Spring 2012

Finite Sample Performance of Semiparametric Binary Choice Estimators

Sean Grover
University of Colorado Boulder, sean.grover@colorado.edu

Follow this and additional works at: https://scholar.colorado.edu/honr_theses

Recommended Citation
https://scholar.colorado.edu/honr_theses/274

This Thesis is brought to you for free and open access by Honors Program at CU Scholar. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
Finite Sample Performance of Semiparametric Binary Choice Estimators

Sean Grover
Undergraduate Economics Honors Thesis
University of Colorado
Boulder, CO
sean.grover@colorado.edu

Advisor:
Carlos Martins-Filho
Department of Economics
University of Colorado
Boulder, CO
carlos.martins@colorado.edu

Committee:
Martin Boileau
Department of Economics
University of Colorado
Boulder, CO
martin.boileau@colorado.edu

Sergei Kuznetsov
Department of Mathematics
University of Colorado
Boulder, CO
kuznetsov@colorado.edu

April 17, 2012

Abstract: Strong assumptions needed to correctly specify parametric binary choice probability models make them particularly vulnerable to misspecification. Semiparametric models provide a less restrictive approach with estimators that exhibit desirable asymptotic properties. This paper discusses the standard parametric binary choice models, Probit and Logit, as well as the semiparametric binary choice estimators proposed in Ichimura (1993) and Klein and Spady (1993). A Monte Carlo study suggests that the semiparametric estimators have desirable finite sample properties and outperform their parametric counterparts when the parametric model is misspecified. The semiparametric estimators show only moderate efficiency loss compared to correctly specified parametric.

Acknowledgements: I especially wish to thank Professor Martins-Filho for making this project possible by continually pushing my abilities and knowledge to new heights. The undergraduate economics honors program provided an excellent basis for research and I would like to thank Professor Boileau and Professor McKinnish for their assistance and involvement in the process. I thank my great family and friends for constant support and encouragement.
1 Introduction

Modeling economic topics such as labor force participation, employment status, and home ownership involves dependent variables qualitative in nature. Examination of relationships concerning qualitative choice originally came forth from the biological sciences where experiments often measure a subject’s tolerance to treatment as related to some underlying threshold (Gourieroux, 2000), (Cramer, 1991). Study of qualitative economic variables related to some underlying threshold give rise to models of discrete choice where variables lacking natural order are coded into quantitative categories (Cramer, 1991). Coefficients of such models inherently imply probabilistic interpretations. For example, a researcher may be interested in a borrower’s probability of mortgage default as determined by covariates like credit score, loan-to-value ratio and income. This mortgage default question relies on a coded binary dependent variable: $y = 1$ if a default occurs and $y = 0$ if not.

While standard regression models can be specified for problems such as default probability, two probability models have come to dominate the field concerning binary discrete choice. The standard parametric Probit and Logit models see wide use in applied research concerning questions involving binary dependent variables (Cramer, 1991). Although popular, these two models rely on particularly stringent assumptions for specification. Under misspecification, the parametric models produce inconsistent estimates (Li and Racine, 2006). A few semiparametric binary choice models have been developed under weaker assumptions to account for the likely inability to correctly specify a parametric model. This paper will examine estimators for binary choice models, including Probit and Logit, as well as two semiparametric, namely the semiparametric least squares (SLS) estimator proposed in Ichimura (1993) and the semiparametric maximum likelihood (SML) estimator proposed in Klein and Spady (1993). A Monte Carlo study provides the basis of examination through comparison of each estimate’s bias and mean squared error (MSE) under different data generating processes.
The goal of conducting this Monte Carlo study is to answer the question: what comments can be made concerning the finite sample performance of semiparametric and parametric binary choice estimators under controlled conditions? Specifically, we make comparisons between semiparametric and parametric estimates in two situations: 1. under correct specification of the parametric model, and 2. under misspecification of the parametric model. The Monte Carlo study also facilitates an evaluation of the estimator of Ichimura compared to that of Klein and Spady.

Throughout this study, we expect the performance of semiparametric estimators to improve with increased sample size as these estimators’ properties rely on asymptotic nonparametric density results (Li and Racine, 2006). As previous work and theory suggests, we expect correctly specified parametric estimators to outperform the semiparametric estimators, specifically under small sample size, but do also propose that semiparametric efficiency losses are small in scale. In the case of misspecified parametric form, we expect dramatically improved semiparametric estimates. We reserve prior judgements on comparisons between the two semiparametric estimators.

Some previous Monte Carlo studies have examined finite sample performance of semiparametric and parametric binary choice estimators. Ichimura (1993) as well as Klein and Spady (1993) both evaluate the finite sample performance of their proposed estimators through simulations. Both of these studies find the semiparametric estimators to outperform a misspecified Probit and underperform a correctly specified Probit with modest loss in efficiency. DeLuca (2008) compares various semiparametric estimators, including Klein and Spady SML, to parametric Probit. He concludes on similar results in examination under correctly specified and misspecified Probit cases and suggests use of semiparametric models. Noting the limited implementation of semiparametric binary choice estimators in applied settings, Horowitz (1993) as well as Gerfin (1998) compare Probit results to select semipara-
metric estimates including those of Klein and Spady. Results from both of these works also suggest use of semiparametric estimators in applied research.

This study contributes to the literature by comparing parametric and semiparametric estimators under broad experimental conditions. Due to difficulties in designing a Monte Carlo study and correct implementation of the choice semiparametric estimators, no extensive study has compared the four discussed estimation procedures. In fact, no study has compared the estimator of Ichimura to that of Klein and Spady either. Under a general DGP, we hope to effectively evaluate finite sample performance of the standard parametric estimates from Probit and Logit, as well as the semiparametric estimators of Ichimura and Klein and Spady. The remaining sections of this study follow as such: section 2 provides a theoretical description of the choice models and estimation techniques, section 3 outlines the DGP and estimation procedures used in the Monte Carlo experiments, and section 4 outlines the Monte Carlo results and evaluates finite sample performance of the selected binary choice estimators.

2 Binary Choice Single Index Models

Applied work in economics commonly seeks answers to questions regarding a qualitative outcome. Interest may concern a homeowner’s decision to default on their mortgage or a firm’s decision in issuing corporate debt. In these cases, the researcher wishes to understand the processes involved in such discrete choices as determined by certain independent variables. This paper specifically examines the case of a binary dependent variable where \( y_i \) takes only values of 0 or 1. Analyzing the relationship between covariates \( x_i \) and Bernoulli random variable \( y_i \) is inherently a study of response probability.

