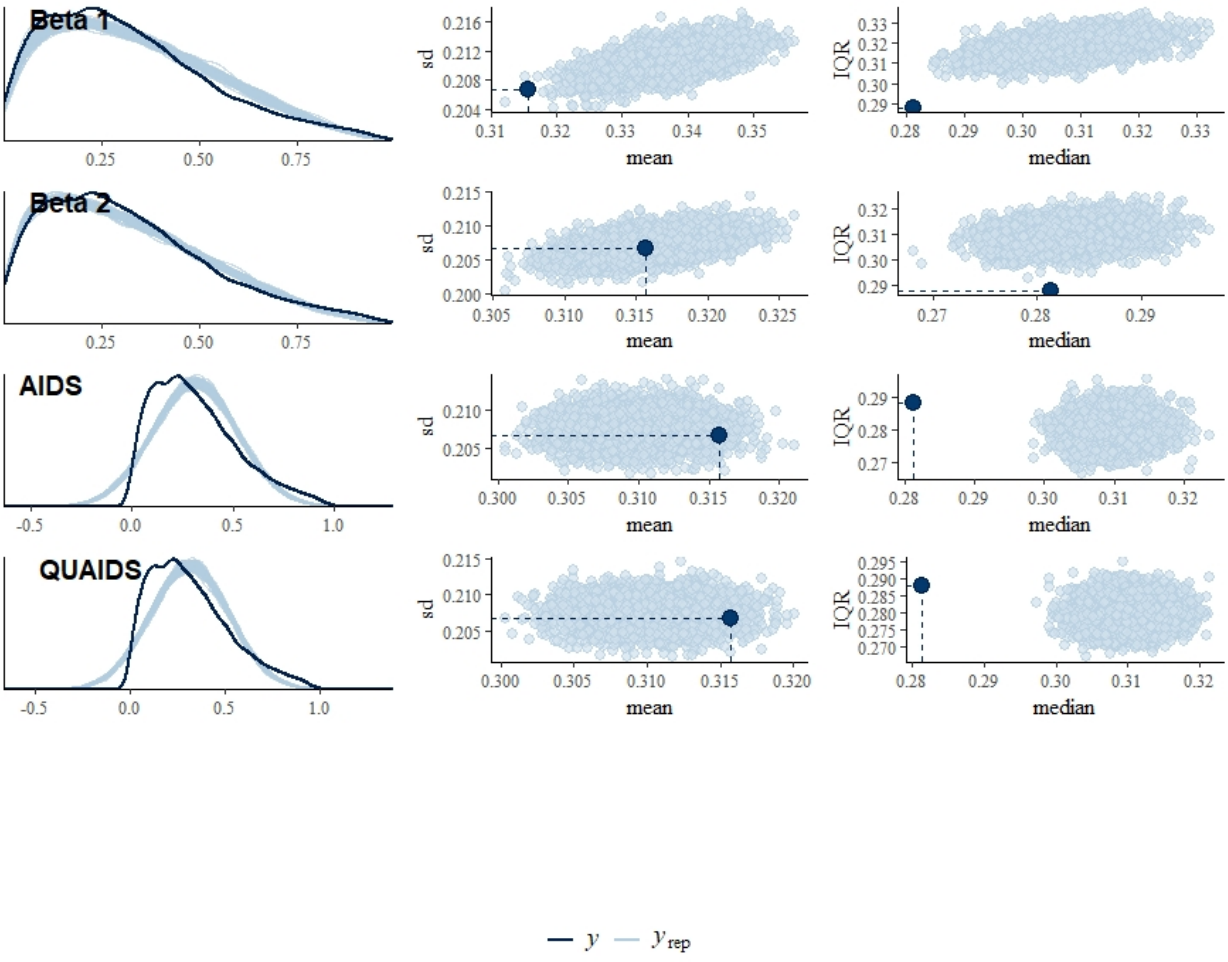

Figure 1.2: Predictive Posterior Checks for Food Away 2017

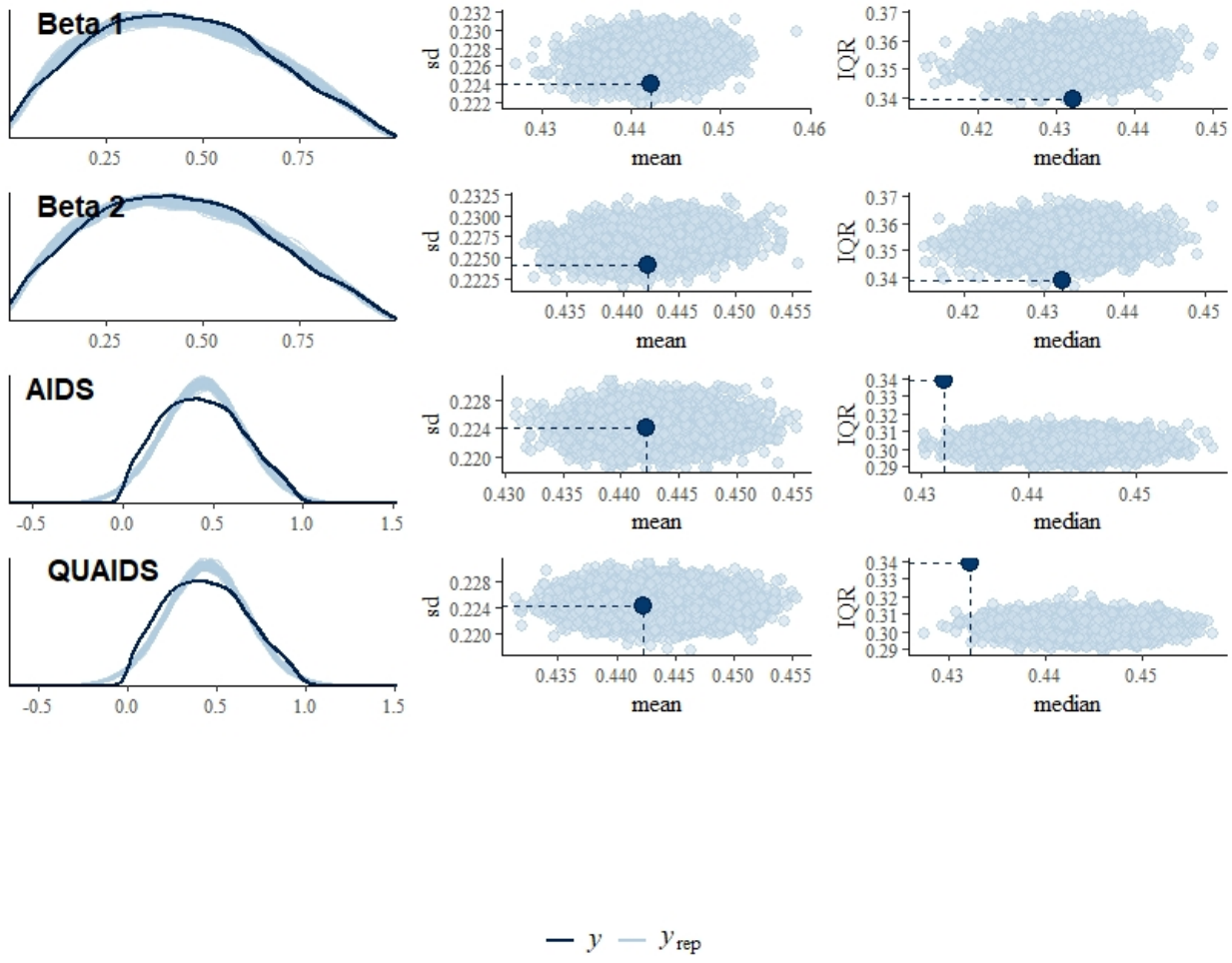

Figure 1.3: Predictive Posterior for 2018 Food Home data using estimates trained with 2012-17 data

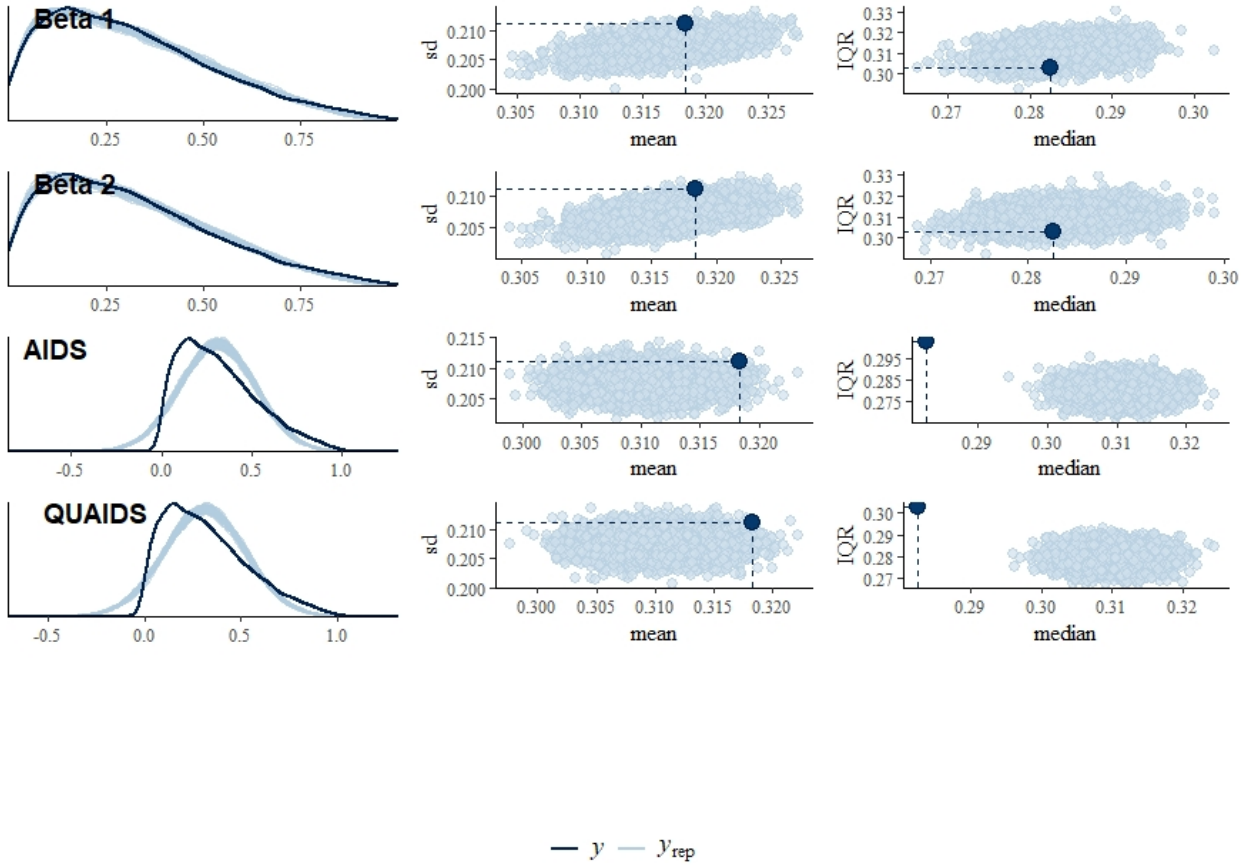

Figure 1.4: Predictive Posterior for 2018 Food Away data using estimates trained with 2012-17 data

to avoid having to consider systematic differences in gasoline demand due to demographic differences.

1.4.4 Nonparametric quantile demand model

The model developed by Blundell et al., 2017 is a nonparametric quantile regression of demand for gasoline. Their model satisfied the following identification condition

$$P(W - G_\alpha(P, Y) \leq 0 | P, Y) = \alpha \tag{1.49}$$

$$G_\alpha(P, Y) = \sum_{m_1} \sum_{m_2} c_{m_1, m_2} B_{m_1}(P) B_{m_2}(Y) \tag{1.50}$$

Where $B(x)$ are b-splines in x, c_{m_1,m_2} are the estimated coefficients. Blundell et al., 2017 used cross-validation to select the number of knots for each predictor (m_1 and m_2). W is the budget share and G_α is the estimated budget share[16]; P the log price of gasoline; Y the log of consumers income. To estimate the model, they numerically solved

$$\sum_i \rho_\alpha(W_i - G_\alpha(P_i, Y_i)) \tag{1.51}$$

$$\rho_\alpha(x) = (\alpha - \mathbf{1}(x < 0))x \tag{1.52}$$

Blundell et al., 2017 fit this model both unconstrained and constrained to satisfy the slutsky condition:

$$\frac{\partial G_\alpha}{\partial p} + G_\alpha \frac{\partial G_\alpha}{\partial y} \leq G_\alpha(1 - G_\alpha) \tag{1.53}$$

This constraint, according to economic theory, should be true at every (P, Y) pair. However, that represents an uncountable number of constraints. To make the above constrained problem tractable, Blundell et al., 2017 restricted the constraint to a finite grid in the (P, Y) plane. With the slutsky constraint imposed, Blundell et al., 2017 argued that the model

[16] the notation in Blundell et al., 2017 is followed in this and the following section

could afford calculation of welfare for individual consumers by assuming that each quantile represented a single consumer across the range of prices; this is the same assumption used to calculate welfare for the beta regression. To this end, they used their model to calculate the dead weight loss of a price increase in gasoline.

In the next section, we compare the results of fitting the restricted NHTS sample with the slutsky constrained b-spline model above[17] and the beta regression model with quadratic income term.

1.4.5 National Household Travel Survey Results

In figures 5-7, the predictions from the Slutsky constrained b-spline model (red), the beta regression with quadratic income term (blue), and the raw data (green) are plotted for each of the three income quartiles ($42,500, $57,500, $72, 500) and three consumers, representing the three quartiles in the preference distributions ($\alpha = 0.25, 0.5, 0.75$ in the specification in the welfare analysis section). The raw data is filtered to observations where the income is around the quartile of interest ($5,000 on either side).

The general location of the b-spline and beta regression medians are basically the same for all three income quartiles. The slopes are not quite the same across the range of $log(p)$. However, the quartiles, α, get progressively different as the income quantile increases. It seems the beta regression may be overestimating the spread of the distributions as income increases. This could be due to assuming ϕ is constant across incomes. The variance of the beta distributions is

$$\frac{\mu(1-\mu)}{\phi+1} \tag{1.54}$$

estimating ϕ as a function of income could perhaps rectify this difference. This is not explored in this paper.

