Bayesian Inference for Semiparametric Dynamic Choice Quantal Response Equilibrium Models

Kevin M. Quinn* Anton H. Westveld†

Abstract
Building on work of (McKelvey & Palfrey 1995, McKelvey & Palfrey 1996, McKelvey & Palfrey 1998) and (Signorino 1999) on quantal response equilibrium models and (Newton, Czado & Chappell 1996, Ishwaran & Zarepour 2000, Ishwaran & James 2001) and (Ishwaran & Zarepour 2002), on semiparametric Bayesian methods, we develop Bayesian inference for a particular class of semiparametric strategic choice models. Unlike previous approaches that have assumed the disturbances entering into actors' random utility functions follow a particular, known distribution such as a Gaussian distribution or a type I extreme value distribution, we instead assume only that the differenced disturbances have a right-continuous distribution function with fixed median and interquartile range. We model such a random distribution function using an approximation to the centrally standardized Dirichlet process prior of (Newton, Czado & Chappell 1996). Model fitting is accomplished via Markov chain Monte Carlo. We present results from Monte Carlo experiments and a simulated data example. Our Monte Carlo results show that incorrectly specifying the distribution of the actors' utility disturbances within a parametric model can dramatrically bias quantities of interest such as the conditional expectation function and fitted probabilities. Our simulated data example demonstrates that our semiparametric approach works well on a difficult example in which the underlying true distribution function of the differenced disturbances is highly skewed.

Key Words: Bayesian semiparametric estimation, multi-agent choice, quantal response equilibrium, social interaction data.

1. Introduction

In recent years there has a been a great deal of interest in empirical tests of formal (oftentimes game-theoretic) models of behavior. Several methods of testing have been advocated and used. These include: the comparison of equilibrium predictions to observational data (Enelow & Hinich 1984, Erikson & Romero 1990, Quinn & Martin 2002) and to experimental data (Fiorina & Plott 1978, Camerer & Weigelt 1988, McKelvey & Ordeshook 1990), comparative statics analysis (Segal, Cameron & Cover 1992), and the specification and estimation of the structural model of interest (McKelvey & Palfrey 1995, McKelvey & Palfrey 1996, McKelvey & Palfrey 1998, Signorino 1999, Mebane 2000)[1].

Perhaps the most exciting general approach has been the use of quantal response equilibrium (QRE) models developed by (McKelvey & Palfrey 1995, McKelvey & Palfrey 1996, McKelvey & Palfrey 1998) along with the extensions proposed by (Signorino 1999, Signorino & Yilmaz 2003). QRE is a game theoretic solution concept that directly incorporates a stochastic component into actors utility functions. Such random disturbances can be thought of as capturing elements of bounded rationality or heterogeneity of tastes. QREs have a number of nice properties including the fact that all outcomes have positive probability of occurring under a QRE. This eliminates the

*Professor of Law; UC Berkeley School of Law; 490 Simon #7200, University of California, Berkeley, CA 94720-7200; *e-mail*: `kquinn@law.berkeley.edu`. Quinn's research was supported by NSF grant SES 01-36679.

†Assistant Professor of Statistics; Department of Mathematical Sciences; University of Nevada, Las Vegas; 4505 S. Maryland Parkway, Box 454020, Las Vegas, NV 89154-4020; *e-mail*: `anton.westveld@unlv.edu`.

[1]For a comprehensive review and synthesis of various methods of empirically testing formal models see (Morton 1999).

so-called zero-likelihood problem. This problem happens when a single observation that has zero probability of occurring under a particular solution concept is observed and consequently sets the likelihood function equal to 0 and falsifies the model.

While QRE models are, in principle, quite general[2], empirical applications of such models often rely on very specific, non-innocuous assumptions regarding the distributions of the unobserved disturbances assumed to enter into the actors' random utility functions. Throughout the paper, we refer to these distribution functions as "link functions"[3]. As we show later in the paper, incorrect specification of the link function can produce substantial bias in estimates of action probabilities, outcome probabilities, and conditional expectation functions. This problem is compounded by the fact that since the random disturbances are never directly observed it is not possible to use their empirical distribution to inform ones choice of link.

In this paper, we solve this problem by relaxing the assumption of a known parametric form for the link function. Rather than conditioning on one of the infinitely many distribution functions as our link, we put a prior probability measure over the space of distribution functions and allow the data to speak to the form of the link function. Our methods work well on a wide variety of data generating processes, are both more flexible and more numerically stable than methods such as scobit (Nagler 1994, Altman & McDonald 2003), and do not require strong assumptions about the form of the link function. Our approach is semiparametric in that it assumes the systematic component of the actors' utilities follows a linear parametric form while at the same time making only weak assumptions about the form of the stochastic component of the utility functions. Our approach is also Bayesian in that we use a prior probability measure over the space of distribution functions to allow for learning about the true, unobserved link.

There are at least two different motivations for our semiparametric approach. The first is the usual reason for performing semiparametric inference. Namely, we would like to estimate the effect of x on y but we are unwilling or unable to specify portions of the probability model. In our case we leave the link function only minimally specified (it is some distribution function) and proceed to make inferences based on this partial model. Another reason for adopting our approach focuses more on its Bayesian aspects than on its semiparametric aspects. In some situations, such as experimental work, it may actually be of some interest to learn if, and to what extent, the disturbance distributions differ across particular strata of actors. For instance, does the stochastic portion of utility look different for the wealthiest 20% of the population than for the least wealthy 20%? Are there differences in the stochastic portion of utility between men and women? Are there differences based on the education of the actors? These are but some of the questions that could, in principle, be answered by our approach.

This paper is organized as follows. Section 2 provides a brief overview of QRE models and discusses the game form that we will be using throughout the examples in the paper. The next section investigates the effects of misspecifying the link function in QRE models via a series of Monte Carlo experiments. Here we see that, in some cases, the misspecification bias in fitted probabilities can be as large as approximately ± 0.3. Section 4 describes our semiparametric model and describes how inference can be performed using Markov chain Monte Carlo (MCMC). Section 5 presents results from a simulated data example. The results demonstrate that our method works well even on difficult problems where the link function is highly skewed. The last section concludes.

[2]For instance, all of the theorems in (McKelvey & Palfrey 1995, McKelvey & Palfrey 1998) are proven under quite general conditions.

[3]This is different from the terminology used in the generalized linear modeling literature where what we are calling a "link function" is referred to as an "inverse link function".

2. QRE Models

QRE is a game-theoretic solution concept developed by (McKelvey & Palfrey 1995, McKelvey & Palfrey 1996, McKelvey & Palfrey 1998). As (McKelvey & Palfrey 1996) discuss, the impetus for their work was to find a, "... simple extension of of the standard model of Nash equilibrium which can explain some of the systematic violations of Nash equilibrium [observed in experimental games]" (p. 187). While there are multiple interpretations of QRE behavior (see (Chen, Friedman & cois Thisse 1997) and (McKelvey & Palfrey 1996)), the basic idea is that there is a stochastic component to agents' utility functions and that agents play best-responses to the expected actions of the other agents. For a comprehensive treatment of the QRE solution concept we refer the reader to the work of (McKelvey & Palfrey 1995, McKelvey & Palfrey 1996, McKelvey & Palfrey 1998). In what follows we restrict our attention to dynamic choice (extensive form) games of perfect information with binary action sets at each node[4].