No rule prohibits examining such probability relationships in a standard regression setting. Take the linear model \( y_i = x_i \beta + u_i \), with \( \{y_i, x_i\}_{i=1}^n \) under classical linear regression
assumptions. Several issues arise out of this linear consideration. Notice the ambiguity in a discrete left hand side equated with a continuous right hand side. Also, the parameter estimates will reflect constant changes in response probability corresponding to unit changes in an independent variable, not likely in most probability settings (Gourieroux, 2000). Also, unless restrictions are placed on $\beta$, the parameter estimates may imply probabilities outside the interval $[0,1]$. More effective probability models exist.

In a discrete choice setting, consider an underlying regression on a continuous unobserved latent variable $y_i^*$ influenced by observed independent variables $x_i$: $y_i^* = x_i \beta + \varepsilon_i$. The researcher only observes the binary $y_i$ which takes value based on a threshold defined by the latent variable $y_i^*$:

$$y_i = \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{if } y_i^* < 0 \end{cases}$$

(1)

**Assumptions** 1. $\{x_i, y_i\}_{i=1}^n$ is independent and identically distributed (i.i.d.). 2. $\varepsilon_i$ is i.i.d. and has some symmetric Cumulative Distribution Function (CDF) denoted $F$. 3. $E[\varepsilon_i|x_i] = 0$ as well as $\varepsilon_i$ and $x_i$ independent.

Under these assumptions we move forward in deriving the condition mean of binary $y$:

$$E[y_i|x_i] = 1 \times P(y_i = 1|x) + 0 \times P(y_i = 0|x) = P(y_i^* \geq 0)$$

(2)

$$= P(x_i \beta + \varepsilon \geq 0) = P(\varepsilon \geq -x_i \beta)$$

$$= 1 - P(\varepsilon \leq -x_i \beta)$$

$$= 1 - F(-x_i \beta)$$

$$= F(x_i \beta)$$

(3)

Thus, the conditional mean of $y_i$ is equal to a function $F$ a CDF, of the scalar index $x_i \beta$ and from this distinction, we define the binary choice single index model:

$$P(y_i = 1|x_i) = E[y_i|x_i] = F(x_i \beta)$$

(4)
2.1 Parametric Estimation

Notice that in the derivation of the binary choice single index model, we do not specify a form for \( F \), the CDF of \( \varepsilon_i \). We only make the assumption that \( F \) exists and is symmetric.

Under parametric estimation, the researcher specifies \( F \) to guarantee identifiability of \( \beta \). The most common choices for \( F \) lead to the popular Probit or Logit probability models.

**Probit:** Consider the distributional assumption: \( \varepsilon_i \sim N(0, 1) \). Then the error term has standard normal distribution with \( \mu = 0 \) and \( \sigma^2 = 1 \) yielding:

\[
F(x_i' \beta) = \Phi(x_i' \beta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_i' \beta} e^{-\frac{1}{2}t^2} dt.
\]

Estimation under this assumption results in the Probit model: \( E(y_i|x_i) = \Phi(x_i' \beta) \) with Probit parameter estimates \( \hat{\beta} \).

**Logit:** Consider the assumption that the \( \varepsilon_i \) has standard logistic distribution with \( \mu = 0 \) and \( \sigma^2 = \pi^2/3 \) yielding:

\[
F(x_i' \beta) = \Lambda(x_i' \beta) = \frac{1}{1 + e^{-x_i' \beta}}.
\]

Estimation under this assumption results in the Logit model: \( E(y_i|x_i) = \Lambda(x_i' \beta) \) with Logit parameter estimates \( \hat{\beta} \).

**Maximum Likelihood Estimation:** Moving forward in parametric estimation of the parameter vector \( \beta \), we again use \( F \) notation where we have now assumed that \( F = \Phi \) or \( \Lambda \) for the Probit and Logit cases respectively. Under i.i.d. assumptions, the product of individual densities defines the likelihood equation:

\[
L(\beta) = f(y_1, \ldots, y_n | \beta) = \prod_{i=1}^{n} f(y_i | \beta)
\]

where the likelihood equation is a function of the unknown parameter vector \( \beta \) with given data \( \{y_i, x_i\}_{i=1}^{n} \). To ensure an estimable parameter vector we need identification: the pa-

---

\(^1\)In the parametric Probit and Logit cases, a constant term \( \alpha \) could be added to the model yielding \( F(\alpha + x_i' \beta) \) but this case is not considered in this paper for reasons involving the semiparametric case discussed later.
rameter vector $\beta$ is identified if for any other parameter vector $\beta^* \neq \beta$, for some data, $\mathcal{L}(\beta^*) \neq \mathcal{L}(\beta)$ (Greene, 2003). Identification of $\beta$ is established under both distributional assumptions. In a binary choice setting with $F$, the likelihood equation becomes:

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} F(x_i'\beta)^{y_i} (1 - F(x_i'\beta))^{1-y_i}. \quad (8)$$

We take the natural logarithm of $\mathcal{L}(\beta)$ to arrive at the log-likelihood function:

$$\ln \mathcal{L}(\beta) = \sum_{i=1}^{n} (y_i \ln[F(x_i'\beta)] + (1 - y_i) \ln[1 - F(x_i'\beta)]). \quad (9)$$

Under this framework, Maximum Likelihood estimation (MLE) provides a means of choosing parameter estimates that will maximize the likelihood of occurrence for the observed data. The MLE estimator $\hat{\beta}$ is defined as:

$$\hat{\beta} = \arg\max_{\beta} \ln \mathcal{L}(\beta). \quad (10)$$

Note that closed form expressions of these maximum likelihood estimator is not available because the likelihood equation is non-linear in parameters. Iterative numerical optimization provides a means of approximating solutions through the first order conditions.