[17] Results of fitting the b-spline model are available from the accompanying website for Blundell et al., 2017here. In what follows, the data and results for the b-spline model are pulled from this site.

It is hard to compare the two methods save visually. Numerical measures, like ELPD which was used in the previous empirical exercise, requires the same in-sample data to be predicted. It is not clear what data should be predicted to appropriately compare the quantile and beta regression. This paper does not attempt to solve this problem. Instead, we shall just point out that the predictions are visually close together and both seem to fit the data well. This fit does have vastly different costs for each model. The b-spline model required selection of knots; Blundell et al., 2017 selected about 4 knots per variable per quantile, amounting to a total of 48 parameters fit over all three income quartiles. The b-spline model would have to estimate 16 additional parameters for every additional income quantile the researcher has interest in. The beta regression estimated 5 parameters for the entire distribution (this would increase to 6 if ϕ was modeled as a function of income).

Table 1.5: Estimates of deadweight loss from the Beta regression and B-spline quantile regressions: evaluated for three incomes at three different quantiles in the conditonal distributions

	Beta			*B − spline*		
Income	25%	50%	75%	25%	50%	75%
\$42,500	4.97	11.39	20.57	0.85	18.57	1.17
\$57,500	4.39	11.95	23.15	29.18	30.20	39.40
\$72,500	3.81	12.34	25.43	12.74	16.32	20.64

Table 5 presents the measures of DWL for both the b-spline and beta regression models. The magnitudes of the estimated DWL are roughly comparable, save for the middle income group. The most salient differences are the large difference in dead weight loss estimate for middle income groups between the two models and that the beta regression estimates are monotonic in both α and m, save for at the 25^{th} percentile. The b-spline regression identifies middle consumers as hurt the most by the price increase[18]

[18]However, it is not entirely clear how Blundell et al., 2017 are holding utility constant in their evaluation

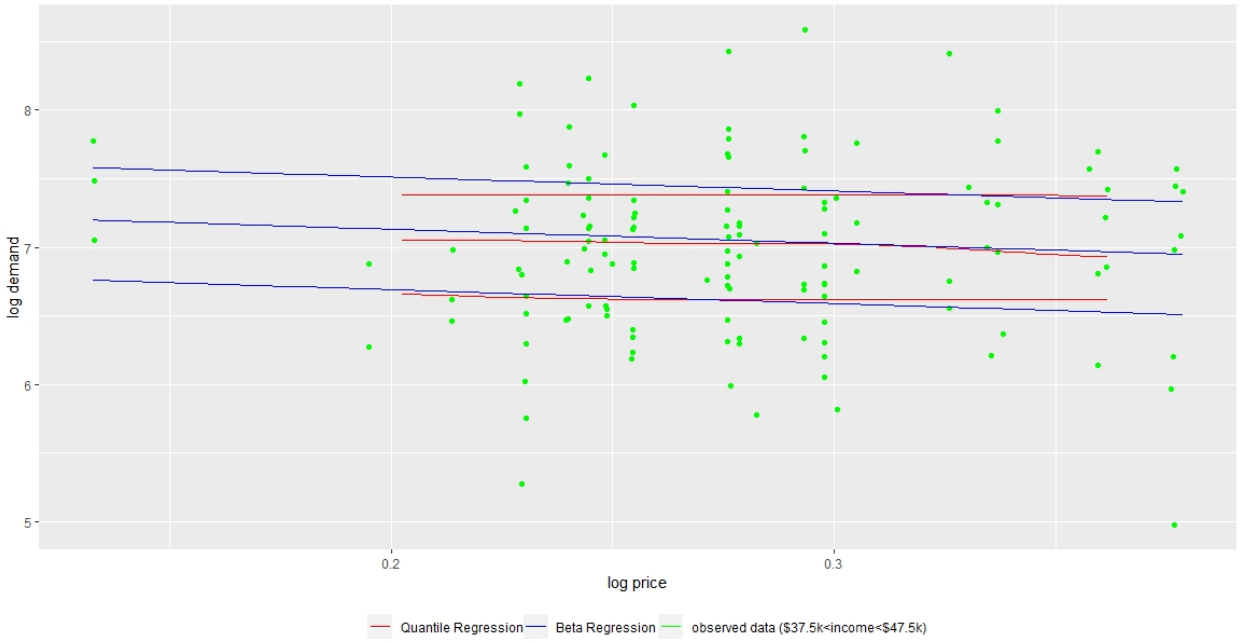

Figure 1.5: Comparison of the beta regression to the b-spline quantile regression, low income group

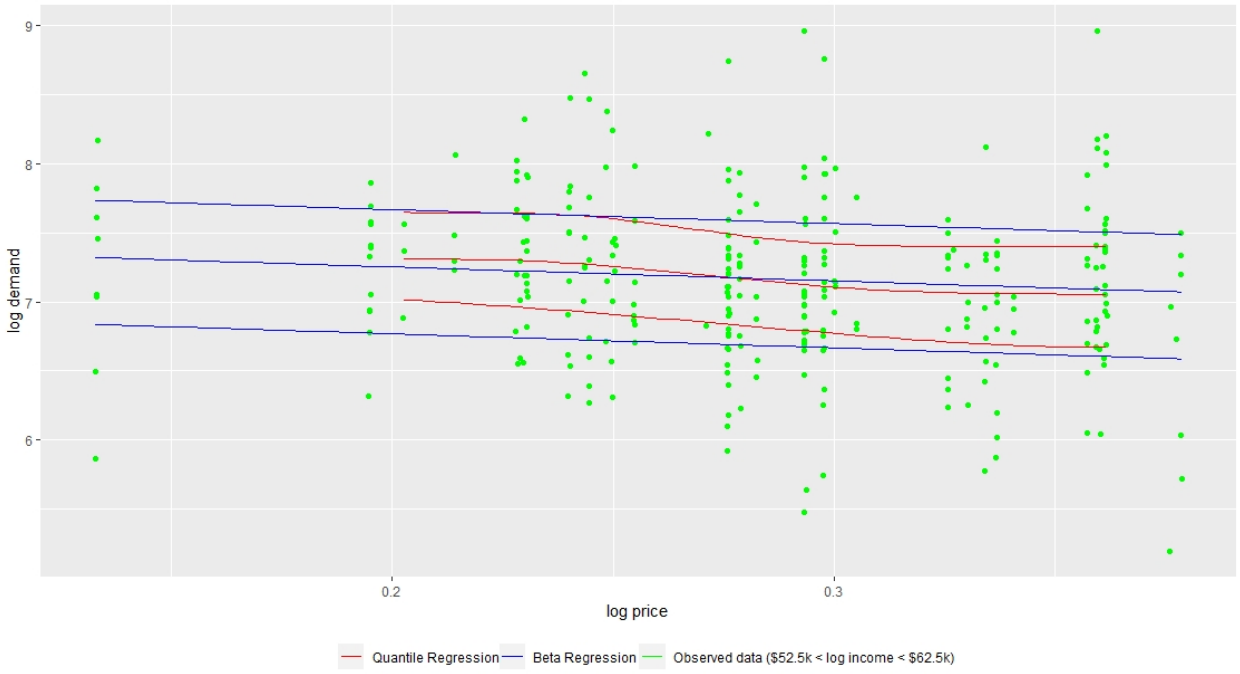

Figure 1.6: Comparison of the beta regression to the b-spline quantile regression, middle income group

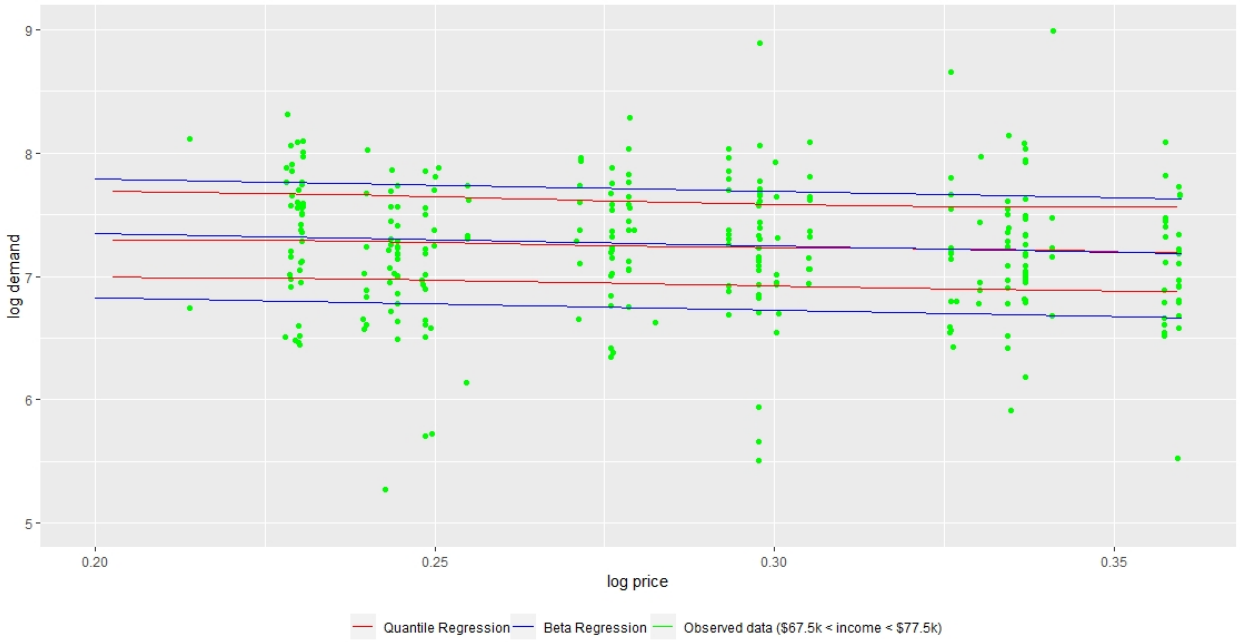

Figure 1.7: Comparison of the beta regression to the b-spline quantile regression, high income group

1.5 Concluding Remarks

In this paper, an old model for estimating demand has been dusted off and given new life by connecting it to utility maximization explicitly. The Dirichlet/Beta regression specification of consumer demand, first introduced by Woodland, 1979 and J. M. Fry et al., 1996 was shown to follow from a very simple Cobb-Douglas utility specification wherein the preference parameters were assumed to follow a Dirichlet distribution the parameters of which were assumed to be functions of income and price, though one could add any other variables

of CV, nor is it clear how they could do so. Their equation is

$$DWL = e(p_1) - e(p_0) - (p_1 - p_0)H(t, e(t)) \tag{1.55}$$

Where H is the Marshallian demand curve and e the expenditure function derived numerically from the differential equation:

$$\frac{de(t)}{dt} = H(t, e(t))\frac{dp}{dt} \tag{1.56}$$

This only works if preferences are homothetic or the differential equation is solved along a line path between p_1 and p_0 (Hausman, 1981, Chipman and Moore, 1976). The issue is there is an uncountably infinite number of paths from p_1 to p_0. It is only justified to assume the unique straight line between the prices if preferences are homothetic. Without homothetic preference, the Hicksian demand or indirect utility function is needed by which utility is kept constant. It is not clear how, given the specification of the b-spline model, the Hicksian demand or indirect utility can be recovered.

one thinks could influence demand (demographics for example). Derivation of demand distributions followed easily. Estimation of these demand distributions was then shown to be a well studied instance of the generalized linear model.

The simple form of the utility function additionally allows for relatively straightforward calculation of a number of economically interesting values, elasticities (both individual and population), market demand, and measures of welfare. A Bayesian estimation technique further allowed us to evaluate the distribution of all these measures, via Monte Carlo sampling of the posterior distribution and Bootstrapping the predictors' empirical distribution. A pair of empirical exercises showed that the beta regression performed as well or outperformed the standard demand models (AIDS and QUAIDS) and performed comparably to a much more demanding nonparametric quantile regression.

The overall performance and simplicity of the Dirichlet/Beta regression model recommends itself for future work estimating demand for individual products as well as systems of demand.

Chapter 2

Comparison of Cost-benefit Ratio of Primary Care between Nurse-led and Physician-led Delivery Systems

2.1 Introduction

2.1.1 a. Description of problem

There continues to be significant health disparities for vulnerable populations. Nurse-led clinics (NLCs), also known as Nurse Managed Health Centers and Community Nursing Centers, are health care delivery models developed in the 1970's that most often serve vulnerable, underserved populations and are classified as part of the US health system safety net (Esperate et al., 2012;J. et al., 2014; JM et al., 2011). Nurse-led clinics most often serve populations suffering from health disparities in chronic conditions. Little is known about the cost-benefits of primary care delivered in NLC health care delivery models. Mainstream primary care is currently delivered in the medical model where cost-benefit studies do exist.

Many health disparities are associated with avoidable complications as a result of untreated or undertreated chronic diseases. As Americans are living longer, the incidence

of multiple disease comorbidities increases thereby escalating human and healthcare costs. African Americans are no exception. In fact, health disparities exist for African Americans across most chronic diseases, along with higher rates of co-morbidities. One chronic disease with major health disparities in prevalence and outcomes in African American populations is hypertension. African American (AA) adults (ages 20+) have the highest rates of hypertension in the United States (US), with the prevalence for non-Hispanic US black males being 58.3% compared to 51.0% for US white males and 57.6% for non-Hispanic US black females being 57.6% compared to 40.8% for US white females. Blacks also experience high rates of exacerbations and complications related to hypertension (Virani et al., 2021). Solutions are needed to eliminate this health disparity for African Americans and improve individual and population health outcomes in an economically efficient manner to meet the recommendation of reducing the per capita cost of health care found in two current frameworks for healthcare system improvement, the Triple Aim (DM and ad Whittington J., 2008; M and K., 2012) or the Quadruple Aim (T and C., 2014; R et al., 2015). Cost-benefit analysis can provide guidance for health care system reforms. Health disparities are costly to the United States economy (TA et al., 2011). A cost analysis of the nurse-led clinic primary care model vs. traditional medical model physician-led primary care is needed to discover any potential cost and value of primary care differences.