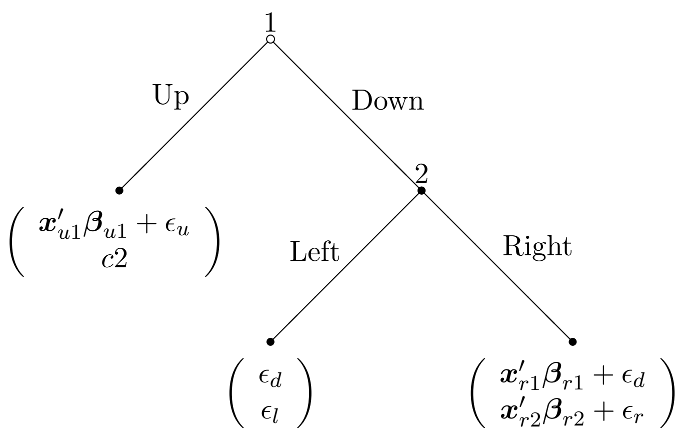

Figure 1: *The example game used throughout the manuscript.*

In the empirical examples that follow we will be working with the game form in Figure 1. This game features two players (labeled 1 and 2) each of whom can take one of two actions. To find the QRE of this game we begin with player 2's decision problem. If player 2 chooses Left she receives a payoff of ϵ_l while if she chooses Right she receives a payoff of $\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2} + \epsilon_r$. Here $\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2}$ represents the systematic portion of utility that is assumed to be observed by all players, while ϵ_l and ϵ_r represent the stochastic portion of utility. The realizations of ϵ_l and ϵ_r are assumed to be observed by player 2 but not by player 1. Player 1 is assumed to know the distribution of $\epsilon_l - \epsilon_r$.

Since player 2 is assumed to be a utility maximizer she will choose Right if

$$\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2} + \epsilon_r \geq \epsilon_l$$

or equivalently if

$$\epsilon_l - \epsilon_r \leq \boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2}.$$

Assume that $\epsilon_l - \epsilon_r$ has distribution function $G_2(\cdot)$. Then the probability that player 2 will choose Right in equilibrium is

$$\Pr(\epsilon_l - \epsilon_r \leq \boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2}) = G_2(\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2})$$

[4]In principle, our semiparametric methods can be applied to dynamic choice (extensive form) games of perfect information with non-binary action sets.

and, of course, the probability that player 2 will choose Left in equilibrium is $1 - G_2(\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2})$. Player 1's choice problem involves the choice between the certain payoff involved with the choice of Up and the uncertain payoff associated with the choice of Down. A bit of simple algebra reveals that the expected payoff to player 1 of choosing Down is

$$G_2(\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2})\boldsymbol{x}'_{r1}\boldsymbol{\beta}_{r1} + \epsilon_d.$$

Since player 1 is assumed to be a utility maximizer he will choose Up if

$$\boldsymbol{x}'_{u1}\boldsymbol{\beta}_{u1} + \epsilon_u \geq G_2(\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2})\boldsymbol{x}'_{r1}\boldsymbol{\beta}_{r1} + \epsilon_d$$

or equivalently if

$$\epsilon_d - \epsilon_u \leq \boldsymbol{x}'_{u1}\boldsymbol{\beta}_{u1} - G_2(\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2})\boldsymbol{x}'_{r1}\boldsymbol{\beta}_{r1}.$$

Assume $\epsilon_d - \epsilon_u$ has distribution function $G_1(\cdot)$. Then the probability that player 1 will choose Up in equilibrium is

$$\Pr(\epsilon_d - \epsilon_u \leq \boldsymbol{x}'_{u1}\boldsymbol{\beta}_{u1} - G_2(\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2})\boldsymbol{x}'_{r1}\boldsymbol{\beta}_{r1}) = G_1(\boldsymbol{x}'_{u1}\boldsymbol{\beta}_{u1} - G_2(\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2})\boldsymbol{x}'_{r1}\boldsymbol{\beta}_{r1})$$

and the probability that 1 chooses Down in equilibrium is simply $1 - G_1(\boldsymbol{x}'_{u1}\boldsymbol{\beta}_{u1} - G_2(\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2})\boldsymbol{x}'_{r1}\boldsymbol{\beta}_{r1})$. In what follows, the outcome variable is denoted y. The outcome of the ith game is coded as follows:

$$y_i = \begin{cases} 1 & \text{iff player 1 chooses Up} \\ 2 & \text{iff player 1 chooses Down and player 2 chooses Left} \\ 3 & \text{iff player 1 chooses Down and player 2 chooses Right.} \end{cases}$$

Assuming independence across repetitions of the game, and allowing $i = 1, \ldots, n$ to index these repetitions, the likelihood function for $\boldsymbol{y}$ is:

$$L(\boldsymbol{\beta}_{u1}, \boldsymbol{\beta}_{r1}, \boldsymbol{\beta}_{r2}|\boldsymbol{y}) = \prod_{i=1}^{n} p_{1i}^{\mathbb{I}(y_i=1)} p_{2i}^{\mathbb{I}(y_i=2)} p_{3i}^{\mathbb{I}(y_i=3)} \tag{1}$$

where $\mathbb{I}(a = b)$ is the indicator function that equals 1 if and only if $a = b$ and equals 0 otherwise, and p_{1i}, p_{2i}, and p_{3i} are the outcome probabilities. If the distribution functions $G_1(\cdot)$ and $G_2(\cdot)$ are known then the outcome probabilities take the following forms:

$$\begin{aligned} p_{1i} &= G_1(\boldsymbol{x}'_{u1}\boldsymbol{\beta}_{u1} - G_2(\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2})\boldsymbol{x}'_{r1}\boldsymbol{\beta}_{r1}), \\ p_{2i} &= (1 - p_{1i})(1 - G_2(\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2})), \\ p_{3i} &= (1 - p_{1i})G_2(\boldsymbol{x}'_{r2}\boldsymbol{\beta}_{r2}). \end{aligned}$$

However, $G_1(\cdot)$ and $G_2(\cdot)$ will typically not be known so arbitrary distribution functions $F_1(\cdot)$ and $F_2(\cdot)$ will be used in their place. In most parametric models $F_1(\cdot)$ and $F_2(\cdot)$ are assumed to be either normal distributions or logistic distribution functions. In addition, it is typically assumed that $F_1(\cdot) = F_2(\cdot)$. In this paper, we refer to $F_1(\cdot)$ and $F_2(\cdot)$ as "link" functions. The consequences of specifying an $F_1(\cdot) \neq G_1(\cdot)$ and an $F_2(\cdot) \neq G_2(\cdot)$ are taken up in the next section.

3. The Effects of Mis-Specified Link Functions in QRE Models

In order to investigate the effects of mis-specification of link functions in QRE models, first single agent statistical choice models are discussed. The work done on the mis-specification of link functions for binary regressions models is directly applicable since single agent binary choice statistical models motivated via random utility theory are simply binary regression models. Next, a Monte Carlo experiment was conducted for multi-agent statistical choice models, specifically QRE models, to expand the results obtained from single agent choice.