Asymptotic properties make MLE quite desirable. In our parametric setting, if the distributional assumptions on $\varepsilon_i$ are correctly asserted and certain regularity conditions are met then the MLE $\hat{\beta}$ will obey the following:

**Properties of Maximum Likelihood Estimates:**

- Consistent: $\text{plim} \ \hat{\beta} = \beta$
- Asymptotically Efficient: $\text{Var}(\hat{\beta})$ achieves Cramer – Rao Lower Bound
- Asymptotically Normal: $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N[0, (-E[\partial^2 \ln L/\partial \beta \partial \beta']^{-1})]$

(Greene, 2003) \hspace{1cm} (11)

The necessary regularity conditions as well as proofs of these properties will not be presented in this paper. For formal definitions and proofs refer to Greene (2003) Chapter 16. Availability of t-tests and confidence interval construction make these asymptotic properties particularly appealing.
2.2 Semiparametric Approach

Parametric estimation of the binary choice single index model relies on a key and stringent assumption concerning the distribution of $\varepsilon$. Different distributional assumptions change the conditional mean of $Y$, so under misspecification of the error term’s distribution, the estimates may be biased, inconsistent and their asymptotic distribution may be unknown. Thus, the rigid assumptions of the parametric approach come at great risk. To avoid masking important features of the data while obtaining robust estimates, we seek a more flexible approach. In modeling the conditional mean function, we need not specify the distribution of the error term while still maintaining identifiability of the model and consistency of parameter estimates (Ichimura and Todd, 2007). The semiparametric single index model:

$$E[y_i|x_i] = g(x_i'\beta).$$

This model receives semiparametric classification because while not specifying a functional form for the distribution function $g$, the model still relies on the parametric index $x_i'\beta$. In this less restrictive setting, an estimate for the unknown link function $g$ will be needed before parameter estimates may be obtained.

Estimation procedures of the choice semiparametric binary choice estimators presented shortly rely on a kernel estimator of $g$. Common kernel estimators for the unknown link function are based on local constant averaging. We estimate the conditional mean function by constructing a locally weighted average of those $y_i$ associated with values of the covariates $x_i$ on an interval of chosen length. Construction of the weights means determining the “closeness” of $x_i$ to $x$ in a local window defined by the bandwidth $h$. Then, evaluation by the kernel function $K$ gives greater weight to those dependent variables which are close in terms of the independent variables (Li and Racine, 2006). The kernel function $K$ is continuous, symmetric about 0 and integrates to 1. Therefore, $K$ provides weight, shape, and continuity to the conditional expectation. A general Nadaraya Watson kernel estimator
for $E[y_i|x_i]$ is given by:

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{(x_i-x)}{h}\right) y_i}{\sum_{i=1}^{n} K\left(\frac{(x_i-x)}{h}\right)}.$$  \hspace{1cm} (13)

Nonparametric estimation heavily relies on bandwidth selection. Choosing a bandwidth too large results in an oversmoothed regression estimator with large bias and small variance that obscures key features of the data. A bandwidth too small results in an undersmoothed regression estimator with small bias and large variance that may have many jumps and local extrema which make parameter estimation difficult (Pagan and Ullah, 1999). Bandwidth selection will be discussed along with corresponding semiparametric estimators.

Before moving forward in discussion of specific semiparametric estimators, we require certain identifiability conditions to ensure an estimable function $g$ and unknown parameter vector $\beta$:

**Identification:**

1. The vector $x_i$ must not contain a constant and must contain at least one continuously distributed variable. $\beta$ must be normalized: $||\beta|| = 1$.
2. $g$ is continuous, differentiable and not constant in the domain of $x_i'\beta$.
3. Varying the discrete components of $x_i$ will not divide the support of $x_i'\beta$ into disjoint subsets.
   (Pagan and Ullah, 1999)

We see that $\beta$ can contain no location parameter and is only identifiable up to a scale so we impose the scale normalization in assuming that $\beta$ has unit length. At least one continuous $x_i$ will ensure existence of a specific $g$ for estimation on the support of $x_i'\beta$. Conditions on $g$ identify a unique $\beta$.

**Ichimura’s Estimator:** From $E[y_i|x_i] = g(x_i'\beta)$, we have $y_i = g(x_i'\beta) + \varepsilon_i$ and then $\varepsilon_i = y_i - g(x_i'\beta)$. Ichimura motivates the estimation procedure with several key observations: (i) The variation in $y$ results from both the variation in $g(x_i'\beta_0)$ and $\varepsilon$, where $\beta_0$ is the true parameter. (ii) In consideration of the contour line $g(x_i'\beta_0) = c$, the variation in $y$ results solely from the variation in $\varepsilon$. (iii) Under $g(x_i'\beta) = c$ with $\beta \neq \beta_0$, the value of $g(x_i'\beta_0)$
changes, so again the variation in \( y \) results from variation in both \( g(x' \beta_0) \) and \( \varepsilon \).

From this, Ichimura identifies \( \beta_0 \) through the conditional variance:

\[
\text{Var}[y|g(x' \beta) = c] = E[y - E[y|g(x' \beta) = c]^2|g(x' \beta) = c] \tag{14}
\]

which measures the variation in \( y \) on each contour line \( c \) for which \( g(x' \beta) = c \). Ichimura devises a semiparametric estimation model built on a sample analog of this conditional variance which determines parameter estimates by minimizing the objective function:

\[
SLS(\beta) = \sum_{i=1}^{n} [y_i - g(x'_i \beta)]^2 \tag{15}
\]

with respect to \( \beta \). As in the maximum likelihood case, placing a distributional assumption on \( \varepsilon \) makes the objective function that of nonlinear least squares. In semiparametric form, without assuming a known \( g \), Ichimura calls the estimation Semiparametric Least Squares (SLS)\(^2\). Thus, \( g(x'_i \beta) \) must be estimated. Notice that not only is \( g \) unknown but so is \( \beta \). Therefore, kernel density methods do not estimate \( g(x'_i \beta_0) \) directly but rather for a given \( \beta \), we can find an estimate for \( E[y_i|x'_i \beta] = E[g(x'_i \beta_0)|x'_i \beta] \) (Li and Racine 2007). Ichimura proposes use of the leave-one-out Nadaraya Watson estimator, \( \hat{g}_{-i}(x'_i \beta) \). The leave-one-out method omits \( x_i \) and estimates \( g(x'_i \beta) \) using all other observations (Pagan and Ullah, 1999).

\[
\hat{g}_{-i}(x'_i \beta) = \frac{\sum_{j \neq i}^{n} K \left( \frac{(x_j - x_i)' \beta}{h} \right) y_j}{\sum_{j \neq i}^{n} K \left( \frac{(x_j - x_i)' \beta}{h} \right)} \tag{16}
\]