Although there is a substantial body of research comparing primary care health care costs and patient outcomes for patients with nurse practitioners versus physician providers (Laurant et al., 2018; Horrocks et al., 2002), there is limited research for examining these costs in different primary care delivery system models, i.e., nurse-led, or physician-led. An integrative literature review by Joo and Huber, 2013 of 18 articles from 2000-2013, gave strong evidence that nurse-led care management is effective in reducing health care costs and increasing continuity of care. The authors concluded that nurse-led community-based care management is a health care model that can reduce duplication of services and fragmented health care. Another study conducted by Bicki et al., 2013 comparing nurse-led care to

usual care found that a walk-in clinic run by nurses saved $760 per patient by avoiding emergency room care, where the uninsured population normally would have gone for care. Based on patient responses, it was estimated that 1 in 4 clinic patients went to the emergency room for care, one patient had fifty-five trips to the emergency room in the last year and ten patients had been to the emergency room within the last month. They found that the most common and cost-effective preventative service was hypertension screening and referral for care, resulting in an anticipated future cost savings of $618,000. Further economic analysis is needed on nurse-led primary care delivery. It can be achieved by exploration and application of existing healthcare data sets, such as the national Medical Expenditures Panel (MEPS). The purposes of this study were to compare cost-benefit ratios of primary care for hypertension between a nurse-led and physician-led primary care delivery models and to develop a replicable method for such comparisons.

2.1.2 Description of Nurse-led Clinic (NLC) Study Setting

The primary setting for this study was a Midwestern, urban, academic Nurse-Led Clinic (NLC). The NLC operates in one of the poorest zip codes in the city and provides primary health care and wellness services for under- and un-insured, primarily African American, populations using the Lundeen, 1993 Community Nursing Center (CNC) Model. The Lundeen CNC Model emphasizes collaborative, coordinated, community-based, culturally humble, caring services (Lundeen, 1993; Lundeen, 1999). Co-located in a longstanding, trusted community-based organization for over 35 years, the NLC maintains a collaborative partnership for services that allow NLC clinicians to address their clients' social determinant of health needs. The UWM (University of Wisconsin Milwaukee) Community Nursing Center (CNC) use the Lundeen CNC Model (Lundeen, 1993), to guide the providers' practice. The Lundeen Model utilizes care management and coordination of all health-related services in the collaborating partner agency, including educational, recreational, vocational, and social services to improve the health of underserved populations. No cost analysis has been

published to date on Lundeen model CNC care.

In what follows, the primary care delivered by the NLC is compared to the primary care delivered by the standard physician-led model. Two statistical strategies are described and implemented to compare the cost adjusted measure of the quality of care between the two models of primary care (Nurse-led, Physician-led). In the first strategy, cost savings realized by the NLC due to decreasing the incidence of major hypertension comorbidities are estimated. The second strategy estimates a naïve-benefit measure adjusted cost of complications that can be interpreted as the cost per comorbidity free year. Both strategies show that the NLC care competes admirably with the standard physician-led model of primary care.

2.2 Method

2.2.1 Data Sources

The current study will assess the relative value of the physician-led care (PLC) health care delivery model and NLC health care delivery model by examining data from two sources. The first, the Medical Expenditure Pane Survey (MEPS) was used as a proxy for the physician lead care, on the grounds that there are very few nurse-led care centers in the United States. So, samples from the U.S. population will likely only capture patients that visit physician-led care centers. The second data source will be referred to as the ACHIS data set, short for "UWM Automated Community Health Information System", a patient documentation system, which includes a relational electronic database utilizing clinical data and the Omaha System (KS., 2005) nursing informatics taxonomy to manage the clinical records at UWM's Community Nursing Centers since 1986 (Lundeen and Friedbacher, 1994).

The MEPS data is "a set of large-scale surveys of families and individuals, their medical providers, and employers across the United States"[1]. The MEPS has two major components:

[1] https://www.meps.ahrq.gov/mepsweb/

the insurance and the household components. This study will pull data primarily from the household component. The household data is collected from "a sample of the families and individuals in selected communities across the United States, drawn from a national representative subsample of households that participated in the prior years' National Health Interview Survey..."[2].

The MEPS collects "detailed information for each person in the household on the following: demographic characteristics, health conditions, health status, use of medical services, charges and source of payments, access to care, satisfaction with care, health insurance coverage, income, and employment." The panel design of the survey involves interviewing respondents over several rounds covering two years.

Data from the MEPS years covering 2006 through 2016 was pulled from the MEPS public website[3], as this was the same decade for which ACHIS data was available. The yearly MEPS data were combined, with the expenditure data adjusted to 2016 dollars using the quarterly Consumer Price Index (CPI) for medical expenses from the Federal Reserve Economic Data (FRED) depository[4]. For the survival analyses described in the following sections, the data was put into the counting process form with age being take as the time variable. Thus, the start time was defined as age=0, censor time=age at interview, and event time=age at first diagnosis of the complication of interest.

The experimental data set for this study was derived from UWM ACHIS data. It consisted of primary care patient electronic health records at the UWM's Community Nursing Centers from 2008 to 2016. The full ACHIS data from 1513 primary care patients was filtered to assure that all participants for this study had blood pressure readings recorded in their record so they could be categorized as either hypertensive or not. The data set sample for this study consisted of 1,305 individual patients who had a systolic blood pressure recorded. from one to 592 visits (M=16.8; SD=35.8) with clinic staff during those years.

[2]https://meps.ahrq.gov/mepsweb/about_meps/survey_back.jsp
[3]https://www.meps.ahrq.gov/mepsweb/
[4]https://fred.stlouisfed.org/

Visits included in-person appointments (40.4%), follow up and care coordination. Patient enrollment at the clinic has a mean tenure of 1.7 years, with a maximum tenure of 12.9 years. The sample population was 84% African American and 10% Caucasian. Their median age at the last visit was 43 years old and more were female (72%). The ACHIS patient data included values for systolic blood pressure and diastolic blood pressure. However, since the MEPS data only has a binary variable on diagnosis of hypertension, a variable was created in the ACHIS data set for diagnosed hypertension, which equals 1 if a patient had a systolic blood pressure value above 140. Table 1 has descriptive statistics for both the MEPS control data set and ACHIS experimental data set.

In addition, in several places in what follows there will be need to calculate weighted lifetime averages. The weights of these averages will represent the probability that an individual can be expected to survive to that age. For example, when calculating the average lifetime costs of having a stroke at 18, future costs must be weighted by the probability that the individual will survive to realize those costs. The ideal would be to use survival probabilities estimated from data. However, neither the ACHIS nor the MEPS data allows for that. A literature search did not find data consistently extensive enough for all our purposes. In lieu of that limitation, the general lifetable for survival probabilities produced by the Centers for Disease Control (CDC) (Arias and Xu, 2019) was used in all weighted lifetime averages.

Variable	MEPS	ACHIS
Unique Patients	100,675	1,305
N (observations)	113,948	5,668
Average number of Interactions (sd)	NA	16.8 (35.7)
Average age at last (sd)	46 (17)	41 (13)
Sex (Male)	46%	28%
White	70%	10%
Black	20%	84%
Hypertensive	34%	35%
Myocardial Infarction	3.8%	0%
Stroke	3.9%	0%
Coronary Heart	5.6%-	0%
Smokes	19%	37%
Diabetic	11.2%	4.2%

2.2.2 Overview of cost-benefit ratio analyses

The cost-benefit ratio between primary care delivered in a NLC delivery system and traditional U.S. PLC delivery system is compared using two different statistical strategies. The first strategy estimates the expected number of hypertension complications the NLC's patients would have experienced under the null hypothesis that the chronic disease complications experienced by patients in the two delivery methods are not quantitatively different. The second strategy estimates the probabilities of developing hypertension complications for each care delivery group separately and compares the two directly. Both strategies' results are then combined with estimates of costs and benefits for each complication in order to summarize the incidence of complications for each care delivery into a single number for comparisons. The costs associated with each complication are estimated using the MEPS

data. These costs are assumed to be the same for patients from both care delivery systems. Benefits are coded on a zero-one scale, where one corresponds to a desirable health state or the absence of a negative health event, and zero corresponds to undesirable health events (stroke, heart attack and coronary heart disease).

Using the first statistical strategy, transition probabilities are estimated for the traditional physician-led primary care delivery model using the Medical Expenditure Panel Survey (MEPS)[5] data set. These transition probabilities are then used to calculate an expected number of strokes, heart attacks, and coronary heart diseases that CNC hypertensive patients may have suffered had they received care through traditional primary care delivery, which would not differ under the null hypothesis that the chronic disease complications experienced by patients in the two delivery methods are not quantitatively different. These expected numbers are then compared to the actual numbers of these hypertension complications observed in the same CNC population. Next, the difference between the expected and actual number for each of the three most common hypertension complications are used as an estimate of the number of complications prevented by receiving care at the CNC.

In the second statistical strategy, the transition probabilities for the Nurse-led primary care delivery model implemented at UWM's Community Nursing Center are estimated using the ACHIS electronic data. The transition probabilities, when combined with costs associated with each complication, derived from the MEPS data, and a binary numerical value of welfare assigned for the presence or absence of each complication, allow for the calculation of a probability weighted cost and welfare for a patient with hypertension utilizing each delivery model. The ratios of these weighted cost/welfare measures are then compared across delivery models.

[5]https://meps.ahrq.gov/mepsweb/

2.2.3 Details of Comparison Strategies

First Comparison Strategy

In the initial comparison strategy, the number and cost of complications patients with hypertension experienced in two groups: those visiting a traditional physician-led primary care facility, our control group, and those visiting the Nurse-led primary care facility, our experimental group, are compared. To generate the comparison, the survival curves for the control group are estimated, controlling for race, sex, hypertension, and diabetes. These survival curves are then applied to estimate the probabilities that the experimental group would have any number of complications events/diagnoses. Finally, these are combined with costs estimates to calculate costs saved by the Nurse-led care delivery model.

Weibull Regression/Survival Curves: To estimate the survival curves for the control group, a Weibull proportional hazards model (Ibrahim et al., 2001) is fit to the MEPS data set. The response variable used is the age at which the complication event was first diagnosed. For those that did not experience a complication, the patient's age at the time of the survey is used as the censoring time or the last time of observation. Thus, with stroke as an example, the response (T) is defined as

$$T = \begin{cases} Age\ at\ stroke & if\ stroke \\ Age\ at\ survey\ Time & if\ no\ stroke \end{cases} \tag{2.1}$$

It is assumed that T follows a Weibull distribution. To capture the effects of predictors, a Generalized Linear Model (GLM) (Dobson and Barnett, 2018) approach was employed:

$$\begin{aligned} T_i &\sim Weibull(shape = \alpha,\ scale = \sigma_i) \\ \sigma_i &= exp\left(-\frac{\mu + X_i\beta}{\alpha}\right) \end{aligned} \tag{2.2}$$

$$h(t|X_I) = \alpha e^{\mu} T^{\alpha-1} e^{X_I \beta} \tag{2.3}$$

Where the X_i are the predictor variables for individual i. The β parameters are the log hazard ratios for the predictors:

$$log\left(\frac{h(t|X_i)}{h(t|X_j)}\right) = \beta(X_i - X_j)$$

Estimation is carried out by a Bayesian technique using Stan (Stan Development Team, 2020b), which implements a Hamiltonian Monte Carlo algorithm to sample from the posterior distribution. As much as possible, the full Bayesian model used minimally informative priors. However, standard deviations of normal priors are set to 1. This constitutes a strong constraint on the prior distribution. This can be justified on the ground that even among patients with hypertension, these complications are relatively rare. Thus, hazard ratios for each predictor can be expected to be relatively close to 1. Thus, the prior on the log hazard ratios, β_k, and the intercept, μ, are given unit standard deviations. Thus, the full model is given by equation 2.2 and the priors:

$$\begin{aligned} \mu &\sim Normal(0, 1) \\ \boldsymbol{\beta} &\sim MVNormal(\mathbf{0}, \mathbf{I}) \\ log(\alpha) &\sim Normal(0, 1) \end{aligned} \tag{2.4}$$

The parametric assumption that T follows a Weibull distribution is a strong assumption. To justify this, the complementary log-log plot of survival time vs. risk is explored. For the Weibull function, this is plotting the log of the cumulative hazard function (the complementary log-log of the survival function) versus the log of survival time. This plot must be linear in survival time:

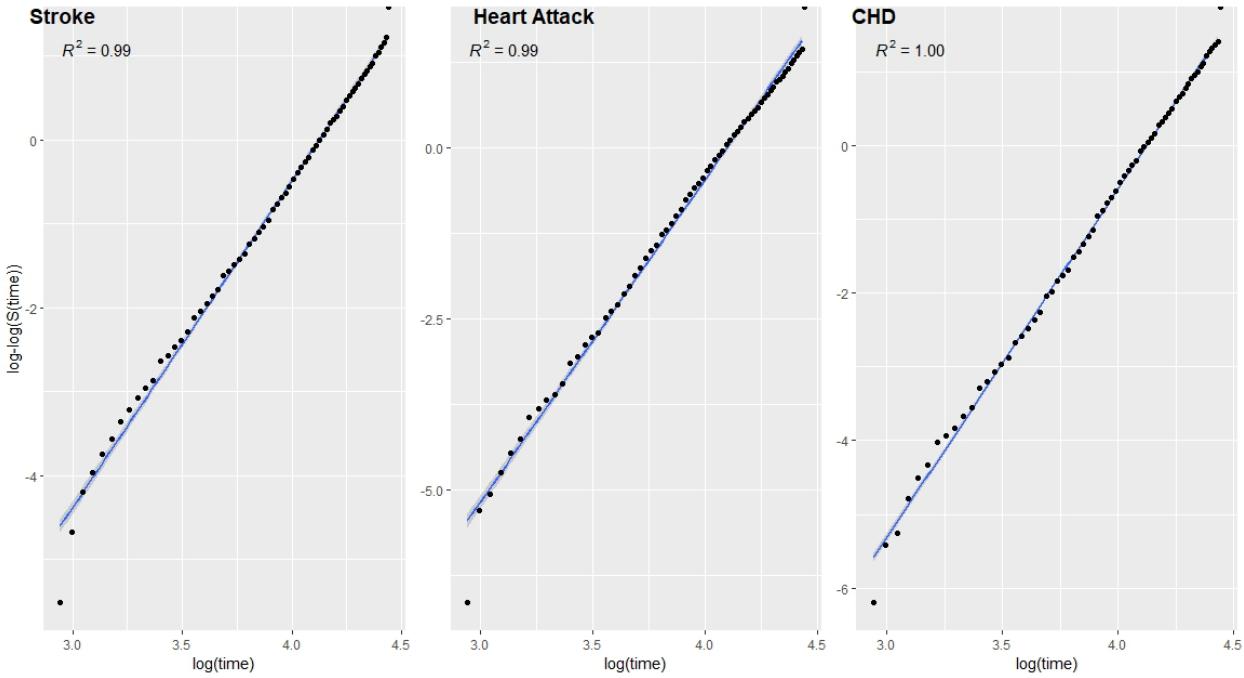

Figure 2.1: Weibull Plots for stroke, heart attack, and CHD

$$\begin{aligned} log(-log(S(T))) &= log(-log(exp(-\left(\frac{T}{\sigma}\right)^{\alpha}))) \\ &= log\left(\left(\frac{T}{\sigma}\right)^{\alpha}\right) \\ &= \alpha log(T) - \alpha log(\sigma) \end{aligned} \tag{2.5}$$

To estimate the log of the cumulative hazard function (H(T)), a Kaplan Meier (KM) curve is fit to the MEPS data. Using this KM curve, S(T) is evaluated for each failure time, T. The complementary log-log is then plotted against the log of survival times. The graphs in figure 2.1 address each of the three hypertension complications. A linear regression is fit to assess the r^2 value; for all three graphs, the $r^2 \geq .99$. Thus, the data fits a Weibull distribution sufficiently well to justify the assumption that T follows a Weibull distribution.