3.1 Single Agent Choice

The effects of link mis-specification in the case of binary regression have been studied extensively through analytical means for asymptotic results as well as Monte Carlo simulations for small samples (Czado & Satner 1992). Since a statistical model for a single agent binary choice can be written in terms of a binary regression, the results garnered from those studies are directly applicable to the research at hand.

(Czado & Satner 1992) show in theorems 3.1-3.4 that an asymptotic bias exists in the predicted probabilities if the data are generated under $G(\cdot)$ but the link function $F(\cdot) \neq G(\cdot)$ is used in the estimation. Furthermore, (Czado & Satner 1992) show via Monte Carlo simulations involving logistic regression that

> there was a substantial increase in the bias of the predicted probabilities in all cases studied when the link was mis-specified compared with that when the [true] link was logistic. Logistic estimation is affected more by skewness than kurtosis of the true link function. (pp. 229-230)

We expect similar results to hold for multi-agent QRE models. We turn our attention to this next.

3.2 Multi Agent Choice

Since the QRE model is an extension of the single agent statistical choice model presented above we would expect similar results in terms of the bias of the predicted probabilities. In fact, we might expect the bias to be worse for predicted probabilities occurring higher in the game tree due to further complexity of the QRE models compared to the single agent statistical choice models. The complexity also makes asymptotic results similar to (Czado & Satner 1992) fairly taxing. Therefore, we conducted a Monte Carlo study in order to investigate the effects of link mis-specification for QRE models. Although data were generated using several different link functions and several different sample sizes, for reasons of space, we only we report the results from data generated with three link functions and a sample size of 1000. In each case, the link function used for estimation is the logit link and maximum likelihood estimates were obtained. The focus on the logit link in this paper is due to its nice mathematical properties, leading to its wide empirical implementation first in the area of single-agent choice (Ben-Akiva & Lerman 1985), and then in the area of multi-agent choice. The game form assumed throughout the Monte Carlo experiment is that presented in Figure 1. In each case the Monte Carlo experiment consisted of 500 Monte Carlo replications. $\boldsymbol{x}_{u1}$ is a 1000×2 matrix that consists of a vector of 1s along with vector of draws from a standard normal distribution. $\boldsymbol{x}_{r1}$ and $\boldsymbol{x}_{r2}$ have a similar structure with the added proviso that the normal draws in $\boldsymbol{x}_{r1}$ and $\boldsymbol{x}_{r2}$ are independent of each other and the normal draws in $\boldsymbol{x}_{u1}$.

The three link functions used to generate the data are

1. $G^l(z) = \exp(z)/(1 + \exp(z))$ (logistic)
2. $G^p(z) = \Phi(z|0, 1)$ (probit)
3. $G^m(z) = 0.02\Phi(z|{-4.5}, 3.25) + 0.1\Phi(z|{-3.5}, 2) + 0.2\Phi(z|{-2}, 2) + 0.3\Phi(z|0, 1) + 0.35\Phi(z|0.75, 0.35) + 0.03\Phi(z|4.25, 2.25)$ (mixture of normals)

where $\Phi(z|\mu, \sigma)$ represents the distribution function of a normal random variable with mean μ and standard deviation σ. In the Monte Carlo experiments, we assume that $G_1(\cdot) = G_2(\cdot)$. Figure 2

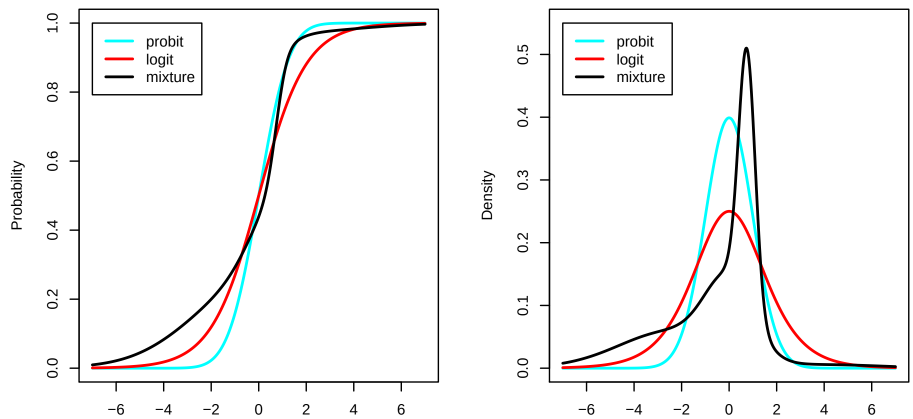

Figure 2: *Comparison of three link functions and the associated densities — for the differenced disturbances.*

depicts these three link functions, as well as the associated density functions, for the differenced disturbances.

The coefficient values used to generate the outcome data in the experiment with $G^l(\cdot)$ as the data generating link are $\boldsymbol{\beta}_u = \boldsymbol{\beta}_{r1} = \boldsymbol{\beta}_{r2} = (0, 1.7)'$. When the probit link was used to generate the data the true coefficient values were assumed to be $\boldsymbol{\beta}_u = \boldsymbol{\beta}_{r1} = \boldsymbol{\beta}_{r2} = (0, 1)'$. Finally, when the mixture of normals link was used to generate the data the true coefficient values were $\boldsymbol{\beta}_{u1} = \boldsymbol{\beta}_{r1} = \boldsymbol{\beta}_{r2} = (0, 2.75)'$.

Figures (3,4,5) show the bias versus the true predicted probabilities for the actions {Up (same as $Y = 1$), Down; Left, Right} and the three different outcomes $\{Y = 1,\ Y = 2,\ Y = 3\}$. The light blue vertical lines are Monte Carlo 90% confidence intervals for the true probability minus the estimated fitted probability. These predicted probabilities for the different actions and outcomes were obtained by allowing a particular x variable to vary across the range from -4 to 4. The particular x variable used in each plot is noted on the plot's x-axis. The other variables were set to zero. Figure 3 shows that when the true link function $G(\cdot)$ is equal to the link function $F(\cdot)$ used in estimation, in this case both are logistic, the estimated probabilities are not noticeably biased (smaller sample sizes did show some bias). When the true link function is the probit link but the model is estimated using a logistic link, there does exist a slight bias. Finally the results from the experiment in which the data were generated using the mixture of Gaussians link agrees with the results of (Czado & Satner 1992) in that the degree of skewness of the true link function $G(\cdot)$ in comparison to the link function $F(\cdot)$ used in the estimation does greatly affect the bias.

4. A Semiparametric QRE Model

As noted earlier, the basic idea behind our semiparametric model is that instead of assuming a particular link function we assign a prior probability measure to the space of link functions. Before discussing our implementation we briefly discuss how one might put a prior on the space of distribution functions which is of fundamental importance to the field of Bayesian nonparametrics and semiparametrics.