Asymptotic theory requires the use of the leave-one-out method. Hall et al. (1993) suggest choosing \( h \) by cross validation in the SLS case. The bandwidth is introduced as a free parameter to the SLS criterion and then selected simultaneously with \( \hat{\beta} \). Because the denominator of \( \hat{g}_{-i}(x'_i \beta) \) may be near 0, Ichimura introduces a trimming set \( 1(x_i \in A_n) \): where \( A_n = \{ x : ||x - x^*|| \leq 2h \text{ for some } x^* \in A_\delta \} \) and \( A_\delta = \{ x_i : \hat{f}(x'_i \beta) \geq \delta \text{ for all } \beta \in B \} \) with

\(^2\)Notice that this derivation does not rely on specification of a binary dependent variable. Ichimura SLS may use a continuous or qualitative \( y \).
δ > 0, density estimate $\hat{f}$, and $B$ is the set where $\beta$ takes values. Then, $\hat{\beta}$ is determined by minimizing $SLS(\beta, h)$ with respect to $\beta$.

$$SLS(\beta, h) = \sum_{i=1}^{n} [y_i - \hat{g}_i(x_i' \beta, h)]^2 1(x_i \in A_n)$$  \hspace{1cm} (17)

Note the kernel estimator, a function of $\beta$, nested within the criterion function. The resulting function depends only on the parameters $\beta$ and $h$ for minimization. Under assumptions regarding $K$ and regularity conditions, Ichimura (1993) shows consistency and asymptotic normality of the SLS estimator of $\beta$. For a formal outline of assumptions and proofs of these properties, refer to Ichimura (1993) or Li and Racine (2006) Chapter 8. Again, note that no closed form expression for $\hat{\beta}$ is available due to non-linearity of SLS on the parameter $\beta$. Iterative numerical optimization provides a means of solving the SLS minimization criterion.

**Klein and Spady’s Estimator:** Specifically under a binary choice setting, Klein and Spady (1993) propose a semiparametric single index model:

$$P(y_i = 1|x_i) = E[y_i|x_i] = g(x_i' \beta)$$  \hspace{1cm} (18)

and find parameter estimates by employing maximum likelihood estimation. The nonparametric component of their model also places weak assumptions on error distribution and requires estimation of $g(x_i' \beta)$. Originally, Klein and Spady suggest either a bias reducing kernel or adaptive kernel with local smoothing. In practice, as with Ichimura SLS, the leave-one-out Nadaraya Watson kernel estimator is widely used (Pagan and Ullah, 1999). Pagan and Ullah point out that in a binary choice setting, the numerator of $\hat{g}_i(x_i' \beta)$ is an estimator of $g(x_i' \beta)$ conditional on $y_i = 1$ while the denominator estimates the unconditional density. With this density estimation technique, the bandwidth may be selected again by cross validation. Our use of cross validated bandwidth in the Klein and Spady estimator is a new method as the literature does not reflect use of this technique. Introduction of bandwidth as a free parameter results in maximization of the log-likelihood function in terms of both $\beta$
and $h$ simultaneously.

$$\ln \mathcal{L}(\beta, h) = \sum_{i=1}^{n} y_i \ln[\hat{g}_{-i}(x_i' \beta, h)] + (1 - y_i) \ln[1 - \hat{g}_{-i}(x_i' \beta, h)].$$ (19)

Thus $\hat{\beta}$ is the SML estimate of the parameter vector $\beta$ which maximizes the likelihood of occurrence for the observed data. Again, the kernel estimator has been nested within the log-likelihood function to make the resulting function depend only on $\beta$ and $h$. Klein and Spady also propose introduction of a trimming set. While aiding in asymptotic proofs, they note that the trimming set seems to not matter in applied settings so we do not discuss it here. With trimming, under assumptions regarding $K$, and regularity conditions, Klein and Spady show that their SML estimator is consistent, semiparametrically efficient and asymptotically normal. For a formal outline of assumptions and proofs of these properties, refer to Klein and Spady (1993).

3 Monte Carlo Design

Monte Carlo methods provide a means of implementing estimation techniques repeatedly under controlled conditions to evaluate finite sample performance. This Monte Carlo study aims to compare the semiparametric binary choice estimators proposed in Ichimura (1993) and Klein and Spady (1993) respectively, to standard parametric Probit and Logit estimators. Estimated bias and MSE provide the basis for comparison under finite sample limitations with two cases in mind: 1. parametric model having correctly specified error term distribution, and 2. parametric model having misspecification of error term distribution. This section reports the development and implementation of the DGP used in Monte Carlo simulations.

With finite sample performance in mind, we consider three modest sample sizes in the experiments: $n_1 = 250, n_2 = 500$ and $n_3 = 750$. The latent model takes the form:

$$y_i^* = x_{i1} \beta_1 + x_{i2} \beta_2 + \varepsilon_i$$ (20)
The covariate $x_{1i}$ takes values from a standard uniform distribution on the interval $[0,1]$ with $\mu = 1/2$ and $\sigma^2 = 1/12$. The covariate $x_{2i}$ takes values from a standard normal distribution where $\mu = 0$ and $\sigma^2 = 1$. We also scale the variance of $x_{2i}$ with another experiment parameter: $\sigma_{x_2}^2 = 1/2$ and $\sigma_{x_2}^2 = 1/5$. Importantly, we specify the true parameter values: $\beta_1 = \sqrt{3}/2$ and $\beta_2 = 1/2$.

This study considers three experimental cases for the distribution of $\varepsilon_i$. In the first case, we generate the error term with random numbers from a standard normal distribution where $\mu = 0$ and $\sigma^2 = 1$. In experiments involving this case, the Probit model will be correctly specified. Case 2 generates the error term with random numbers from a standard logistic distribution with $\mu = 0$ and $\sigma^2 = \pi^2/3$. Here, the Logit model will be correctly specified. The third case for distributional form of the error term takes values from a student t distribution with $\mu = 0$ and degrees of freedom $\nu = 9$ which leads to $\sigma^2 = 9/7$. Degrees of freedom are kept low to make a clear distinction between standard normal and student t distributions. In the case of student t distributed error term, neither the Probit nor Logit model will have correct specification. We also scale the variance of these error terms with another experiment parameter: $\sigma_{\varepsilon_1}^2 = 1/3$ and $\sigma_{\varepsilon_2}^2 = 1$.