Application of Survival Curves Once the Weibull model is fit, multiple survival curves for the control group are calculated. The survival curve estimations yielded probabilities of

complication for any time, measured in age of the patient. Using these survival curves, the question, 'What is the probability the patient does not have a stroke in 1, 2, 3, ... years?', is answered for each patient in the control group. Under the null hypothesis that there is no difference in quality of care between the two groups, the same survival curves are used for the experimental group.

The survival curves are used to calculate the probabilities that the patients in our experimental group experienced complications under the null hypothesis that they received the same "quality" care as the control group. However, to compare the two groups it was deemed better to explore aggregate results. To do this, the distributions of the number of complications the experimental group experience under the null hypothesis is compared to the actual number of complications. Since each patient in the experimental group is not observed for the same amount of time, the probability is calculated that each of the patients in the experimental group would experience a complication during the time they were observed (as measured by their age). These probabilities are then fed into a Poisson-binomial model (Fernandez and Williams, 2010), which is a model of the sum of independent Bernoulli trials that do not necessarily have the same probabilities. Defining

$$Y_i = \begin{cases} 1 & \textit{patient } i \textit{ has a complication} \\ 0 & \textit{patient } i \textit{ has no complication} \end{cases} \tag{2.6}$$

The Poisson-binomial is the distribution of $\sum_i Y_i$ with probability mass function (pmf)

$$P(Y = k) = \sum_{A \in F_k} \prod_{i \in A} p_i \prod_{j \in A^c} 1 - p_j \tag{2.7}$$

Where F_k is the set of subsets of the patient list of size k, in our case this is a list of all possible ways the experimental group can have exactly k complications. The p_i and p_j

are the probabilities the i and j patients, respectively, have the complication:

$$p_i = P(Y_i = 1) = S(T_{first\ visit}|X_i) - S(T_{last\ visit}|X_i) \tag{2.8}$$

This distribution allows calculation of the probability of observing the number of complications we observed under the null hypothesis that the quality of care/outcomes is the same between the two groups. The expected number of outcomes is then combined with the costs of each complication to calculate expected costs saved and intervals for costs saved.

The adoption of the Bayesian estimation procedure affords the ability to consider all the proceeding uncertainty in the final calculation of these costs savings estimates and intervals. The estimation procedure described in the previous section provided posterior distributions for the model parameters. With these posteriors the predictive survival curves (letting $\theta = \{\alpha\ \mu\ \beta\}$, for ease of notation) were derived:

$$\begin{aligned} 1 - p_i = S(T_i|X_i) &\sim exp\left(-\left(\frac{T_i}{\sigma_i}\right)^{\alpha}\right) f(\theta|T) \\ \sigma_i &= exp\left(-\frac{\mu + X_i^T\beta}{\alpha}\right) \end{aligned} \tag{2.9}$$

This results in a distribution for each T for each patient. These were then multiplied by the probability mass function (pmf) of the Poisson binomial. Thus,

$$Y \sim \sum_{A \in F_k} \prod_{i \in A} p_i \prod_{j \in A^c} 1 - p_j d\mathbf{P} \tag{2.10}$$

Where $d\mathbf{P}$ is the distribution of $\{p_1 ... p_n\}$. Under the assumption that the patients are independent, the expected number of complications and the quantiles, ψ, are calculated:

$$\begin{aligned} \mathbf{E}(Y) = \sum_{k=0}^{n} k \sum_{A \in F_k} \prod_{i \in A} \int p_i dP_i \prod_{j \in A^c} \int 1 - p_j dP_j \\ \sum_{k=0}^{l} \sum_{A \in F_k} \prod_{i \in A} \int p_i dP_i \prod_{j \in A^c} \int 1 - p_j dP_j = \psi \end{aligned} \tag{2.11}$$

Letting $\psi = 0.05, 0.95$ and solving for l in this last equation yields a 90% highest density interval for the number of complications the experimental group should experience under the null hypothesis that the two groups receive equal quality care.

Second Comparison Strategy

In the first comparison strategy, the comparison of expected and actual health outcomes for hypertension patients is made by holding a null hypothesis that the control and experiment groups received similar quality care. The second comparison strategy compares the expected cost-benefit ratios for each group directly; though, not completely dispensing with the null hypothesis. Thus, for the second comparison strategy, survival curves are estimated using the same method described in the previous sections. The survival curves that are estimated for the control group were used as such. For the experimental group, the same Bayesian estimation procedure with an adjustment to the priors is used.

From a statistical point of view, one problematic feature was found in the 2008-2016 NLC data used for this study: the population had no complications from hypertension documented. This led to all observations being censored. Standard frequentist statistical methods are not able to handle this dearth of variability in the response variable. The Bayesian technique adopted for this study still allows the estimation of posterior distributions for our parameters. While there may be a tendency not to trust these estimates, in the following sections, the adjustments to the priors in the Bayesian model and their related rational will be presented.

Adjustments to the Bayesian Priors Recall the Bayesian model from the previous sections, 2.2 and 2.4:

$$\begin{aligned} T_i &\sim Weibull(shape = \alpha,\ scale = \sigma_i) \\ \sigma_i &= exp\left(-\frac{\mu + X_i\beta}{\alpha}\right) \\ \mu &\sim Normal(0, 1) \\ \boldsymbol{\beta} &\sim MVNormal(\mathbf{0}, \mathbf{I}) \\ log(\alpha) &\sim Normal(0, 1) \end{aligned} \tag{2.12}$$

For the experimental group, instead of using non-informative prior distributions, prior distributions informed by the posterior distributions obtained from the estimation of the model on the control data are used. Thus, the model priors are given by

$$\begin{aligned} \mu &\sim Normal(\mu_c, \sigma_c^2) \\ \boldsymbol{\beta} &\sim MVNormal(\boldsymbol{\beta}_c, \Sigma_c) \\ log(\alpha) &\sim Normal(\alpha_c, \gamma_c) \end{aligned} \tag{2.13}$$

Where the c subscript indicates the hyperparameter is derived from the control posterior distributions. These are informative priors. They correspond to the null hypothesis that the two groups receive similar quality care. If the study data supports this null hypothesis, the posterior should not be significantly different from these priors. It is only when the study data supports a conclusion that the two groups are receiving different quality care would the posterior diverge from these priors.

The interpretations of the parameters μ and β_k are as the log of the baseline scale parameter ($\sigma^{-\alpha}$) and the log of the hazard ratio for the covariates, respectively. When dealing with a population that has no chance of experiencing the complication, these parameters are $-\infty$ and 0, respectively. To see why this is the case, the interpretations of these parameters are useful. e^{β_k} is the ratio of the probabilities when predictor k changes by one unit over

an infinitesimal time. When there is no variability in the response variable, the predictor variables are not having a discernible effect on the response variable. This only happens when $\beta_k = 0$. For the case of μ, this controls the baseline probability of experiencing a complication. When there are no complications, the empirical survival curve equals 1 over the entire range of the response variable. For the Weibull distribution specified above, the survival curve is given by

$$S(T|X_i) = exp(-T^{\alpha} exp(\mu) exp(X_i^T \beta)) \tag{2.14}$$

For this to equal 1 everywhere (while $\alpha > 0$, and β is finite), $\mu = -\infty$, which is equivalent to the baseline scale parameter equaling 0 ($\sigma = 0$). The parameters should take these values if the experimental group is believed to have no chance of experiencing complications, irrespective of predictor values. Given the nature of the study's response variable and predictors, this is only believable if the study data set would be infinitely large with no complications. For the finite MEPS and, more particularly, the NLC data sets, two relative claims were inferred: 1) The parameter values for the experimental group should be closer to $\mu = -\infty$ and $\beta_k = 0$ than they are for the control group; 2) how much closer should be a function of the sample size, how large the sample size gets without experiencing a complication.

Rationale for the Adjustments to the Bayesian Model Priors Given the dependence on the prior distribution and sample size, it is natural to ask if the estimates that this Bayesian technique produces are reliable. The answer is a qualified yes. The qualifications, not surprisingly, depend on what one is doing with the estimates, but also on the details of the prior distribution and size of the sample. There are two issues that can engender objections to the estimates: 1) strong dependence on prior distributions entails any estimate can be obtained; 2) dependence on sample size entails that estimates will change as more data is obtained. Each of these issues are addressed below.

It is true, that with a strong dependence on prior distributions that any estimate can be

obtained if one simply uses the appropriate prior distribution. This is not a specific feature of the model and data being considered in this study. It is a feature of every Bayesian model. If one were to specify a degenerate prior distribution, the posterior distribution will be equal to the prior distribution. This has led to a great deal of discussion about the appropriate prior distributions to use in Bayesian model (Gelman et al., 2013). So, with respect to the model considered in the previous section, the objection that strong dependence on the prior distribution entails any estimate can be obtained, amounts to the claim that the control group's posterior distribution is a bad prior distribution for the experimental group.

However, the natural null hypothesis for this study entails that the prior distributions selected above for the experimental group are the only appropriate ones. Suppose the null hypothesis were true. In that case, using the control group's posterior distribution for the experimental group's prior distribution is equivalent to combining both data sets and using the minimally informative prior distribution. This is the idea behind sequential estimation in Bayesian statistics. Suppose one wanted to fit a model to a data set, but the data was received at different times. Say the total data set is denoted by X, but received in two tranches, X_1 and X_2. A Bayesian estimation technique could proceed along one of two lines: 1) wait until all the data were obtained, or 2) perform the Bayesian estimation sequentially as the data were received, using the posterior from the first tranche as the prior when training with the second. The Bayesian posterior distribution in the first case would be given by

$$f(\theta|X) = \frac{f(X|\theta)f(\theta)}{f(X)} \tag{2.15}$$

Assuming a random sample, the two tranches are independent, conditional on the model. So, this could be rewritten as

$$f(\theta|X) = \frac{f(X_1|\theta)f(X_2|\theta)f(\theta)}{f(X_1)f(X_2)} \tag{2.16}$$

Associating the terms for each tranche of the data yields:

$$f(\theta|X) = \frac{f(X_1|\theta)f(\theta)}{f(X_1)}\frac{f(X_2|\theta)}{f(X_2)} = \frac{f(X_2|\theta)}{f(X_2)}f(\theta|X_1) \tag{2.17}$$

This result is just how we would estimate the model sequentially. Thus, the two ways of estimating are equivalent. Therefore, using the control group's posterior distribution as prior distribution for the experimental group entails holding an initial assumption that there is no difference between the two groups. Use of any other prior distribution would bias against the null hypothesis. The limited literature comparing physician-lead and nurse lead primary care delivery models do not support any different prior distribution.

In exploring the intuitions concerning the influence of the sample size in a completely censored case, a simpler model: a coin toss, can be considered. Let X be the sum of n coin tosses with probability of success π, which is unknown. Thus $X \sim Binomial(n, \pi)$ and, assuming a Beta prior: $\pi \sim Beta(a, b)$ with $a, b > 1$, the posterior of π given a sequence of coin tosses is given by:

$$f(\pi|X) \propto \pi^X(1-\pi)^{n-X}\pi^{a-1}(1-\pi)^{b-1} = \pi^{a+X-1}(1-\pi)^{b+n-X-1} \tag{2.18}$$

Thus, $\pi|X \sim Beta(a+X, b+n-X)$. If $X = 0$, a situation analogous to this study's case of all observations being censored, the posterior is $\pi|X \sim Beta(a, b+n)$. Then, the posterior mode, which is often numerically similar to the Maximum Likelihood estimate, is

$$mode(\pi|X) = \frac{a+X-1}{x+b+n-2} = o(n) \tag{2.19}$$

This shows how the point estimate is dependent upon the sample size. As the study's sample size increased with no successes, the mode gets pushed closer and closer toward 0. A

calculation of how the mode changes as the sample size increases can be done:

$$\frac{\partial mode(\pi|X)}{\partial n} = \frac{-(a-1)}{(a+b+n-2)^2} = o(n^2) \tag{2.20}$$

Thus, as the number of coin tosses increases with no success recorded, the mode gets pushed toward 0, but more slowly with each additional coin toss. It is also interesting to calculate how the point estimate changes when we get a success. This can be calculated as follows:

$$\frac{\partial mode(\pi|X)}{\partial X} = \frac{1}{(a+b+n-2)^2} = o(n^2) \tag{2.21}$$

Since n increases by 1 every time X increases by 1, it can be seen that for large n the effect of additional successes is negligible on the point estimate. A significant change in the point estimate would only be seen if, after many failures, a large sequence of successes occurred, comparable in size to the number of failures initially observed. This possibility becomes ever less likely as the number of initial failures increases.