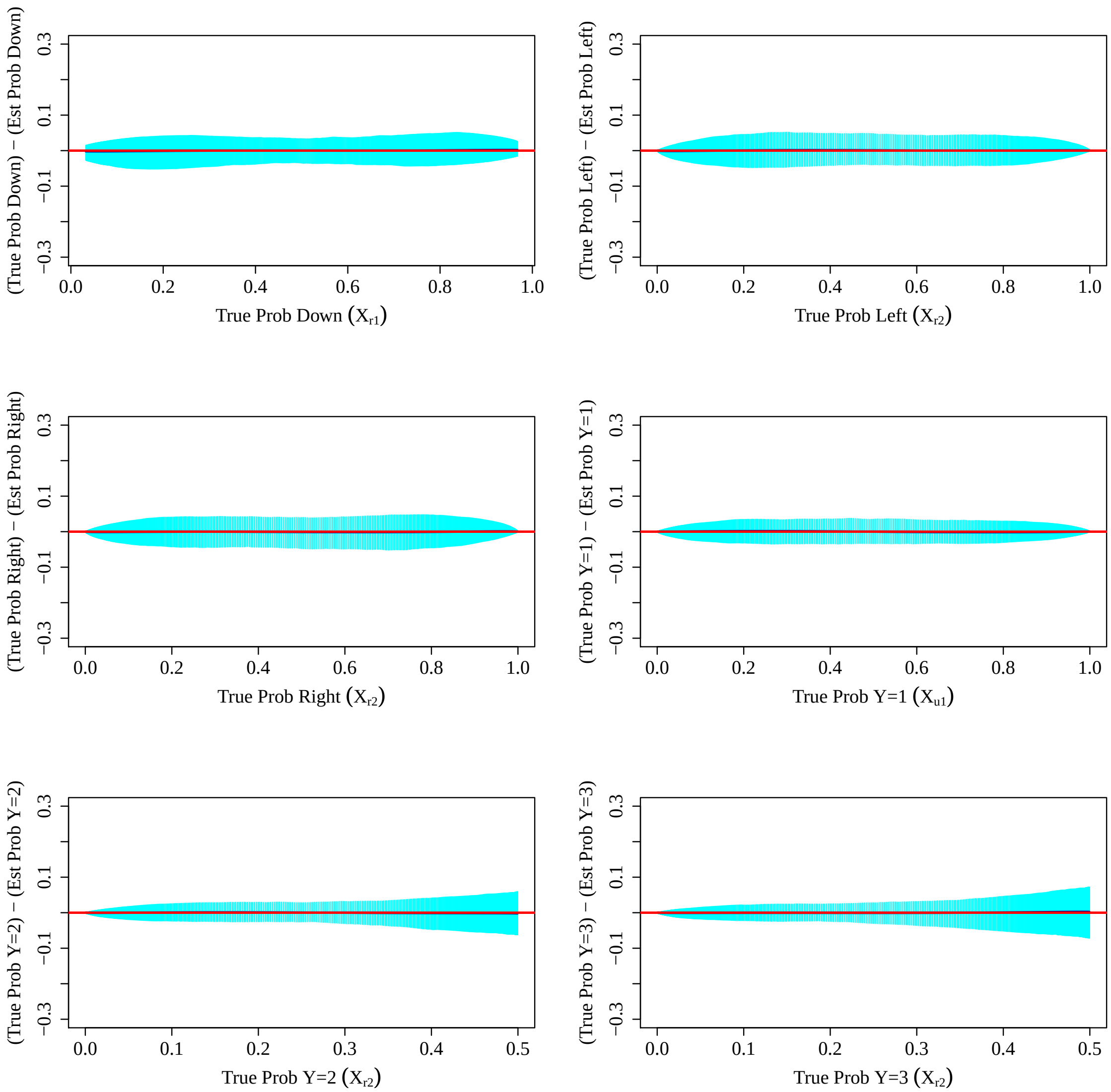

Figure 3: *Plots of the bias in the fitted probabilities from the Monte Carlo experiment in which data were generated with a logit link and a logit link was used in estimation.*

4.1 Bayesian Nonparametrics and Semiparametrics

The most commonly used prior measure defined over the space of distribution functions is the Dirichlet process prior of (Ferguson 1973, Ferguson 1974). To understand the Dirichlet process prior we need to introduce a bit of additional notation. Let $\mathbb{R}$ denote the real line and $\mathcal{B}$ the σ-algebra of Borel subsets of $\mathbb{R}$. Let P denote a random probability measure on $(\mathbb{R}, \mathcal{B})$. The distribution function formed from P is $F(t) = P((-\infty, t])$. We say that P (and hence F) follows a *Dirichlet process* $\mathcal{D}(\alpha H)$ and write $P \sim \mathcal{D}(\alpha H)$ if for every finite, measurable partition $B_1, \ldots, B_k$ $(P(B_1), \ldots, P(B_k)) \sim Dirichlet(\alpha H(B_1), \ldots, \alpha H(B_k))$. Here H is a probability measure on $(\mathbb{R}, \mathcal{B})$ and α is a positive scalar parameter. H is called the centering measure while the measure αH is referred to as the base measure. As α gets larger, the Dirichlet process put more mass near the