Under these experimental designs, we classify the observable $y_i$ as $1$ if $y_i^* \geq 0$ or $0$ if $y_i^* < 0$. Notice that this data generating process has satisfied the assumptions of $\{y_i, x_i\}_{i=1}^n$ i.i.d. and $\varepsilon_i$ i.i.d. with $E[\varepsilon_i|x_i] = 0$. We have also ensured specification by only considering error distributions with $\mu = 0$. Specifically for semiparametric identification, we specify the true parameters such that $||\beta|| = 1$. Both $x_1$ and $x_2$ have continuous distribution with $x_2 \in (-\infty, \infty)$ which satisfies the remaining semiparametric identification conditions. Estimation of the hypothetically unknown parameter vector $\beta$ can proceed with data assumptions and identification conditions in place for both parametric and semiparametric estimation procedures.

The DGP as well estimation procedures are carried forth in MATLAB. Permutations
between sample size (3), distribution of $\varepsilon_i$ (3), variance of $x_2$ (2), and variance of $\varepsilon_i$ (2) makes 36 total experiments. Each experiment consists of 500 repetitions of DGP and estimation. The Leave-one-out Nadaraya Watson kernel estimator employed the Gaussian Kernel: $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$. Due to unavailability of closed form solutions for parameter estimates in all estimation procedures, iterative numerical algorithms are employed to provide approximations. Parametric optimization procedures use OLS parameter estimates as starting values. Semiparametric optimization begins with initial estimates of $\frac{1}{\sqrt{2}}$ for both $\beta_1$ and $\beta_2$ (normalized initial estimates). In the semiparametric cases of Ichimura and Klein and Spady, an initial bandwidth $h = 1$, is concatenated with the matrix of initial parameter estimates prior to optimization for cross validated selection.

4 Monte Carlo Results

This section analyzes estimated bias and MSE in aggregate. Each DGP and estimation procedure calculates bias and MSE for both $\hat{\beta}_1$ and $\hat{\beta}_2$. We estimate the bias and MSE by taking the sample mean across all 500 iterations of each experiment. Discussion of results is organized by experimental parameter with special consideration placed upon error term distribution as designated by our goal of comparing semiparametric estimates to correctly specified and misspecified parametric estimates. Full tables of results follow in the first appendix.

As indicated in Li and Racine (2006), the semiparametric estimates should improve as sample size increases, as should the parametric. Under $\varepsilon_i \sim N(0,1)$, bias of semiparametric estimates for both $\beta_1$ and $\beta_2$ generally improve as sample size increases as do the parametric estimates. MSE of semiparametric and parametric estimates always improves as sample size increases. Consider the experiment from Table 1 where $n = 250, \sigma_{\varepsilon}^2 = 1$, and $\sigma_{x_2}^2 = \frac{1}{2}$ for $\beta_1$: Ichimura yields MSE of 0.0321 compared to 0.0230 for Probit. Under the same
conditions except \( n = 750 \), Ichimura MSE improves to 0.0096 and Probit MSE improves to 0.0037. When the error term has standard logistic distribution, we find similar results. In most cases, the bias for semiparametric and parametric estimates improves with sample size for both parameters. Similarly, the MSE for semiparametric and parametric estimates always improves with larger \( n \). These same results concerning bias and MSE hold when \( \varepsilon_i \sim t(0, \nu = 9) \). From Table 5, in estimating \( \beta_1 \) when \( \sigma_x^2 = 1 \) and \( \sigma_{x_2}^2 = 1/2 \), as sample size increases from 250 to 500, MSE for Klein and Spady improves from 0.0488 to 0.0096 and MSE for Probit improves from 0.0278 to 0.0110.

In considering the variance parameter \( \sigma_{x_2}^2 \) assigned to the covariate \( x_2 \), results do not indicate a firm conclusion as in the case of sample size. First looking at cases, where \( \varepsilon_i \sim N(0, 1) \), the bias and MSE of semiparametric and parametric estimates for \( \beta_1 \) are generally better when \( \sigma_{x_2}^2 = 1/5 \). In the case of \( \beta_2 \), bias and MSE for both Ichimura as well as Klein and Spady improve when \( \sigma_{x_2}^2 = 1/5 \). But in the parametric cases, while bias is generally better with \( 1/5 \), MSE is not. Under experiments with standard logistic error distribution, estimates of \( \beta_1 \) are typically better when \( \sigma_{x_2}^2 = 1/5 \) for both semiparametric and parametric cases. For \( \beta_2 \) however, while semiparametric bias generally improves with \( \sigma_{x_2}^2 = 1/5 \), parametric bias and MSE of both parametric and semiparametric estimates seem to be better with \( \sigma_{x_2}^2 = 1/2 \). The same results follow from experiments with student t distributed error term. For example, looking at estimates of \( \beta_2 \) when \( n = 750 \) and \( \sigma_x^2 = 1 \) from Table 6, Ichimura bias improves from 0.0817 to 0.0462 when \( \sigma_{x_2}^2 \) changes from \( 1/2 \) to \( 1/5 \) while MSE changes from 0.0163 to 0.0207. Under these conditions, Probit bias declines from -0.0426 to -0.0469 and MSE from 0.0078 to 0.0157. It appears that the semiparametric estimators, less restrictive by design, prefer the smaller variance parameter on an average basis but at a cost of variability in estimates while the parametric estimates generally produce worse estimates under experiments when \( \sigma_x^2 = 1/5 \).