There are two intuitions that can be made from this simplified model. The first is that when a parameter has a lower bound, the estimates' sensitivity to increases in sample sizes will decrease. The second is that when the sample size is large, further increases in the sample size will not change the point estimates much, unless the initial sample was drastically unrepresentative.

The use made for these estimates would be the most important determining factor for their reliability. There are two broad categories of use: 1) explanation, and 2) prediction. In explanation, the values for the parameters associated with the study's predictor variables are of interest. So, in this case, the hazard ration for hypertension would be of interest. In prediction only the resulting survival curve would be of value in making predictions for a new patient or group of patients. Given the dependence on the sample size and priors, caution should be exercised about trusting our estimates if explanation was the goal of the use of the estimates. The reason for this caution, is that within the Weibull model, there are

several ways to adjust the parameters to get the survival curve that reasonably fits the data and the clinical expectations that the experimental patients would still have complications at some point. For instance, to adjust the parameters, the β parameters could be left alone and the μ parameter shifted a great deal or the β parameters could be shifted a lot but the μ parameter only a bit. The tradeoff between these two extremes can be influenced by the number of predictors in the model as well; if there are a lot of predictors the model has no meaningful way to decide which coefficients to adjust so likely μ will house most of the adjustment. Thus, if the intention were to use these estimates to explain the relationship between hypertension and its complications, through the hazard ratio, it would be wise to be very cautious.

However, the story is quite different when we intend to make predictions. When the only interest is survival curves, then the estimates will be much less worrisome. Consider again the tradeoff mentioned in the previous paragraph. In the estimates for our study's experimental group, the lack of complications could be handled by shifting the β parameters or the μ parameter, or both. However, no matter how they shift, the resulting survival curves should be roughly equivalent, capturing that the frequencies of complications in the experimental group is much smaller than in the control group. Thus, while advising caution of the point estimates of the parameters, the estimate of the survival curve will not succumb to the same objections. In this study, we are only interested in the resulting survival curves.

2.2.4 Cost and benefit estimation

Expected costs attributable to three major hypertension complications: stroke, heart attack and chronic heart disease are the focus of this study. The expected costs of these complications are the sum of the attributable costs in the year the complications were first diagnosed and attributable costs in subsequent years. The former is generally larger due to treatment in the acute phase of the complication, for instance, due to an emergency room visit while suffering a stroke or heart attack. However, the costs from subsequent years are only realized

if the patient survives past that year. To account for this, the costs in subsequent years are weighted by the probability of surviving to each respective year using general population life tables. Thus, with a and $a\prime$ representing ages, and letting $C(a, a\prime)$ be the cost in year of life $a\prime$ when a complication was first diagnosed at age a; $C(a)$ being total expected lifetime costs for an individual that suffers the complication at age a, and $P(a, a\prime)$ the probability of surviving past age $a\prime$ when the complication was suffered at age a,

$$C(a) = C(a, a) + \sum_{a\prime} C(a, a\prime)P(a, a\prime) \tag{2.22}$$

The costs are a weighted average of the costs over one's life weighted by the probability of surviving. The expected costs for a population suffering complications, C, will be determined by a weighted average of these expected costs to an individual, $C(a)$, and then weighted by the proportion of patients that suffer the complications at each year, $P(a)$:

$$\mathbf{E}(C) = \sum_{a} C(a)P(a) \tag{2.23}$$

To estimate the cost of complications, $C(a, a\prime)$ the method developed in Zhang et al., 2017 is used to estimate the costs attributable to hypertension. The authors used a two-step procedure to estimate total expenditures in the US. In the first step, a logit model was used to estimate the probability of positive expenditures. The second step fit a gamma distribution to the total expenditure of those patients with positive expenditures using a Generalized Linear Model (GLM) procedure. The two-step procedure was employed because the Zhang et alia (2018) wanted to extrapolate their results to the entire population; thus, they needed the proportion of US residents that had positive expenditure. Since the current study is only interested in the cost/expenditures that can be attributed to specific complications of hypertension, the first step is omitted from its method. Only Zhang et alia (2018)'s second step of fitting the GLM was needed.

This second step (Zhang et al., 2017) fit a gamma distribution in which the log of the

mean was modeled with a linear combination of a series of control vectors. Because ten years of data were used in the current study, costs/expenditures are first adjusted into 2016 dollars using the consumer price index for healthcare services from the Federal Reserve bank of St. Louis' data repository (U.S. Bureau of Labor Statistics). Total adjusted expenditures, ADJTOTEXP, are assumed to follow a Gamma distribution:

$$ADJTOTEXP_i \sim Gamma(\alpha, \beta_i) \tag{2.24}$$

For Gamma distributions, the mean is the product of the two parameters, but convention in Gamma GLMs is to hold the scale parameter, α, constant. Thereby, the shape parameter, β, allows for individual heterogeneity. The expected value is connected to the linear predictors via a link function. Zhang et alia (2018) used a log-link function for the Gamma distribution, as is followed in the current study. Thus,

$$\mathbf{E}(ADJTOTEXP_i) = \alpha\beta_{=}g^{-1}(X_i\delta) = e^{X_i\delta} \tag{2.25}$$

X_i contain control variables. In this study, a Bayesian estimation procedure was used. Thus, $normal(0, 1)$ priors were given to the linear coefficients, δ, and an $exponential(1)$ prior for the scale parameter, α.

Under the procedure outlined above, three different models are fit for the current study. For the first model, total adjusted expenditures from all patients are used. The predictive posterior from this model is interpreted as the distribution of baseline healthcare expenditures. The second model is fit using those patients that suffered complications in the same year that they were surveyed. The resulting predictive posterior from this second model is interpreted as the distribution of all expenditures realized by people who had a complication in the year they were surveyed, or "this year". In the Zhang et al., 2017 study, the difference between the average predicted costs from these two models was interpreted as the average expenditures attributable to the complication, $C(a, a)$. Their interpretation is applied in

the current study. In addition, the use of Bayesian estimation techniques in the current study allows the calculation of the convolution of the two predictive posteriors, which, when restricted to positive values, can be interpreted as the distribution of these attributable costs.

For the third model the same method is applied. This time, the gamma-GLM is fit to those respondents that suffered a complication before "this year". The resulting predictive posterior, conditional on age, is interpreted as the distribution of expenditures after the year the complications was suffered. The difference with predictions from the first model are interpreted as the costs attributable to the complication suffered in the years after "this year", $C(a, a\prime)$. Again, the appropriate convolution yields the distribution of these expenditures.

Benefits were captured by a simple binary coding; a 0 indicated that a complication was suffered and a 1 indicated no complication was suffered. This entailed that for an individual patient, the benefits were calculated as the probability of suffering no complication over the tenure they received primary care.

2.3 Results

2.3.1 Using the first strategy

Following the first methodological strategy laid out in the previous sections, the MEPS data is used to fit a Weibull censored survival model with the response variable being the age when a respondent was first diagnosed with one of the complications under consideration (stroke, heart attack, coronary heart disease). The predictors are constrained by the data available in the ACHIS data set; these are race, sex, smoking, hypertension, and diabetes. Race is interacted with the other three variables in order to control for disparities in health outcomes between the races. The fitted model is then applied to the ACHIS data, from which the probability of complications while under observation are obtained. These are fed into the Poisson-Binomial model to obtain the expected number of complications, a highest density interval, and the probability of the observed occurrences. The difference between observed

and expected occurrences are then multiplied by the average cost of the complication, with this average taken over all ages; the cost values can be seen in 2.1. The expected cost savings due to the NLC can be seen in 2.2.

From the table, we can see that the number of complications the NLC's patients experienced were all very near the lower boundary of a 95% highest density interval (HDI). Probabilities for no complications would be sufficient in two cases to reject the null hypothesis (one tailed). Costs saved, $3.7 million, represents costs that were not realized by the NLC patients because that population had no complications. If the population had visited primary care providers that operated under the standard physician lead delivery method, they would have been expected to face that $3.7 million in costs.

Table 2.1: Expected lifetime expenditures attributable to Stroke, Heart Attack, and Coronary Heart Disease.

Age	Stroke	MI	CHD	Age	Stroke	MI	CHD	Age	Stroke	MI	CHD
18	$456,041	$560,790	$490,731	39	$381,553	$400,099	$367,592	60	$215,215	$212,855	$205,057
19	$454,210	$553,867	$485,744	40	$375,901	$392,051	$360,799	61	$204,901	$204,440	$196,303
20	$452,772	$547,385	$480,969	41	$369,413	$382,918	$352,879	62	$195,685	$194,477	$188,770
21	$451,126	$540,587	$475,708	42	$363,459	$374,466	$346,414	63	$185,920	$185,953	$180,701
22	$448,689	$534,542	$470,221	43	$356,467	$364,906	$339,041	64	$176,451	$176,295	$173,571
23	$446,158	$527,380	$464,644	44	$349,036	$356,254	$331,977	65	$165,852	$167,828	$163,835
24	$443,162	$520,028	$459,3451	45	$341,207	$347,973	$325,112	66	$155,704	$158,491	$155,673
25	$441,353	$512,217	$454,079	46	$334,724	$337,769	$318,573	67	$147,448	$149,982	$147,849
26	$438,477	$505,107	$448,595	47	$328,416	$329,538	$310,742	68	$137,179	$141,794	$138,435
27	$435,436	$497,528	$442,663	48	$320,276	$321,228	$302,741	69	$128,545	$133,684	$131,049
28	$431,860	$489,939	$437,302	49	$311,388	$312,148	$294,842	70	$120,762	$124,700	$121,756
29	$427,999	$482,136	$431,076	50	$303,687	$304,539	$287,417	71	$109,866	$116,150	$113,527
30	$424,771	$474,402	$425,324	51	$295,562	$294,703	$278,671	72	$101,367	$106,730	$105,506
31	$421,367	$466,925	$419,487	52	$287,848	$286,322	$271,905	73	$91,808	$99,489	$95,682
32	$416,352	$457,972	$413,638	53	$278,193	$277,418	$262,853	74	$83,584	$89,235	$87,112
33	$413,570	$451,428	$407,052	54	$269,435	$267,945	$255,238	75	$74,712	$82,145	$79,891
34	$408,394	$442,539	$400,553	55	$261,326	$259,108	$247,802	76	$67,005	$73,779	$71,087
35	$403,650	$434,658	$394,261	56	$251,563	$251,035	$239,104	77	$58,108	$66,036	$64,643
36	$397,827	$425,845	$387,651	57	$243,161	$240,869	$230,790	78	$50,457	$58,621	$56,659
37	$393,187	$417,236	$380,792	58	$232,448	$231,462	$222,593	79	$42,215	$51,197	$49,671
38	$387,561	$408,460	$373,875	59	$224,180	$221,966	$213,892	80	$35,199	$43,048	$41,320

These values are the weighted sum of the yearly average cost experienced by someone who suffers the noted complication at the age marked. Weights are given by the probability of surviving to realize the cost.

Table 2.2: Expected lifetime expenditures attributable to Stroke, Heart Attack, and Coronary Heart Disease.

Selected Hypertension Complication	Number of Complications Observed in NLC Study	Expected Number of Complications[a] (95% HDI)	Probability of Observing No (0) Complications[a]	Average Expected Lifetime Cost per Patient of Complication[b]	Costs Saved during NLC Study Period[c]
Stroke	0	3.66 (0, 8)	0.025	$349,812	$1,307,495
Heart Attack	0	2.95 (0, 8)	0.052	$378,088	$1,141,679
Coronary Heart Disease	0	3.79 (1, 8)	0.022	$345,646	$1,338, 852
			Total		$3,705,673

[(a)] Calculated under the null hypothesis that the two delivery methods offer statistically equivalent outcomes.

[(b)] The average is taken over the age distribution at last follow up of the NLC patient population. The expectation is taken based on lifetime costs weighted by probability of surviving so many years after the complication.

[(c)] Expected number of complications multiplied by the average expected lifetime cost per patient.

2.3.2 Using the second strategy

Under the second strategy, the survival curves for both the control and experimental groups are estimated. The survival curves for blacks and whites for each complication are presented in 2.2. We see that the 95% highest density intervals (dotted lines) are tightly packed around the mean survival curves. This is consistent with the discussion in previous sections. Specific coefficient values might not be particularly trustworthy, but the survival curve predictions do exhibit precise estimation.

Estimated survival curves for the non-hypertensive patients show less divergence between the two patient populations. This is not unexpected, but that there is a difference may suggest that the NLC's curve is generally offering non-inferior, if not superior, care to the physician lead care. For the hypertensive patients, on the other hand, we see a large divergence between the two groups.