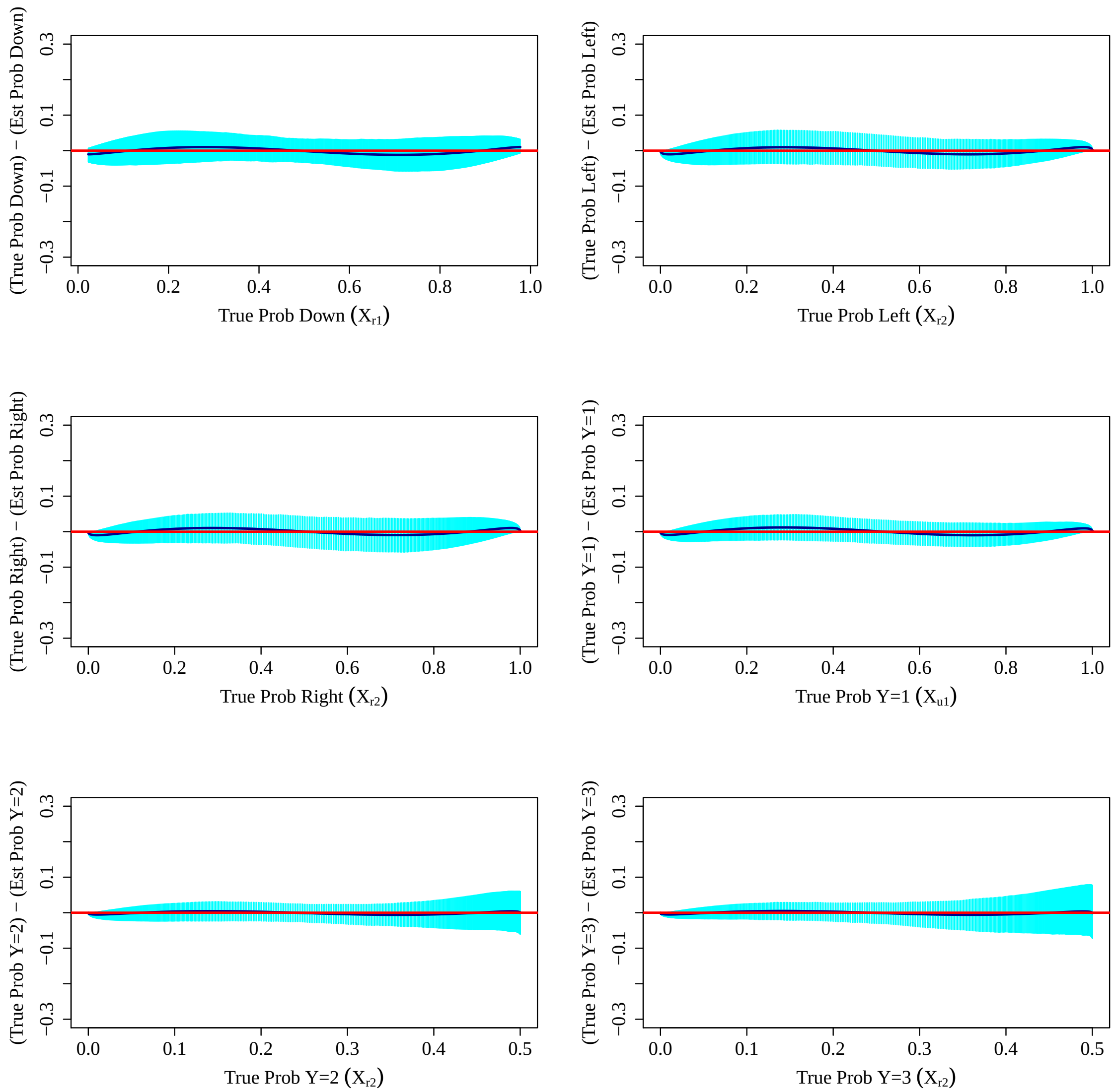

Figure 4: *Plots of the bias in the fitted probabilities from the Monte Carlo experiment in which data were generated with a probit link and a logit link was used in estimation.*

centering measure H. In other words, for large values of α we would expect to see realizations of the process that look very much like H.

(Ferguson 1973, Ferguson 1974) state a number of useful properties of the Dirichlet process. The most important for our work here is that if the support of αH is $\mathbb{R}$ then the support of $\mathcal{D}(\alpha H)$ with respect to convergence in law is the set of all probability measures on $(\mathbb{R}, \mathcal{B})$. This is clearly a very nice property in that $\mathcal{D}(\alpha H)$ can be used to mimic any probability measure with the same support as αH. Another property of the Dirichlet process that is worth mentioning is that a realization from the process is almost surely discrete. This creates difficulties in some applications. Nonetheless, in our application discreteness of the link function does not impose any serious problems. In fact, as we will see now, it is possible to exploit this almost sure discreteness to arrive at a computationally

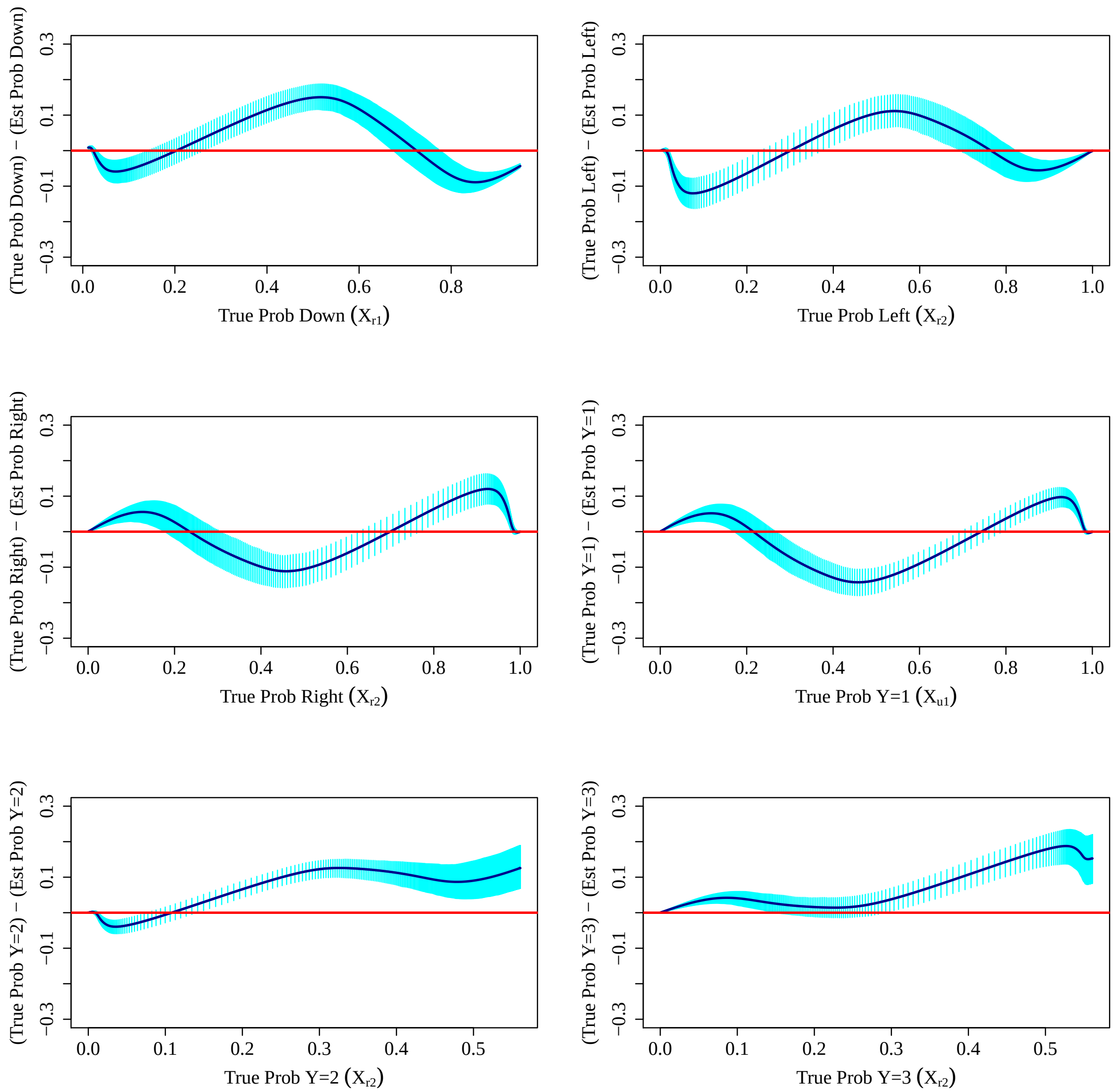

Figure 5: *Plots of the bias in the fitted probabilities from the Monte Carlo experiment in which data were generated with the mixture of normals link discussed in the body of the manuscript and a logit link was used in estimation.*

attractive discrete approximation to the Dirichlet process.