The experiment parameter \( \sigma_x^2 \) produced some interesting results. For estimates of \( \beta_1 \),
bias and MSE for the semiparametric estimates improve when \( \sigma^2 = \frac{1}{3} \). Semiparametric estimates for \( \beta_2 \) see more bias when \( \sigma^2 = \frac{1}{3} \) but better MSE. It appears that overall, the less restrictive semiparametric estimators prefer the smaller variance parameter. In almost all cases, indifferent to error term distribution, when \( \sigma^2 = \frac{1}{3} \), parametric estimates suffered dramatically compared to when \( \sigma^2 = 1 \). The reason for the dramatic decrease in accuracy of parametric estimates when \( \sigma^2 = \frac{1}{3} \) is attributed to model specification. It is important to note that the parametric error term assumptions rely on a specific variance value: \( \sigma^2 = 1 \) in the Probit case and \( \sigma^2 = \frac{\pi^2}{3} \) in the Logit case. Thus, although Probit is correctly specified when \( \varepsilon_i \sim N(0,1) \) and Logit is correctly specified when \( \varepsilon \sim L(0,\frac{\pi^2}{3}) \), by changing the variance of \( \varepsilon \) the models become misspecified by definition. For example, in the case of \( \varepsilon_i \sim N(0,1) \) from Table 1, when the Probit model is correctly specified, with \( n = 750, \sigma^2 = 1 \) and \( \sigma^2 = \frac{1}{2} \) Probit estimates of \( \beta_1 \) show bias of 0.0037 and MSE of 0.0076. But under the same conditions except \( \sigma^2 = \frac{1}{3} \), Probit estimates show bias of 0.6445 and MSE of 0.4260; a severe reduction in accuracy. One exception comes from \( \varepsilon_i \sim L(0,\frac{\pi^2}{3}) \) where Probit estimates improve when \( \sigma^2 = \frac{1}{3} \). Because of the variance change just discussed, the logistic error distribution here actually more resembles that of a standard normal. Notably, with \( \varepsilon_i \sim L(0,\frac{\pi^2}{3}) \) when \( \sigma^2 = \frac{1}{3} \), the Probit estimates are actually more accurate than those of the Logit.

In examining results concerning error term distribution, we expect the semiparametric estimators to outperform misspecified parametric estimators and underperform correctly specified parametric estimators. The results appear to confirm these predictions. When \( \varepsilon_i \sim N(0,1) \), the correctly specified Probit model always produces better estimates than those from the misspecified Logit model. For example from Table 1, when \( n = 750, \sigma^2 = 1 \) and \( \sigma^2 = \frac{1}{5} \), Probit estimates of \( \beta_1 \) show bias of 0.0064 and MSE of 0.0077 while Logit estimates show 0.5615 and 0.3374. Under these same conditions, Ichimura shows bias of -0.0474 and MSE of 0.0090; worse than Probit but dramatically better than Logit. Although
underperforming the correctly specified Probit, the semiparametric estimates do not suffer dramatic efficiency losses and maintain systematic accuracy throughout these experiments.

When $\varepsilon_i \sim L(0, \pi^2/3)$ we expect estimates from the correctly specified Logit model to outperform those of the misspecified Probit and the results indicate this is the case, save the exception discussed earlier. Consider the experiment from Table 4 where $n = 750, \sigma_\varepsilon^2 = 1$ and $\sigma_x^2 = 1/5$. The Logit estimates for $\beta_2$ show bias of 0.0183 and MSE of 00293 while Probit estimates show -0.1800 and 0.0434. Under these conditions, Klein and Spady shows bias of 0.0580 and MSE of 0.0368. Thus, the correctly specified Logit produces the most accurate estimates while the semiparametric Klein and Spady outperforms the misspecified Probit. Overall, the semiparametric estimates produce systematically accurate results throughout these experiments and do not suffer a dramatic loss of efficiency to those of the correctly specified Logit.

With an error term having student t distribution, neither the Probit nor the Logit models has correct specification and we expect the two semiparametric estimators to outperform both parametric. The results generally confirm this. Consider the case from Table 5 where $n = 750, \sigma_\varepsilon^2 = 1$ and $\sigma_x^2 = 1/2$. For $\beta_1$, Ichimura exhibits bias of -0.0621 and MSE of 0.0097. The Logit estimates exhibits bias of 0.4531 and MSE of 0.2255 while Probit has bias of -0.0623 and MSE of 0.0110. Notice that the Probit estimate is quite improved over the Logit and is comparable to that of Ichimura. Probit estimates actually outperform those of Logit in all cases of student t distributed error term. While in most experiments, the semiparametric estimators outperform the misspecified parametric, in some cases with small sample size and $\sigma_\varepsilon^2 = 1$, the Probit even outperforms the two semiparametric. Again this unexpected result follows from a consideration of variance. The DGP used a student t distribution with 9 degrees of freedom in these experiments, which produces a similar distribution to that of standard normal which enables the Probit model to produce fairly accurate estimates despite misspecification. But in consideration of experiments with $\varepsilon_i \sim t(0, \nu = 9)$, the consistently
more accurate semiparametric estimates are favorable to those of the two parametric.

We also wish to make comments and comparisons concerning estimates of Ichimura and those of Klein and Spady. Notice first from Tables 1, 3, and 5 that both semiparametric estimators produce negatively biased estimates for $\beta_1 = \sqrt{3}/2$. From Tables 2, 4, and 6, both semiparametric estimators produce positively biased estimates for $\beta_2 = 1/2$. These bias signs result from the optimization procedure and not from the estimators themselves. The iterative optimization began with normalized initial estimates of $1/\sqrt{2}$ for both parameters which means the initial estimates lay between $1/2$ and $\sqrt{3}/2$. The bias signs imply that the optimization procedure ended before overshooting the true parameter value. In consideration of all experiments, while both semiparametric estimators showed systematically low bias and MSE, Ichimura estimates displayed more accuracy than those of Klein and Spady on an absolute basis. The only experiments where Klein and Spady estimates consistently outperform those of Ichimura come when $\sigma_\varepsilon^2 = 1/3$ and $\sigma_{x_2}^2 = 1/5$. Although select experiments see lower bias and/or MSE from Klein and Spady, as a whole, the estimates from Ichimura are consistently improved. It is worth noting that the differences in bias and MSE observed between estimates of Ichimura and those of Klein and Spady are not so dramatic as to promote choice of one estimator over the other.