From the survival curves the probability of suffering a complication at each age is calculated. These are then multiplied by the expected costs, in 2.1, of each complication. Summing these then gives us average costs to a random individual due to hypertension and

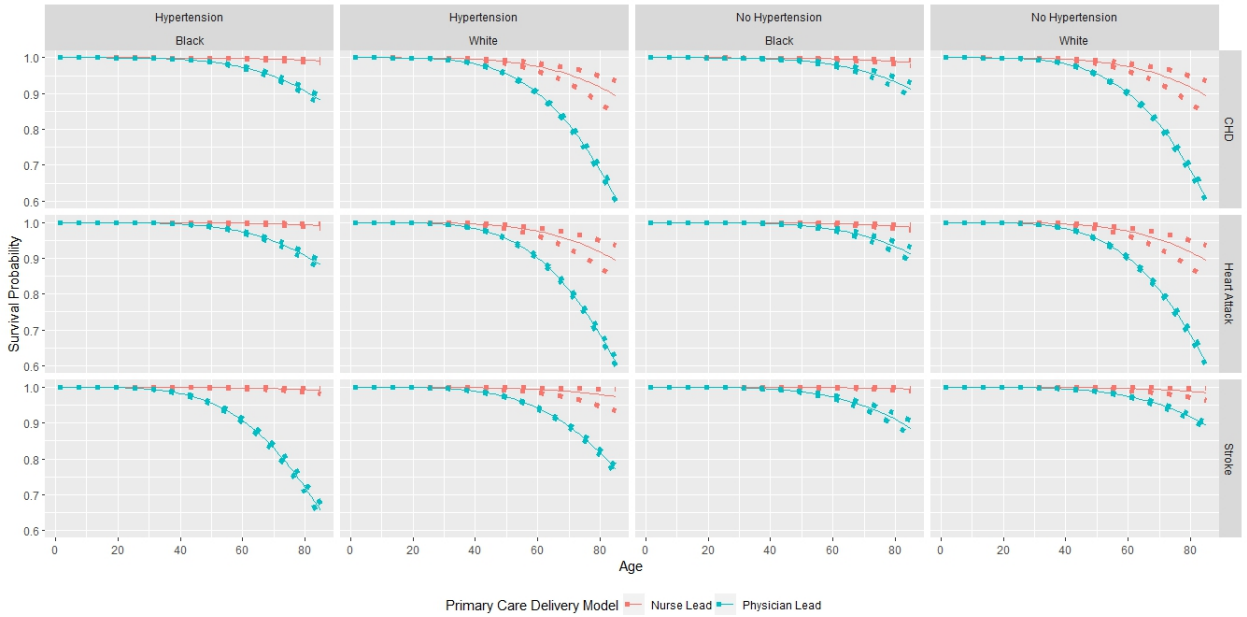

Figure 2.2: Probability of suffering coronary heart disease (CHD), Heart Attack, or Stroke by age and care delivery method for Black individuals.solid lines: mean; dotted lines: 95% highes density interval.

these three complications. Following a similar procedure for benefits results in the expected complication free number of years for a random individual. Taking the ratio of expected costs and expected complication free number of years gives us the cost per complication free year. 2.3 presents the expected cost to benefit ration stratified by race (white and black, only), hypertension status, and care delivery method. We see that the nurse lead care is consistently a fraction of the physician lead care. These cost/benefits ratios are exclusively reflecting the differences in the survival curves in 2.2.

2.4 Study limitations

In assessing any study's limitations, it is useful to keep in mind what would be the ideal data collection and analysis one would like to perform to answer the question of interest. In the case of this study the ideal would be:

- A random sample of the population, stratified by relevant demographic groups, is

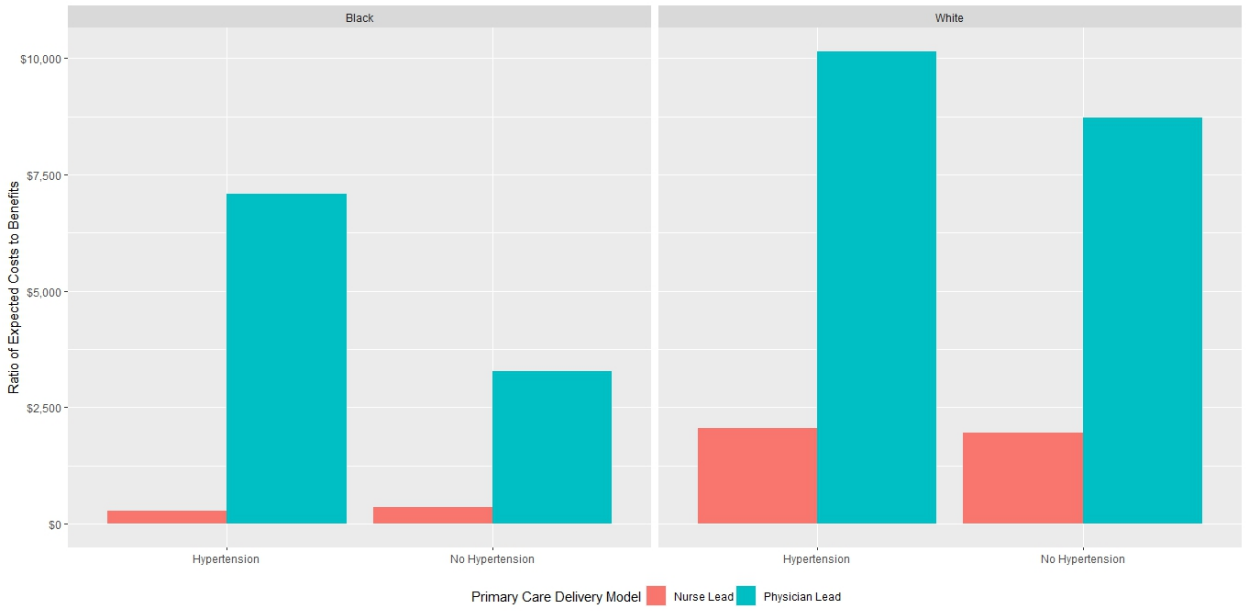

Figure 2.3: Expected Costs per expected complication free year by race and primary delivery care model

randomly assigned to receive primary care from either a nurse led or physician lead facility.

- Patients are followed over a significant period, with lose of follow up only due to death.
- We have complete health records for all patients.
- We have complete receipts of expenditures for all patients.

Physical and ethical constraints make the ideal unfeasible. However, from an inferential perspective, the ideal is the best way to assess the relative cost savings of nurse lead primary care facilities. This study's method can be compared to the ideal point by point to see where deviations from the ideal limits the conclusions that can be drawn.

In the ideal study, patients are randomly selected from the population and randomly assigned to the two groups. In this study, the MEPS respondents are selected randomly, but the ACHIS patients cannot be considered a random sample. The ACHIS data comes from a clinical setting so suffers from selection bias issues, thus is not necessarily representative

of the entire population. So, it might be objected that these results cannot be generalized beyond a population that is similar in make up to the ACHIS patient population.

The ideal study would follow all patients for a uniform amount of time. In neither the ACHIS nor the MEPS data is this followed. The MEPS data is survey responses. Thus, the data may suffer from recollection issues, but also survivor biases. The ACHIS data does better on this regard in that it follows patients over time, but follow up time is, effectively, random, from the researcher perspective. Thus, it might be objected that ACHIS patients may have only visited the nursing center while they were not sick. This cannot be completely ruled out given the data available, even if unlikely for all patients.

The ideal study would have complete medical records and receipts of expenditures for each patient. For neither the MEPS respondents nor the ACHIS patients do we have complete medical records or receipts of healthcare expenditures. For the MEPS respondents, we have more extensive records of expenditures, but we have minimal health information, recollection of the initial age of diagnosis of hypertension or complication. For the ACHIS data, we have no expenditure receipts but clinical records from primary care visits. This entails there is unaccountable uncertainty in the survival curve estimates due to the recollection errors in the MEPS respondents. We also had to assume that expenditure receipts for the MEPS respondents and ACHIS patients would be distributed identically, which may not be true.

The estimates of this study that resulted from the Bayesian survival model must be used carefully. The coefficient estimates are not reliable, which will be reflected in the spread of their posterior distributions, but the posterior survival curves will at worst be lower bounds for the true survival curves. These qualifications do not interfere with using the resulting estimated survival curves in calculating expected costs and benefits, though, we may have to interpret them as lower bounds to the actual costs and benefits realized by a high-risk population that visit NLC primary care facilities.

2.5 Conclusions

This study compares the cost-benefit ratios of primary care for hypertensive patients between a NLC primary care delivery model and physician-led primary care delivery model and developed a replicable method for such comparisons. Each aim will be discussed below.

The cost-benefit ratio comparison completed for this study, demonstrates that hypertensive patients in a high-risk population see significant cost reductions and increased years free of hypertensive complications when visiting a NLC model facility for primary care. This study presents two statistical strategies for comparing the cost-benefits of the NLC primary care delivery model to the physician-led primary care delivery. First, the cost of expected hypertensive complications that the NLC patient population did not experience is calculated, under the assumption that there was no difference between the two delivery models. Second, weighted costs and benefits are calculated for both a population that visited a NLC and physician-led primary care facilities. Both strategies indicate that there were significant cost savings for hypertensive primary care and an increase in complication free years for the hypertensive patients that visited the NLC facility. With health care expenditures taking up 17.7% of Gross Domestic Product (GDP) in the U.S. ($3.8 trillion in 2019) (CMS), government policy has been focused on finding ways to control costs while ensuring quality is maintained or improved. This study shows that NLC primary health care delivery model can accomplish both of those goals.

The second outcome of this study was the development and testing of a replicable economic cost-benefit analysis method using existing, accessible and easily retrievable data. This addressed a gap in the current literature on cost-benefit analysis for NLC primary care delivery model. In order to calculate the expected costs and benefits for the NLC patient population a novel method of estimating survival curves is implemented. The NLC patient population did not experience any complications during the time the population was followed. While this demonstrated evidence of good clinical outcomes, plausible survival curves, often used for cost-benefit analysis, would require complication event data be available. However,

using Bayesian estimation techniques, this study is able to estimate survival curves for the NLC patient population. To do this a null hypothesis is assumed that the two delivery methods were equivalent in terms of the hypertensive complications their patients experienced. There were no previous studies found that leveraged this advantage of Bayesian estimation. The methodology needs to be further tested to be confident in its utility. Having such an easily replicable cost-benefit analysis model will allow further evaluation and comparisons of primary care delivery models which are essential as the US strives to contain costs and improve patient outcomes for the US healthcare system.

Further exploration of NLC across the country could support economic rationale for investment in major health care changes. Such data is needed as the US contemplates decisions of major health care reforms needed to support elimination of health disparities and provide equitable care while controlling healthcare cost inflation. This study provides a starting point for further economic analyses. It shows that a NLC primary care delivery model provids both cost savings and improved outcomes for hypertensive patients.

References for chapter 1

Aitchison, J. (1986). *The statistical analysis of compositional data*. Chapman; Hall.

Akerlof, G. A., & Shiller, R. J. (2009). *Animal spirits*. Princeton University Press.

Allenby, G. M., Arora, N., & Ginter, J. L. (1998). On the heterogeneity of demand. *The American Economic ReviewJournal of Marketing Research*, *35*, 384–389.

Alston, J. M., Foster, K. A., & Gree, R. D. (1994). Estimating elasticities with the linear approximate almost ideal demand system; some monte carlo results. *The Review of Economics and Statistics*, *76*(2), 351–356.

Ayres, I., & Siegelman, P. (1995). Race and gender discrimination in bargaining for a new car. *The American Economic Review*, *85*(3), 304–321. http://www.jstor.org/stable/2118176

Banks, J., Blundell, R., & Lewbel, A. (1997). Quadratic engel curves and consumer demand. *The REview of Economics and Statistics*, *79*(4), 527–539.

Barigozzi, M., Alessi, L., Capasso, M., & Fagiolo, G. (2012). The distribution of household consumption-expenditure budget shares. *Structural Change and Economic Dynamics*, *23*, 69–91.

Becker, G. (1962). Irrational behavior and economic theory. *The Journal of Political Economy*, *70*(1), 1–13.

Benkard, C. L., & Bajari, P. (2005). Demand estimation with heterogenous consumers and unobserved product characterstics: A hedonic approach. *Journal of Political Economy*, *115*(6), 1239–1276.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–890.

Berry, S., Levinsohn, J., & Pakes, A. (2004). Differentiated products demand system from a combination of micro and macro data:the new car market. *Econometrica*, *112*(1), 68–105.

Berry, S. T., & Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, *82*(5), 1749–1797. https://doi.org/https://doi.org/10.3982/ECTA9027

Blundell, R., Horowitz, J., & Parey, M. (2017). Nonparametric Estimation of a Nonseparable Demand Function under the Slutsky Inequality Restriction. *The Review of Economics and Statistics*, *99*(2), 291–304. https://doi.org/10.1162/REST_a_00636

Blundell, R., & Stoker, T. M. (2005). Heterogeneity and aggregation. *yJournal of Economic Literature*, *43*(2), 347–391.

Browning, M., & Chiappori, P. A. (1998). Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica*, *66*(6), 1241–1278. http://www.jstor.org/stable/2999616

Browning, M., & Carro, J. (2007). Heterogeneity and microeconometrics modeling. *Econometric Society Monographs*, *43*, 47.

Buse, A. (1994). Evaluating the linearized almost ideal demand system. *American Journal of Agricultural Economics*, *76*(4), 781–793.

Chipman, J. S., & Moore, J. C. (1976). The scope of consumer's surplus arguments. In M. Tang, N. Georgescu-Roegen, A. Tang, F. Westfield, & J. Worley (Eds.), *Evolution, welfare, and time in economics: Essays in honor of nicholas georgescu-roegen* (pp. 69–125). Lexington Books.

Deaton, A., & Muellbauer, J. (1980a). An almost ideal demand system. *The American Economic Review*, *70*(3), 611-618312–326.

Deaton, A., & Muellbauer, J. (1980b). *Economics and consumer behavior*. Cambridge University Press.

Ferrari, S. L., & Cribari-Neto, F. (2004). Beta regression for modeling rates and proportions. *Journal of Applied Statistics*, *31*(7), 799–815.

Feuerverger, A. (1989). On the empirical saddlepoint approximation. *Biometrika*, *76*(3), 457–464. http://www.jstor.org/stable/2336112

Figueroa-Zuniga, J. I., Arellano-Valle, R. B., & Ferrari, S. L. (2013). Mixed beta regression: A bayesian perspective. *Computational Statistics & Data Analysis*, *61*, 137–147.

Filzmoser, P., Hron, K., & Templ, M. (2018). *Applied compositional data analysis: With worked examples in r*. Springer International Publishing. https://books.google.com/books?id=cw92DwAAQBAJ

Fry, J. M., Fry, T., & McLaren, K. (2000). Compositional data analysis and zeros in micro data. *Applied Economics*, *32*(8), 953–959. https://doi.org/10.1080/000368400322002

Fry, J. M., Fry, T. R., & McLaren, K. R. (1996). The stochastic specification of demand share equations: Restricting budget shares to the unit simplex. *Journal of Econometrics*, *73*, 377–385.

Fry, T. (2011). Applications to economics. In V. Pawlowsky-Glahn & A. Buccianti (Eds.), *Compositional data analysis: Theory and applications*. John Wiley & Sons Ltd.

Hartley, J. E. (1997). *The representative agent in macroeconomics*. Routledge.

Hausman, J. A. (1981). Exact consumers' surplus and deadweight loss. *The American Economic REview*, *71*, 662–676.

Heckman, J. J., & Willis, R. J. (1977). A beta-logistic model for the analysis of sequential labor force participation by married women. *The Journal of Political Economy*, *85*(1), 27–58.