In a series of papers (Ishwaran & Zarepour 2000, Ishwaran & James 2001, Ishwaran & Zarepour 2002), Ishwaran and colleagues discuss a finite-dimensional approximation to the Dirichlet process. More specifically, they consider random probability measures of the form

$$P_m(\cdot) = \sum_{j=1}^{m} p_j \delta_{z_j}(\cdot)$$

where $\delta_{z_j}(\cdot)$ is a unit point mass concentrated at z_j, $(p_1, \ldots, p_m) \sim Dirichlet(\alpha/m, \ldots, \alpha/m)$, and $z_j \overset{iid}{\sim} H,\ j = 1, \ldots, m$ where it is further assumed that $\boldsymbol{p}$ is independent of $\boldsymbol{z}$. We say that such a random probability measure follows a *finite dimensional Dirichlet process prior* and write $P_m \sim \mathcal{D}_m(\alpha H)$. For an independent development of this type of prior see (Walker & Wakefield 1998). (Ishwaran & Zarepour 2002) prove that the $\mathcal{D}_m(\alpha H)$ can be used to approximate integrable functions of $\mathcal{D}(\alpha H)$. They also discuss methods of choosing m in order to achieve a certain amount accuracy of approximation.

So far, we have discussed random probability measures that are not restricted in terms of their location and scale. Such restrictions on the link function are necessary to achieve identification in binary regression models as well as QRE models. (Newton, Czado & Chappell 1996) propose a modification to the Dirichlet process that they call the *centrally standardized Dirichlet process* that has a fixed location and scale — through fixing the median and the interquartile range. Their approach works by breaking the random measure up into four pieces each of which has measure 1/4. They then exploit a method for taking a random sample from a realization of the centrally standardized Dirichlet process to develop a data augmentation scheme for model fitting.

4.2 The Proposed Semiparametric QRE Model

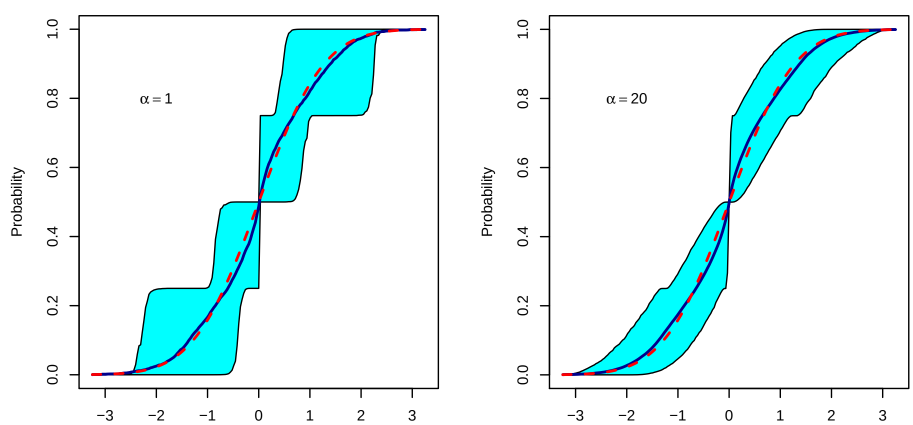

Figure 6: *Illustrations of two centrally standardized finite Dirichlet processes.*

Our proposed semiparametric prior for the link function is a centrally standardized (to use Newton, Czado & Chappell (1996)'s terminology) version of the finite dimensional Dirichlet process prior of (Ishwaran & Zarepour 2000, Ishwaran & James 2001) and (Ishwaran & Zarepour 2002). We need to rely on the finite dimensional version of the centrally standardized Dirichlet process because, unlike the binary regresssion case studied by (Newton, Czado & Chappell 1996), our QRE model is not readily amenable to data augmentation. As such, we need to work directly with realizations from the process within our MCMC scheme. Through the use of a finite dimensional Dirichlet process.

In order to identify the model we need to fix the location and scale of the link function. To do this, we assume that the median of the differenced disturbances is 0 and the interquartile range is $d = 1.349$ which is the interquartile range of a standard normal random variable. While the median and interquartile range are fixed *a priori* we allow for skewness by allowing the third quartile to be governed by a parameter θ. Our prior distribution for θ is:

$$\theta \sim \mathcal{U}nif(0, d).$$

The next step in formulating our prior is to choose a centering measure for the support points (the z_js). The centering measure we use here is based on the normal distribution. More specifically, we assume

$$z_j \stackrel{iid}{\sim} \mathcal{N}(0,1) \quad j = 1, \ldots, m.$$

If we think of the z_js as parameters, this distributional assumption serves as the prior distribution for $\boldsymbol{z}$.

Rather than specify the prior distribution of the probability weights (p_j) directly in terms of the Dirichlet distribution we use an equivalent specification in terms of the gamma distribution. We assume

$$p_j^* \sim \mathcal{G}amma(\alpha/m, 1) \quad j = 1, \ldots, m,$$

and

$$p_j = \begin{cases} \frac{p_j^*}{4\sum_{k\in Q_1} p_k^*} & \text{if } j \in Q_1 \\ \frac{p_j^*}{4\sum_{k\in Q_2} p_k^*} & \text{if } j \in Q_2 \\ \frac{p_j^*}{4\sum_{k\in Q_3} p_k^*} & \text{if } j \in Q_3 \\ \frac{p_j^*}{4\sum_{k\in Q_4} p_k^*} & \text{if } j \in Q_4, \end{cases}$$

where

$$\begin{aligned} Q_1 &= \{j \in \{1,2,\ldots,m\} : z_j \in (-\infty, \theta - d]\} \\ Q_2 &= \{j \in \{1,2,\ldots,m\} : z_j \in (\theta - d, 0]\} \\ Q_3 &= \{j \in \{1,2,\ldots,m\} : z_j \in (0, \theta]\} \\ Q_4 &= \{j \in \{1,2,\ldots,m\} : z_j \in (\theta, \infty)\}. \end{aligned}$$

The distribution function[5] that results is

$$F(t) = \begin{cases} \sum_{j\in Q_1} p_j \mathbb{I}[z_j \le t] & \text{if } t \in (\infty, \theta - d] \\ 0.25 + \sum_{j\in Q_2} p_j \mathbb{I}[z_j \le t] & \text{if } t \in (\theta - d, 0] \\ 0.50 + \sum_{j\in Q_3} p_j \mathbb{I}[z_j \le t] & \text{if } t \in (0, \theta] \\ 0.75 + \sum_{j\in Q_4} p_j \mathbb{I}[z_j \le t] & \text{if } t \in (\theta, \infty]. \end{cases}$$

To generate a realization (i.e., a distribution function) from this process one samples θ from its prior distribution, $\boldsymbol{z}$ from its prior distribution, and $\boldsymbol{p}^*$ from its prior distribution and then forms the associated distribution function.

To get a sense of how the concentration parameter α affects the process we drew 5,000 distribution functions from a process with $\alpha = 1$ and a process with $\alpha = 20$. In both cases we set $m = 250$. This results are presented in Figure 6. The solid dark blue line is the pointwise expectation of the distribution functions generated from each process, while the light blue shaded region is a pointwise central 90% probability band. The dashed red line is the standard normal distribution function which is the distribution function of the centering measure for each process. As we would expect, in each case the realized distribution functions are clustered around the standard normal distribution

[5] Technically, $F(\cdot)$ is a sub-distribution function since $F(\infty) \le 1$ rather than $F(\infty) = 1$.