With the motivation behind this Monte Carlo study in mind, we have found results in agreement with what we expected from the DGP and estimation techniques. Under misspecification of the parametric models, the semiparametric approaches almost always produce more accurate estimates. With correct parametric specification, the semiparametric estimates do see a loss in efficiency but none so dramatic as to preclude their use. In all experiments, the estimators of Ichimura and Klein and Spady consistently exhibit small bias and MSE. This, along with the likely situation in applied settings where a parametric model cannot be correctly specified, makes the estimators of Ichimura as well as Klein and Spady especially appealing compared to the standard parametric Probit and Logit.
5 Conclusion

This paper has outlined and discussed binary choice estimators under more restrictive parametric and less restrictive semiparametric settings. With asymptotic properties of said estimators proven in previous literature, we examined their finite sample performance through implementation of a Monte Carlo simulation. On a basis of bias and MSE, the expected results hypothesized, based on theory, see confirmation through this study. The semiparametric estimators provide consistent estimates throughout the experiments dominating the misspecified parametric models and underperforming correctly specified parametric models at a modest loss of efficiency. Results indicate that semiparametric estimates improve with increased sample size. It appears that the model of Ichimura, in general, outperforms that of Klein and Spady in a finite sample setting, although not dramatically in an absolute sense. The results of this finite sample evaluation indicate that the semiparametric estimators of Ichimura as well as Klein and Spady have a promising future in applied economic problems involving binary choice.

Future study of finite sample performance may wish to keep values of variance parameters closer to 1 to prevent egregious misspecification of parametric models. As this study has compared the semiparametric estimator of Ichimura to that of Klein and Spady, future research might make comparisons to other semiparametric binary choice estimators such as that of Lewbel (2000).
### Appendix 1: Tables

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\sigma^2_{\epsilon}$</th>
<th>$\sigma^2_{\tau_2}$</th>
<th>Ichimura</th>
<th>Klein and Spady</th>
<th>Probit</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>Bias</td>
<td>-0.0760</td>
<td>-0.0800</td>
<td>0.0179</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0321</td>
<td>0.0343</td>
<td>0.0230</td>
</tr>
<tr>
<td>250</td>
<td>1</td>
<td>$\frac{1}{5}$</td>
<td>Bias</td>
<td>-0.0582</td>
<td>-0.0591</td>
<td>0.0079</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0354</td>
<td>0.0359</td>
<td>0.0215</td>
</tr>
<tr>
<td>250</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{2}$</td>
<td>Bias</td>
<td>-0.0651</td>
<td>-0.0695</td>
<td>0.6630</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0114</td>
<td>0.0117</td>
<td>0.4771</td>
</tr>
<tr>
<td>250</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{5}$</td>
<td>Bias</td>
<td>-0.0463</td>
<td>-0.0469</td>
<td>0.6575</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0099</td>
<td>0.0098</td>
<td>0.4646</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>Bias</td>
<td>-0.0650</td>
<td>-0.0679</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0146</td>
<td>0.0153</td>
<td>0.0111</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>$\frac{1}{5}$</td>
<td>Bias</td>
<td>-0.0552</td>
<td>-0.0558</td>
<td>0.0049</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0134</td>
<td>0.0133</td>
<td>0.0126</td>
</tr>
<tr>
<td>500</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{2}$</td>
<td>Bias</td>
<td>-0.0661</td>
<td>-0.0694</td>
<td>0.6464</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0075</td>
<td>0.0081</td>
<td>0.4360</td>
</tr>
<tr>
<td>500</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{5}$</td>
<td>Bias</td>
<td>-0.0449</td>
<td>-0.0447</td>
<td>0.6345</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0062</td>
<td>0.0060</td>
<td>0.4186</td>
</tr>
<tr>
<td>750</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>Bias</td>
<td>-0.0642</td>
<td>-0.0660</td>
<td>0.0037</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0096</td>
<td>0.0096</td>
<td>0.0076</td>
</tr>
<tr>
<td>750</td>
<td>1</td>
<td>$\frac{1}{5}$</td>
<td>Bias</td>
<td>-0.0474</td>
<td>-0.0486</td>
<td>0.0064</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0090</td>
<td>0.0092</td>
<td>0.0077</td>
</tr>
<tr>
<td>750</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{2}$</td>
<td>Bias</td>
<td>-0.0640</td>
<td>-0.0679</td>
<td>0.6445</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0059</td>
<td>0.0063</td>
<td>0.4260</td>
</tr>
<tr>
<td>750</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{5}$</td>
<td>Bias</td>
<td>-0.0434</td>
<td>-0.0438</td>
<td>0.6405</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0051</td>
<td>0.0050</td>
<td>0.4228</td>
</tr>
</tbody>
</table>
Table 2: $\hat{\beta}_2$ Bias and MSE: $\varepsilon_i \sim N(0, 1))$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\sigma^2$</th>
<th>$\sigma^2_{x_2}$</th>
<th>Ichimura</th>
<th>Klein and Spady</th>
<th>Probit</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>1</td>
<td>$1/2$</td>
<td>Bias</td>
<td>0.0672</td>
<td>0.0712</td>
<td>0.0049</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0323</td>
<td>0.0330</td>
<td>0.0156</td>
</tr>
<tr>
<td>250</td>
<td>1</td>
<td>$1/5$</td>
<td>Bias</td>
<td>0.0177</td>
<td>0.0185</td>
<td>-0.0061</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0477</td>
<td>0.0479</td>
<td>0.0416</td>
</tr>
<tr>
<td>250</td>
<td>$1/3$</td>
<td>$1/2$</td>
<td>Bias</td>
<td>0.0835</td>
<td>0.0899</td>
<td>0.3820</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0179</td>
<td>0.0187</td>
<td>0.1700</td>
</tr>
<tr>
<td>250</td>
<td>$1/3$</td>
<td>$1/5$</td>
<td>Bias</td>
<td>0.0517</td>
<td>0.0530</td>
<td>0.3793</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0186</td>
<td>0.0184</td>
<td>0.1886</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>$1/2$</td>
<td>Bias</td>
<td>0.0780</td>
<td>0.0817</td>
<td>-0.0008</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0201</td>
<td>0.0206</td>
<td>0.0075</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>$1/5$</td>
<td>Bias</td>
<td>0.0585</td>
<td>0.0596</td>
<td>0.0072</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0237</td>
<td>0.0236</td>
<td>0.0172</td>
</tr>
<tr>
<td>500</td>
<td>$1/3$</td>
<td>$1/2$</td>
<td>Bias</td>
<td>0.0932</td>
<td>0.0975</td>
<td>0.3792</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0139</td>
<td>0.0147</td>
<td>0.1556</td>
</tr>
<tr>
<td>500</td>
<td>$1/3$</td>
<td>$1/5$</td>
<td>Bias</td>
<td>0.0593</td>
<td>0.0595</td>
<td>0.3759</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0123</td>
<td>0.0120</td>
<td>0.1646</td>
</tr>
<tr>
<td>750</td>
<td>1</td>
<td>$1/2$</td>
<td>Bias</td>
<td>0.0853</td>
<td>0.0882</td>
<td>0.0027</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0163</td>
<td>0.0165</td>
<td>0.0057</td>
</tr>
<tr>
<td>750</td>
<td>1</td>
<td>$1/5$</td>
<td>Bias</td>
<td>0.0563</td>
<td>0.0580</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0167</td>
<td>0.0169</td>
<td>0.0113</td>
</tr>
<tr>
<td>750</td>
<td>$1/3$</td>
<td>$1/2$</td>
<td>Bias</td>
<td>0.0932</td>
<td>0.0985</td>
<td>0.3766</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0118</td>
<td>0.0127</td>
<td>0.1490</td>
</tr>
<tr>
<td>750</td>
<td>$1/3$</td>
<td>$1/5$</td>
<td>Bias</td>
<td>0.0601</td>
<td>0.0609</td>
<td>0.3697</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0101</td>
<td>0.0100</td>
<td>0.1532</td>
</tr>
</tbody>
</table>
Table 3: $\hat{\beta}_1$ Bias and MSE: $\varepsilon_i \sim L(0, \pi^2/3)$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\sigma^2_\varepsilon$</th>
<th>$\sigma^2_{x_2}$</th>
<th>Ichimura</th>
<th>Klein and Spady</th>
<th>Probit</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>Bias</td>
<td>-0.1322</td>
<td>-0.1380</td>
<td>-0.3295</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0845</td>
<td>0.0858</td>
<td>0.1299</td>
</tr>
<tr>
<td>250</td>
<td>1</td>
<td>$\frac{1}{5}$</td>
<td>Bias</td>
<td>-0.1197</td>
<td>-0.1239</td>
<td>-0.3263</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0903</td>
<td>0.0930</td>
<td>0.1270</td>
</tr>
<tr>
<td>250</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{2}$</td>
<td>Bias</td>
<td>-0.0976</td>
<td>-0.0969</td>
<td>0.0514</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0426</td>
<td>0.0415</td>
<td>0.0290</td>
</tr>
<tr>
<td>250</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{5}$</td>
<td>Bias</td>
<td>-0.0752</td>
<td>-0.0740</td>
<td>0.0532</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0317</td>
<td>0.0306</td>
<td>0.0276</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>Bias</td>
<td>-0.1101</td>
<td>-0.1073</td>
<td>-0.3254</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0610</td>
<td>0.0579</td>
<td>0.1161</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>$\frac{1}{5}$</td>
<td>Bias</td>
<td>-0.0678</td>
<td>-0.0696</td>
<td>-0.3183</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0368</td>
<td>0.0378</td>
<td>0.1109</td>
</tr>
<tr>
<td>500</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{2}$</td>
<td>Bias</td>
<td>-0.0697</td>
<td>-0.0742</td>
<td>0.0447</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0143</td>
<td>0.0160</td>
<td>0.0132</td>
</tr>
<tr>
<td>500</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{5}$</td>
<td>Bias</td>
<td>-0.0441</td>
<td>-0.0441</td>
<td>0.0547</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0109</td>
<td>0.0109</td>
<td>0.0143</td>
</tr>
<tr>
<td>750</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>Bias</td>
<td>-0.0773</td>
<td>-0.0807</td>
<td>-0.3259</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0267</td>
<td>0.0291</td>
<td>0.1126</td>
</tr>
<tr>
<td>750</td>
<td>1</td>
<td>$\frac{1}{5}$</td>
<td>Bias</td>
<td>-0.0679</td>
<td>-0.0698</td>
<td>-0.3280</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0251</td>
<td>0.0261</td>
<td>0.1137</td>
</tr>
<tr>
<td>750</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{2}$</td>
<td>Bias</td>
<td>-0.0583</td>
<td>-0.0615</td>
<td>0.0493</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0076</td>
<td>0.0082</td>
<td>0.0105</td>
</tr>
<tr>
<td>750</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{5}$</td>
<td>Bias</td>
<td>-0.0444</td>
<td>-0.0447</td>
<td>0.0517</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0080</td>
<td>0.0081</td>
<td>0.0104</td>
</tr>
</tbody>
</table>
Table 4: $\hat{\beta}_2$ Bias and MSE: $\varepsilon_i \sim L(0, \pi^2/3) $