Hildenbrand, W. (1994). *Market demand: Theory and emprical evidence*. Princeton University Press.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions, volume 2*. John Wiley & Sons Ltd.

Lewbel, A. (2001). Demand system with and without errors. *The American Economic Review*, *91*(3), 611–618.

Lyssiotou, P., Pashardes, P., & Stengos, T. (1999). Preference heterogeneity and the rank of demand systems. *Journal of Business and Economic Statistics*, *17*, 248–252.

Maier, M. J. (2014). *Dirichletreg: Dirichlet regression for compositional data in r* (Research Report Series / Department of Statistics and Mathematics No. 125). WU Vienna University of Economics and Business. Vienna. http://epub.wu.ac.at/4077/

McCullagh, P., & Nelder, J. (1989). *Generalized linear models*. CRC Press. https://books.google.com/books?id=SPSDDwAAQBAJ

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, *3*, 303–328.

Monti, G., Figueras, G., Pawlowsky-Glahn, V., & Egozcue, J. J. (2011). The shifted-scaled dirichlet distribution in the simplex.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer New York. https://books.google.com/books?id=yexFAAAAQBAJ

Ng, K. W., Tian, G.-L., & Tang, M.-L. (2011). *Dirichlet and related distributions: Theory, methods, and applications*. Wiley.

Okrent, A. M., & Alston, J. M. (2012). The demand for disaggregated food away from home and food at home products in the united states.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Pearson, K. (1897). Mathematical contributions to the theory of evolution; on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London*, *60*(359-367), 489–498. https://doi.org/10.1098/rspl.1896.0076

Poi, B. P. (2012). Easy demand-system estimation with quaids. *The Stata Journal*, *12*, 433–446.

Sennhenn-Reulen, H. (2018). Bayesian regression for a dirichlet distributed response using stan. *arXiv preprint arXiv:1808.06399*.

Smith, B., & Rayens, W. (2002). Multivariate positive dependence on the simplex. *Statistics*, *36*(1), 75. https://ezproxy.lib.uwm.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=aph&AN=10955563&site=ehost-live&scope=site

Stigler, G. J. (1961). The economics of information. *The Journal of Political Economy*, *69*(3), 213–225.

Tang, M., Georgescu-Roegen, N., Tang, A., Westfield, F., & Worley, J. (1976). *Evolution, welfare, and time in economics: Essays in honor of nicholas georgescu-roegen*. Lexington Books. https://books.google.com/books?id=Awu8AAAAIAAJ

U.S. Census Bureau. (n.d.). Consumer expenditure survey.

U.S. Department of Transportation, Federal Highway Administration. (n.d.). *2001 national household travel survey*. https://nhts.ornl.gov

van der Merwe, S. (2018). A method for Bayesian regression modelling of composition data. *ArXiv e-prints*.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, *27*, 1413–1432.

Woodland, A. (1979). Stochastic specification and the estimation of share equations. *Journal of Econometrics*, *10*, 361–383.

References for chapter 2

Arias, E., & Xu, J. (2019). United states life tables, 2017. https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_07-508.pdf

Bicki, A., Silva, A., Joseph, V., Handoko, R., Rico, S., Burns, J., & De Groot, A. S. (2013). A nurse-run walk-in clinic: Cost-effective alternative to non-urgent emergency department use by the uninsured. *Journal of Community Health*.

by Commodity: Health Care Services [WPU51], P. P. I. (n.d.). U.S. Bureau of Labor Statistics. https://fred.stlouisfed.org/series/WPU51

Data, N. H. E. (n.d.). Centers for Medicare and Medicaid Services (CMS). https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical

DM, B., & ad Whittington J., N. T. (2008). The triple aim: Care, health, and cost. *Health Aff (Millwood)*.

Dobson, A. J., & Barnett, A. G. (2018). *An introduction to generalized linear modelsn*.

Esperate, M., Hanson-Turton, T., Richardson, M., Debisette, A., & Rupinta, C. (2012). Nurse-managed health centers: Safety-net care through advanced nursing practice. *Journal of the American Academy of Nurse Practitioners*.

Fernandez, M., & Williams, S. (2010). Closed-form expression for the poisson-binomial probability density function. *IEEE Transactions on Aerospace and Electronic Systems*.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*.

Horrocks, S., Anderson, E., & Salisbury, C. (2002). Systematic review of whether nurse practitioners working in primary care can provide equivalent care to doctors. *BMJ*, *324*(7341), 819–823.

Ibrahim, J. G., Chen, M.-H., & Sinha, D. (2001). *Bayesian survival analysis*.

J., H., B., Z., & MJ., B. (2014). Evidence based characteristics of nurse-managed health centers for quality and outcomes. *Nurs Outlook*.

JM, P., C, T., B, P., & R., B. (2011). Comparison of nurse managed health centers with federally qualified health centers as safety net providers. *Policy Polit Nurs Practice*.

Joo, J. Y., & Huber, D. L. (2013). An integrative review of nurse-led community-based case management effectiveness. *International Nursing Review*.

KS., M. (2005). *The omaha system: A key to practice, documentation, and information management*.

Laurant, M., van der Biezen, M., Wijers, N., Watananirun, K., Kontopantelis, E., & van Vught, A. J. (2018). Nurses as substitutes for doctors in primary care. *Cochrane Database Syst Rev*, *7*, CD001271.

Lundeen, S. (1993). Comprehensive, collaborative, coordinated, community-based care: A community nursing center model. *Community Health*.

Lundeen, S. (1999). An alternative paradigm for promoting health in communities: The lundeen community nursing center model. *Family & Community Health*.

Lundeen, S., & Friedbacher, B. (1994). The automated community health information system (achis): A relational database application of the omaha system in community nursing centers. *Nursing informatics*, 23–31.

M, S., & K., N. (2012). *A guide to measuring the triple aim: Population health, experience of care, and per capita cost* (tech. rep.). Institute for Healthcare Improvment.

R, S., JM, M., & L., L. (2015). The quadruple aim: Care, health, cost and meaning in work. *BMJ Qual Saf*.

Stan Development Team. (2020a). *Rstan: The r interface to stan*. https://mc-stan.org

Stan Development Team. (2020b). *Stan modeling language users guide and reference manual*. https://mc-stan.org

T, B., & C., S. (2014). From triple to quadruple aim: Care of the patient requires care of the provider. *Ann Fam Med*.

TA, L., D, G., & P., R. (2011). Estimating the economic burden of racial health inequalities in the united states. *Int J Health Serv*.

The triple aim: Care, health, and cost. (n.d.). Institute for Healthcare Improvement. http://www.ihi.org/Engage/Initiatives/TripleAim/Pages/default.aspx

Virani, S. S., Alonso, A., Aparicio, H. J., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, A. P., Chamberlain, A. M., Cheng, S., Delling, F. N., Elkind, M. S., Evenson, K. R., Ferguson, J. F., Gupta, D. K., Khan, S. S., Kissela, B. M., Knutson, K. L., Lee, C. D., Lewis, T. T., ... Tsao, C. W. (2021). Heart disease and stroke statistics—2021 update. *Circulation*, *143*(8), e254–e743. https://doi.org/10.1161/CIR.0000000000000950

Wolf, P., D'Agostino, R., Belanger, A., & Kannel, W. (1991). Probability of stroke: A risk profile from the framingham study. *Stroke*.

Zhang, D., Wang, G., Zhang, P., Fang, J., & Ayala, C. (2017). Medical Expenditures Associated With Hypertension in the U.S., 2000-2013. *Am J Prev Med*, *53*(6S2), S164–S171.

Appendix A

Properties of the Cobb-Douglas utility function

Consider three models of the consumer's problem

	A.1	A.2	A.3
Utility	$\max_{\mathbf{c}} \prod_i^d c_i^{\alpha_i}$	$\max_{\mathbf{c}} \prod_i^d c_i^{\alpha_i}$	$\max_{\mathbf{c}} \prod_i^d c_i^{\alpha_i}$
Budget	$\langle \mathbf{p}, \mathbf{c} \rangle \leq m$	$\langle \mathbf{p}, \mathbf{c} \rangle \leq m$	$\langle \mathbf{p}, \mathbf{c} \rangle \leq m$
	-	$\boldsymbol{\alpha} = \boldsymbol{\alpha}(\mathbf{X})$	$\boldsymbol{\alpha} = e^{\langle \mathbf{X}, \boldsymbol{\beta} \rangle}$

The following table presents the elasticities of these three models.

Elasticities	Income	Own Price	Cross Price
A.1	1	-1	0
A.2	$1 + m\frac{\partial \alpha_i(\mathbf{X})/\partial m}{\alpha_i(\mathbf{X})}$	$p_i \frac{\partial \alpha_i(\mathbf{X})/\partial p_i}{\alpha_i(\mathbf{X})} - 1$	$\frac{p_j}{\alpha_i(\mathbf{X})} \frac{\partial \alpha_i(\mathbf{X})}{\partial p_j}$
A.3	$1 + \beta_m m$	$\beta_i p_i - 1$	$\beta_j p_j$

Thus, despite the Cobb-Douglas form, we can derive a great deal of different elasticity behaviors by assuming different forms for $\boldsymbol{\alpha}(\boldsymbol{X})$.

Appendix B

A few notes on the Dirichlet distribution

The Dirichlet family of distributions is used to model compositional data, which is concerned with the proportions of constituent parts, i.e. for

$$\begin{aligned} \mathbf{X} &\sim Dirichlet(\mathbf{h}_d) \\ \mathbf{X} \in \mathbf{S}^{d-1}(k) &= \{x : \forall i x_i \geq 0, \sum_i^d x_i = k\} \end{aligned} \tag{B.1}$$

Thus, the support of the Dirichlet distribution is the simplex. The standard form of the Dirichlet has support on the unit simplex, $k = 1$, but there are many generalizations, see Ng et al., 2011. The density of the Dirichlet distribution is

$$f(\mathbf{x}) = \frac{\Gamma(h_+)}{\prod_i^d \Gamma(h_i)} \frac{\prod_i^d x_i^{h_i - 1}}{k^{h_+}}, \ for \ \mathbf{x} \in S^{d-1}; 0 \ otherwise \tag{B.2}$$

where $h_+ = \sum_i^d h_i$. When the h_i are all unity, this becomes the uniform distribution on the simplex. When $n = 2$, this reduces to the standard Beta distribution. In fact, due to some nice properties of the Dirichlet, it can be shown that the marginals all follow a Beta regression:

$$\alpha \sim Dirichlet(\mathbf{h}) \Rightarrow \ \alpha_i \sim Beta(h_i, h_+ - h_i).$$

The simplest of the Dirichlet distributions has a vector of parameters, $\mathbf{h}$, with which the functionals of the distribution can be calculated. For example, the means of the elements are given by

$$\mathbf{E}[X_i] = \frac{h_i}{h_+}. \tag{B.3}$$

Variances can be calculated similarly:

$$Var[X_i] = \frac{h_i(h_+ - h_i)}{h_+^2(h_+ + 1)} \tag{B.4}$$

Likewise, the covariance between any two elements is calculated by

$$Cov(X_i, X_j) = \frac{-h_i h_j}{h_+^2(h_+ + 1)}. \tag{B.5}$$

However, it should be noted that the covariance matrix of a Dirichlet random vector does not have the same interpretation as in other settings. Because of the adding up condition, these covariances will always have a negative bias, as indicated by the above equation. This makes it difficult to interpret the covariance matrix in a standard way or in a way economically meaningful. This has often been taken as a reason not to use the Dirichlet for modeling demand systems. However, this is a misunderstanding of how the geometry of the simplex affects the covariances of random vectors in the simplex. Smith and Rayens, 2002 prove that no distribution on the simplex can have positive covariances in the way most people intuit demand systems should. The intuition can be banished by considering a single composition and its inverse. Let

$$X_1 \sim F \tag{B.6}$$

$$X_2 = 1 - X_1 \tag{B.7}$$

The covariance of these two variables will always be negative, no matter what distribution

describes either one of them.

$$Cov(X_1, X_2) = Cov(X_1, 1 - X_1) = -Var(X_1) \tag{B.8}$$

The details get a bit more complicated when we increase the dimensions, but the intuitions still hold. the covariances will tend negative because of the geometry of the simplex, irrespective of the distribution assumed.

This topic deserves a bit more discussion. The negative covariances are likely taken to be a reason to reject the Dirichlet distribution because this does not seem to cohere with the existence of complements. We recall that complements are products or product categories that tend to be purchased together. Often this is defined in terms of the cross-price elasticities, complements are products that have negative cross-price elasticities. A canonical example is hot dogs and hot dog buns. When the price of hot dogs increases, consumers respond by purchasing fewer hot dogs but also fewer hot dog buns. What might be missed here is the recognition that complements is always a three variable relation. It is not always the same third variable. Suppose that both prices remain constant and a consumer buys more hot dogs, we might expect that she would buy more hot dog buns as well. However, to do so one of two things must be true: 1) total expenditure must increase, or 2) our consumer must decrease the amount she spends on some third product. We can never get a complement relation that only involves two variables, the quantity of or expenditure on two complementary products. The covariance operator cannot capture this three variable relation, especially in its theoretical formation. The covariance effectively treats all the other variables as if they were constant at their conditional means. This makes it very obvious why the covariances must be negative. If everything else is held constant (all prices, total expenditure, quantities of all other products), then if I increase my purchases of hot dogs I must necessarily decrease my purchases of hot dogs. This is what the covariance is capturing. No distribution on the simplex will ever get away from this geometric fact about the simplex.

On a somewhat contradictory path, Maier, 2014 reports that the negative covariances does not necessarily hold when predictors are used to model the parameters of the Dirichlet distribution, especially sample covariances from simulations. This fits in with the discussion above, though. In our case, estimating budget shares, we may be inclined to use income/total expenditure and prices as predictors. When we do this, taking covariances involves budget shares that are varying along side incomes and prices that are potentially varying. Thus, our sample covariance is capturing a lot of other relationships that are not captured in our theoretical covariance. To my knowledge, it has not been answered to everyone satisfaction why/how this happens. It is likely that the geometry of the simplex is responsible.