(centering measure). When $\alpha = 1$ there is quite a bit of variability around the standard normal distribution function, while with $\alpha = 20$ the realized distribution functions tend to cluster more tightly.

To complete the model we specify the improper, uniform prior $p(\boldsymbol{\beta}_u, \boldsymbol{\beta}_{r1}, \boldsymbol{\beta}_{r2}) = 1$. While this lack of prior information about the coefficient vectors is consistent with the spirit of our semiparametric approach there is no reason that an informative prior could not be used if prior information was available. With this specification of the link function, the likelihood function for a QRE model now depends on θ, $\boldsymbol{z}$ and $\boldsymbol{p}^*$ as well as $\boldsymbol{\beta}$. Nonetheless, the form of the likelihood function is basically the same, the only real difference being what link function is used. The posterior for this model is formed in the usual way.

4.3 Model Fitting and Inference

Model fitting is accomplished via MCMC through updating parameters in the following blocks: $[\theta|\boldsymbol{z}, \boldsymbol{p}^*, \boldsymbol{\beta}_u, \boldsymbol{\beta}_{r1}, \boldsymbol{\beta}_{r2}]$, $[\boldsymbol{z}|\boldsymbol{p}^*, \boldsymbol{\beta}_u, \boldsymbol{\beta}_{r1}, \boldsymbol{\beta}_{r2}, \theta]$, $[\boldsymbol{p}^*|\boldsymbol{\beta}_u, \boldsymbol{\beta}_{r1}, \boldsymbol{\beta}_{r2}, \theta, \boldsymbol{z}]$, $[\boldsymbol{\beta}_u|\boldsymbol{\beta}_{r1}, \boldsymbol{\beta}_{r2}, \theta, \boldsymbol{z}, \boldsymbol{p}^*]$, $[\boldsymbol{\beta}_{r1}|\boldsymbol{\beta}_{r2}, \theta, \boldsymbol{z}, \boldsymbol{p}^*, \boldsymbol{\beta}_u]$, $[\boldsymbol{\beta}_{r2}|\theta, \boldsymbol{z}, \boldsymbol{p}^*, \boldsymbol{\beta}_u, \boldsymbol{\beta}_{r1}]$.

5. Results from Simulated Data

To get a sense as to how our semiparametric model performs, we decided to look at a simulated dataset. We wanted to look at a difficult case, i.e., a dataset that was generated using a link function that was fairly far away from the standard normal centering distribution in our semiparametric prior, but that was also not so perverse as to be completely unrealistic. To this end, we generated one dataset of 1,000 observations using the mixture of normals link and same general setup as used in the Monte Carlo experiments of section 3. We then fit our semiparametric model to these data using a prior parameterization of $m = 100$, $\alpha = 5$. Additionally, the centering distribution is the normal distribution. For reasons of simplicity we assumed that the distribution of $\epsilon_l - \epsilon_r$ is the same as the distribution of $\epsilon_d - \epsilon_u$ both in the data generating process as well as estimation. In actual empirical applications it would typically be wise to relax this assumption and assume distinct finite dimensional Dirichlet process priors for the two link functions. The output is based on 10,000,000 MCMC scans, collected after a burn-in of 50,000 scans. Every 1,000th draw was stored.

Figure 7 presents the results from this model. As noted in passing, one of the byproducts of our approach is a posterior over the the space of link functions which allows us to make inferences about the shape of the link function underlying the data generating process. The panel in the top left corner presents the pointwise posterior expectation of the link function along with pointwise 90% posterior credible bands. The true link function is superimposed after being rescaled to have median 0 and interquartile range 1.349. These normalizations are the same normalizations used to identify our semiparametric model. As can be seen from this figure, the semiparametric model does a very good job of estimating the link function. The pointwise posterior mean is very close to the true link and the true link is entirely within the pointwise 90% credible interval with the exception of the extreme right tail.

The remaining 5 panels plot the estimated conditional expectation function of various action and outcomes probabilities as a function of particular covariates. Again these predicted probabilities are obtained by allowing a particular x variable to range from -4 to 4, while the other variables are set to zero. To get a sense of how well our model performs relative to a logit QRE specification we also calculated the MLEs of the logit QRE model and superimposed the appropriate conditional expectation functions based on these MLEs. As we would expect given the accuracy of the link function estimates, the estimated conditional expectation functions from the semiparametric model are very close to the true conditional expectation functions. This is true for both expected actions as well as expected outcomes. On the other hand, the logit estimates are, at times, quite far from the true functions. For instance, when looking at the conditional expectation of Down given $\boldsymbol{x}_{r1}$ the logit probabilities are off by 0.1 for large ranges of $\boldsymbol{x}_{r1}$. Similar amounts of bias are evident in the plots of $Y = 2$ and $Y = 3$.

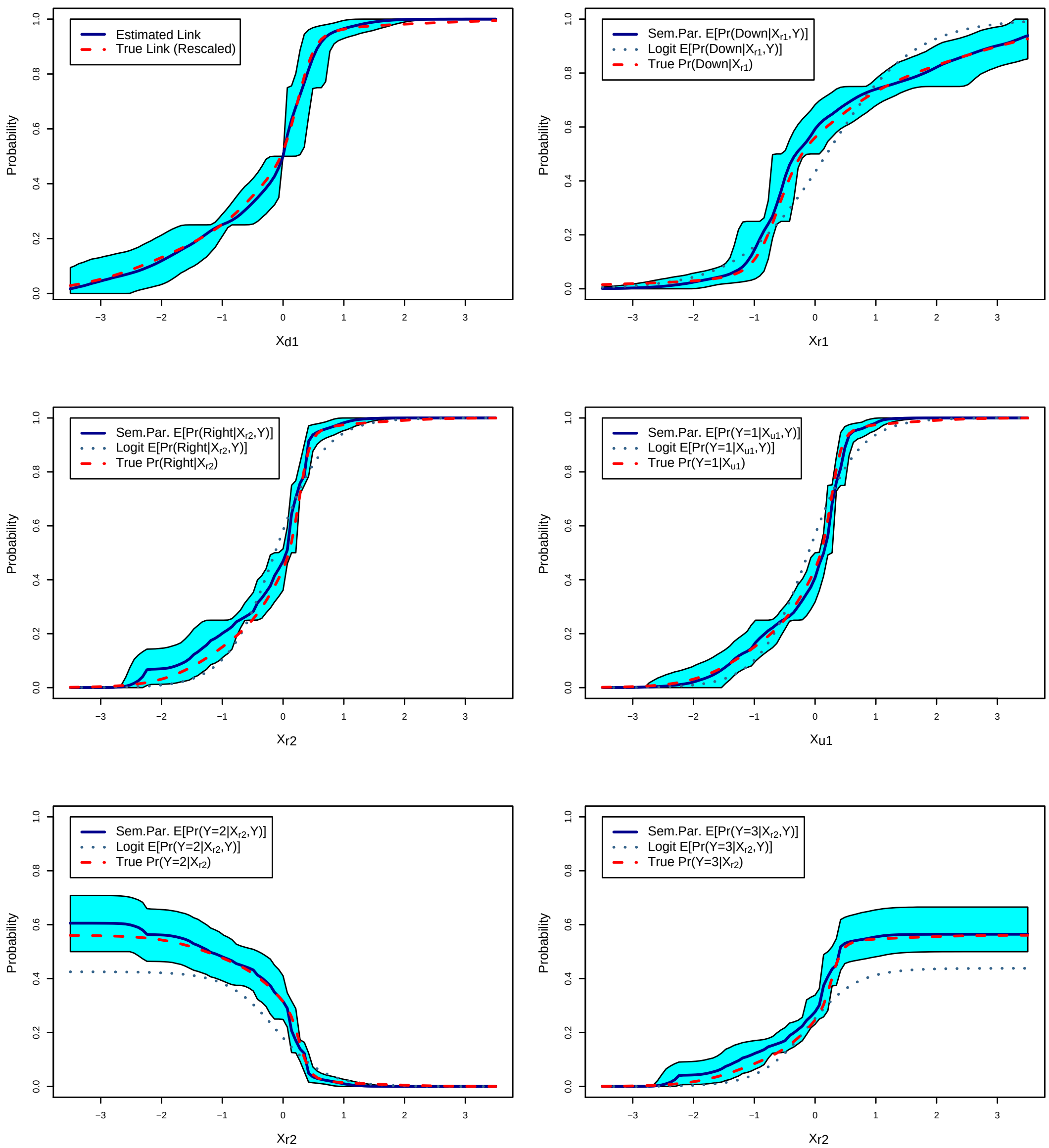

Figure 7: *Results from the semiparametric fit to the simulated data example.*

6. Discussion

In this paper, we have presented a framework for Bayesian semiparametric inference for QRE models. The approach offers numerous advantages such as greater robustness to misspecification of the stochastic portion of random utility functions and the ability to actually estimate the distribution of the difference of the latent disturbances (up to scale and location). While our method is somewhat computationally expensive, it does, based on our limited experience, appear to work well and to be numerically stable.

We are very much in favor of the recent movement among social scientists to derive statistical models directly from theoretical models. Nonetheless, we feel it is extremely important, whether one's goal is purely point estimation or hypothesis testing, that the statistical model be as accurate a representation of the theoretical model as possible. This process involves making very specific assumptions derived directly from the theoretical model where possible and, just as importantly, it also involves making the weakest possible assumptions when the theoretical model is silent regarding some aspect of the statistical model. Failure to be faithful to the theoretical model in these ways can lead to biased estimates and tests *even when the underlying theoretical model is correct.* We see our method as one way to achieve such faithfulness to the theoretical model.

Currently, we are extending our methodology to dynamic choice (extensive form) games with non-binary action sets, as well as simultaneous choice (normal form) games for binary and non-binary action sets (which require the link functions to be absolutely continuous to satisfy the QRE existence proofs).

References

Altman, Micah & Michael P. McDonald. 2003. "Replication with Attention to Numerical Accuracy." *Political Analysis* 11:302–307.

Ben-Akiva, Moshe & Steven R. Lerman. 1985. *Discrete Choice Analysis.* The MIT Press.

Camerer, Colin & Keith Weigelt. 1988. "Experimental Tests of a Sequential Equilibrium Reputation Model." *Econometrica* 56:1–36.

Chen, Hsiao-Chi, James W. Friedman & Jacques-François Thisse. 1997. "Boundedly Rational Nash Equilibrium: A Probabilistic Choice Approach." *Games and Economic Behavior* 18:32–54.

Czado, Claudia & Thomas J. Satner. 1992. "The effect of link misspecification on binary regression inference." *Journal of Statistical Planning and Inference* 45(November):213–231.

Enelow, James & Melvin Hinich. 1984. *The Spatial Theory of Voting: An Introduction.* Cambridge: Cambridge University Press.

Erikson, Robert S. & David W. Romero. 1990. "Candidate Equilibrium and the Behavioral Model of the Vote." *American Political Science Review* 84:1103–1126.

Ferguson, Thomas S. 1973. "A Bayesian Analysis of Some Nonparametric Problems." *The Annals of Statistics* 1:209–230.

Ferguson, Thomas S. 1974. "Prior Distributions on Spaces of Probability Measures." *The Annals of Statistics* 2:615–629.

Fiorina, Morris P. & Charles R. Plott. 1978. "Committee Decisions Under Majority Rule: An Experimental Study." *American Political Science Review* 72:575–598.

Ishwaran, Hemant & Lancelot F. James. 2001. "Gibbs Sampling Methods for Stick-Breaking Priors." *Journal of the American Statistical Association* 96:161–173.

Ishwaran, Hemant & Mahmoud Zarepour. 2000. "Markov Chain Monte Carlo in Approximate Dirichlet and Beta Two-Parameter Process Hierarchical Models." *Biometrika* 87:371–390.

Ishwaran, Hemant & Mahmoud Zarepour. 2002. "Exact and Approximate Sum-Representations for the Dirichlet Process." *Canadian Journal of Statistics* 30:269–283.

McKelvey, Richard D. & Peter C. Ordeshook. 1990. Information and Elections: Retrospective Voting and Rational Expectations. In *Experimental Foundations of Political Science*, ed. Donald R. Kinder & Thomas R. Palfrey. Ann Arbor: University of Michigan Press.

McKelvey, Richard D. & Thomas R. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior* 10(July):6–38.

McKelvey, Richard D. & Thomas R. Palfrey. 1996. "A Statistical Theory of Equilibrium in Games." *The Japanese Economic Review* 47(2):186–209.

McKelvey, Richard D. & Thomas R. Palfrey. 1998. "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics* 1(1):9–41.

Mebane, Walter R. 2000. "Coordination, Moderation, and Institutional Balancing in American Presidential and House Elections." *American Political Science Review* 94:37–57.

Morton, Rebecca B. 1999. *Methods and Models: A Guide to the Empirical Analysis of Formal Models in Political Science*. Cambridge: Cambridge University Press.

Nagler, Jonathan. 1994. "Scobit: An Alternative Estimator to Logit and Probit." *American Journal of Political Science* 38:230–255.

Newton, Michael A., Claudia Czado & Rick Chappell. 1996. "Bayesian Inference for Semiparametric Binary Regression." *Journal of the American Statistical Association* 91:142–153.

Quinn, Kevin M. & Andrew D. Martin. 2002. "An Integrated Computational Model of Multiparty Electoral Competition." *Statistical Science* 17:405–419.

Segal, Jeffrey A., Charles M. Cameron & Albert D. Cover. 1992. "A Spatial Model of Roll Call Voting: Senators, Constituents, Presidents, and Interest Groups in Supreme Court Nominations." *American Journal of Political Science* 36:96–121.

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93(2):279–297.

Signorino, Curtis S. & Kuzey Yilmaz. 2003. "Strategic Messpecification in Regression Models." *American Journal of Political Science* 47:551–566.

Walker, Stephen & Jonathan Wakefield. 1998. "Population Models With a Nonparametric Random Coefficient Distribution." *Sankhya, Series B* 60:196–214.