Another reason that has been offered against the Dirichlet claims that assuming the Dirichlet for budget shares entails that the expenditure levels are independent Gamma distributed (Barigozzi et al., 2012). This position is founded upon, what I will call, a scholastically equivalent sampling method. Consider, n independent Gamma distributed random variables, with the same scale parameter,

$$Y_i \sim Gamma(\alpha_i, \theta) \tag{B.9}$$

The sum of these n random variables will also be Gamma,

$$Y_+ = \sum Y_i \sim Gamma(\sum \alpha_i, \theta) \tag{B.10}$$

If we take the ratio of the n Y_i random variables with Y_+, the distribution of the resulting vector of proportions will be Dirichlet, i.e.

$$\left(\frac{Y_i}{Y_+}\right)_{i=1}^{n} \sim Dir(\boldsymbol{\alpha}) \tag{B.11}$$

This fact is often used to sample from a Dirichlet distribution. However, the objection to using the Dirichlet requires the entailment to go in the other direction. As is, the direction

of entailment is

$$Y_i \sim Gamma(\alpha_i, \theta) \ \& \ Y_i \perp Y_j \ for \ i \neq j \Rightarrow \left(\frac{Y_i}{Y_+}\right)_{i=1}^{n} \sim Dir(\boldsymbol{\alpha}) \tag{B.12}$$

The problem with this position is that direction that is needed to justify the claim is not true. One need only refer to Sklar's theorem to show that it cannot be true. Following Nelsen, 2007, Sklar's theorem says that for any joint distribution H with marginals F_i (not necessarily from the same family), there is a copula C such that

$$H(\mathbf{X}) = C(F_1(X_1), ..., F_n(X_n)) \tag{B.13}$$

if the F_i are continuous, then the copula is unique. This entails that one can, theoretically, use any marginals with a unique copula to get out a Dirichlet distributed random vector. It's likely that there will not be an analytic expression for many of these copulas, but they nonetheless exist.The above Gamma is a special case of Sklar's theorem with H being the Dirichlet distribution, the F_i being ratios of Gammas, and C being the copula of independence. There is no necessity here. One can use any F_is one would like as long as one uses the appropriate copula. Thus, the force of this objection to using the Dirichlet dissolves.

There are a number of generalizations of the Dirichlet distribution. In particular, the scaled Dirichlet distribution, Monti et al., 2011, will be of particular use. This form of the scaled Dirichlet is defined on a transformed simplex, which I call a weighted simplex:

$$S_w^{d-1}(k) = \{x : \forall i x_i \geq 0, \sum_i^d \beta_i x_i = k\} \tag{B.14}$$

where β_i are known parameters. A random vector $\mathbf{X} \in S_w^{d-1}(k)$ is a scaled Dirichlet random vector, $\mathbf{X} \sim SD(\boldsymbol{h}, \boldsymbol{\beta})$ with density

$$f(\mathbf{x}) = \frac{\Gamma(h_+)}{\prod_i^d \Gamma(h_i)} \frac{\prod_i^d \beta_i^{h_i} x_i^{h_i - 1}}{(\sum_i^d \beta_i x_i)^{h_+}} for \ \mathbf{x} \in S_w^{d-1}(k) \tag{B.15}$$

This is a non-standard formulation of the scaled Dirichlet, in particular the scaled Dirichlet is usually still defined on the simplex.

Appendix C

Concavity of the log-likelihood of the Dirichlet GLM model

The concavity of the log-likelihood is shown by proving the second derivatives are always less than or equal to 0. Recall the log-likelihood,

$$logL = \sum_{i}^{N} -log\left(B(\phi\boldsymbol{\mu_i})\right) + \sum_{i}^{n}\sum_{j}^{d}(\phi\mu_{ij} - 1)log(s_{ij}) \tag{C.1}$$

The first derivative of this is

$$\frac{\partial logL}{\partial \beta_{jk}} = \sum_{i}^{N} X_{ik}\mu_{ij}\phi\left(log(s_{ij}) - \psi(\phi\mu_{ij})\right) \tag{C.2}$$

The second derivative is

$$\frac{\partial^2 logL}{\partial \beta_{jk}^2} = \sum_{i}^{N} X_{ik}^2\mu_{ij}\phi\left(log(s_{ij}) - \psi(\phi\mu_{ij}) - \psi_1(\phi\mu_{ij})\mu_{ij}\right) \tag{C.3}$$

The sign of this is going to depend on the relative values of the polygamma functions, ψ and ψ_1. The trigamma function, ψ_1, is everywhere positive. Since $\phi\mu_{ij} \geq 0$, there is a relatively small range within which the two functions have different signs. When $\phi\mu_{ij} \geq 1.46...$, the

positive root of ψ, the two polygamma functions are positive and thus the second derivative of the loglikelihood is negative. When this condition is not satisfied, the sign of the second derivative depends intimately on the relative values of the polygamma functions and the mean, μ_{ij}. However, since $|\psi_1| > |\psi|$ between 0 and the positive root of ψ, the log-likelihood is concave in the β.

The derivative of the loglikelihood with respect to ϕ is a bit trickier. The second derivative is

$$\frac{\partial^2 logL}{\partial \phi^2} = N\psi_1(\phi) - \sum_i^N \sum_j^d \psi_1(\phi\mu_{ij})\mu_{ij}^2 \tag{C.4}$$

Again, ϕ and $\phi\mu_{ij}$ are positive and ψ_1 is a decreasing function in the positive reals. Thus, the sign of this one is ambiguous. For this to be positive, the following inequality must be satisfied,

$$N\psi_1(\phi) > \sum_i^N \sum_j^d \psi_1(\phi\mu_{ij})\mu_{ij}^2 \tag{C.5}$$

If we replace the μ_{ij} with their average value, $\frac{1}{d}$, then $\sum_j^d \mu_{ij}^2 = \frac{1}{d}$; the above inequality then reduces to

$$d\psi_1(\phi) > \psi_1(\frac{\phi}{d}) \tag{C.6}$$

This is never satisfied because of bounds for the polygamma function:

$$\frac{(m-1)!}{x^m} + \frac{m!}{2x^2} \leq (-1)^{m+1}\psi_m(x) \leq \frac{(m-1)!}{x^m} + \frac{m!}{x^2} \; for \; x > 0 \tag{C.7}$$

For $m = 1$, becomes

$$\frac{d}{2\phi^2} \leq d\psi_1(\phi) \leq \frac{d}{\phi^2} \leq \frac{d^2}{2\phi^2} \leq \psi_i(\frac{\phi}{d}) \leq \frac{d^2}{\phi^2}. \tag{C.8}$$

For $d \geq 2$, the inequality goes in the wrong direction. Thus, we conclude that the log-likelihood is concave.

Appendix D

Derivation of the Consumption Distribution

Starting from the assumptions of the model

$$\begin{aligned} &\max_{\mathbf{c}} \Pi_i^n c_i^{\alpha_i} \\ &\textit{subject to } \langle \mathbf{p}, \mathbf{c} \rangle \leq m \\ &\boldsymbol{\alpha} \sim \textit{Dirichlet}(\mathbf{X}) \end{aligned} \tag{D.1}$$

This leads to demand curves of the form

$$c_i = \frac{m}{p_i}\alpha_i \Rightarrow \alpha_i = \frac{c_i p_i}{m} \tag{D.2}$$

Thus, by the change of variables formula, the density of the consumption vectors is given by

$$f_{\mathbf{c}} = f_{\boldsymbol{\alpha}}\left(\frac{c_i p_i}{m}\right)\left|det\left(\frac{\partial \boldsymbol{\alpha}}{\partial \mathbf{c}}\right)\right| \tag{D.3}$$

The Jacobian of the consumption share curves is a diagonal matrix with $\frac{p_i}{m}$ along the diagonal. Thus, the Jacobian determinant is just the trace of this matrix, or

$$\left| det\left(\frac{\partial \boldsymbol{\alpha}}{\partial \mathbf{c}}\right)\right| = \prod_i \frac{p_i}{m} \tag{D.4}$$

Using this result and the density of $\frac{p_i}{m}$, we get

$$f_{\mathbf{c}} = \frac{\Gamma(h_+)}{\prod_i^d \Gamma(h_i)} \prod_i^d \left(\frac{c_i p_i}{m}_i\right)^{h_i - 1} \prod_i^d \frac{p_i}{m} \tag{D.5}$$

Associating terms, gives us

$$f_{\mathbf{c}|\mathbf{x}} = \frac{\Gamma(h_+)}{\prod_i^d \Gamma(h_i)} \frac{\prod_i^d p_i^{h_i} c_i^{h_i - 1}}{m^{h_+}} \tag{D.6}$$

This is the form of a scaled Dirichlet density function. However, its support is not on a standard simplex. The support of this density is

$$\mathbf{C} = \{\mathbf{c} : \forall i c_i \geq 0, \langle \mathbf{p}, \mathbf{c} \rangle = m\} \tag{D.7}$$

This is best described as a weighted simplex, where the weights are given by the prices of the individual goods.

The marginals of this distribution follow the four parameter Beta distribution, Johnson et al., 1995, $Beta(h_i, h_+ - h_i, 0, \frac{m}{p_i})$, which is a linear transformation of the standard Beta. This can be proven by noting that the marginal distributions of the preference parameters are themselves Beta, Monti et al., 2011. Thus, α_i has density

$$f_{\alpha_i} = \frac{1}{B(h_i, h_+ - h_i)} \alpha_i^{h_i - 1} (1 - \alpha_i)^{h_+ - h_i - 1} \tag{D.8}$$

Where $B(a, b)$ is the Beta function,

$$B(a, b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \tag{D.9}$$

The change of variable formula is again used to get the marginal distribution of demand from the density of consumption share.

$$\begin{aligned} f_{c_i} &= f_{\alpha_i}(\frac{c_i p_i}{m}) \left|\frac{p_i}{m}\right| \\ &= \frac{1}{B(h_i, h_+ - h_i)} \left(\frac{c_i p_i}{m}\right)^{h_i - 1} \left(1 - \frac{c_i p_i}{m}\right)^{h_+ - h_i - 1} \left|\frac{p_i}{m}\right| \\ &= \frac{1}{B(h_i, h_+ - h_i)} c_i^{h_i - 1} \left(\frac{m}{p_i} - c_i\right)^{h_+ - h_i - 1} \left(\frac{m}{p_i}\right)^{-(h_+ - 1)} \end{aligned} \tag{D.10}$$

This is the density of the four parameter Beta, $Beta(h_i, h_+ - h_i, 0, \frac{m}{p_i})$. QED.

Appendix E

Derivation of Elasticities

The derivation of ϵ_1 and ϵ_3 is relatively straight forward. Starting from the definition of ϵ_3,

$$\frac{\partial log \mathbf{E}(Y)}{\partial log x} \tag{E.1}$$

Using the law of iterated expectations

$$\begin{aligned} &= \frac{\partial log \mathbf{E}(\mathbf{E}(Y|\mathbf{X}))}{\partial log x} \\ &= \frac{\partial log \mathbf{E}(N m e^{\beta \mathbf{X}})}{\partial log x} \end{aligned} \tag{E.2}$$

If we assume the size of the market is independent of $\mathbf{X}$, it can be pulled out of the expectation:

$$= \frac{\partial log(N)}{\partial log x} + \frac{\partial log \mathbf{E}(m e^{\beta \mathbf{X}})}{\partial log x} \tag{E.3}$$

The second summand is just ϵ_1. The assumption of independence between market size and $\mathbf{X}$, also entails the first summand is 0. Thus, $\epsilon_1 = \epsilon_3$.

$$= \frac{\partial \mathbf{E}(m e^{\beta \mathbf{X}}) / \partial log x}{\mathbf{E}(m e^{\beta \mathbf{X}})} \tag{E.4}$$

If we assume the moment generating function of $\mathbf{X}$ is finite everywhere - which is not an unreasonable assumption given that an economic variable must be truncated to a compact set, otherwise its production process would contradict the laws of thermodynamics - and thus measurable, therefore dominated by a measurable function, namely itself, the derivative can by pulled inside the expectation:

$$= \frac{\beta_x \mathbf{E}(me^{\beta \mathbf{X}})}{\mathbf{E}(me^{\beta \mathbf{X}})} = \beta_x \tag{E.5}$$

The derivation of ϵ_2 is also straightforward. Again, starting with the definition

$$\mathbf{E}\left(\frac{\partial logY}{\partial logx}\right) \tag{E.6}$$

Start with the law of iterated expectations:

$$= \mathbf{E}\left(\mathbf{E}\left(\frac{\partial logY}{\partial logx}\middle|\mathbf{X}\right)\right) \tag{E.7}$$

Since Y is a four-parameter beta distributed random variable, we know that it's log is measurable and dominated by itself, thus, the derivative can be pulled out of the expectation:

$$= \mathbf{E}\left(\frac{\partial}{\partial logx}\mathbf{E}\left(logY|\mathbf{X}\right)\right) \tag{E.8}$$

To evaluate this, we need to be more explicit about the relationship between the four parameter beta and the standard beta. If $Y \sim Beta(\alpha, \beta, 0, m)$, then $Y = mW$ where $W \sim Beta(\alpha, \beta)$. Since $m \in \mathbf{X}$, it can be pulled outside the conditional expectation:

$$= \mathbf{E}\left(\frac{\partial}{\partial logx} logm + \mathbf{E}\left(logW|\mathbf{X}\right)\right) \tag{E.9}$$

The expected value of the log of a beta random variable is known

$$= \mathbf{E}\left(\frac{\partial}{\partial logx}(logm + \psi(e^{\beta\mathbf{X}}) - \psi(\phi))\right) \tag{E.10}$$

In the empirical exercises, it is assumed that ϕ is independent of the $\mathbf{X}$, thus the derivative $\frac{\partial}{\partial logx}\psi(\phi) = 0$; the derivative $\frac{\partial}{\partial logx}logm = \delta_m(x) = 1\ if\ x = m;\ 0\ otherwise.$

$$= \mathbf{E}\left(\delta_m(x) + \beta_x e^{\beta\mathbf{X}}\psi^{(1)}(e^{\beta\mathbf{X}})\right) \tag{E.11}$$

$\psi(x)$ and $\psi^{(1)}(x)$ are the digamma and trigamma functions, the first and second derivatives of the log of the gamma function.