| $n$  | $\sigma_\varepsilon^2$ | $\sigma_{\varepsilon | x}^2$ | Ichimura | Klein and Spady | Probit | Logit |
|------|-----------------|-----------------|-----------|-----------------|--------|-------|
| 250  | 1               | $\frac{1}{2}$   | Bias      | 0.0778          | 0.0853 | -0.1818 | 0.0173 |
|      |                 |                 | MSE       | 0.0668          | 0.0680 | 0.0474 | 0.0391 |
| 250  | 1               | $\frac{1}{5}$   | Bias      | 0.0320          | 0.0359 | -0.1762 | 0.0253 |
|      |                 |                 | MSE       | 0.0851          | 0.0857 | 0.0669 | 0.0961 |
| 250  | $\frac{1}{3}$   | $\frac{1}{2}$   | Bias      | 0.0895          | 0.0899 | 0.0369 | 0.3947 |
|      |                 |                 | MSE       | 0.0370          | 0.0363 | 0.0189 | 0.2062 |
| 250  | $\frac{1}{3}$   | $\frac{1}{5}$   | Bias      | 0.0558          | 0.0552 | 0.0484 | 0.4098 |
|      |                 |                 | MSE       | 0.0427          | 0.0422 | 0.0406 | 0.2758 |
| 500  | 1               | $\frac{1}{2}$   | Bias      | 0.0833          | 0.0819 | -0.1929 | -0.0005 |
|      |                 |                 | MSE       | 0.0464          | 0.0460 | 0.0443 | 0.0189 |
| 500  | 1               | $\frac{1}{5}$   | Bias      | 0.0307          | 0.0323 | -0.1909 | 0.0013 |
|      |                 |                 | MSE       | 0.0498          | 0.0505 | 0.0517 | 0.0406 |
| 500  | $\frac{1}{3}$   | $\frac{1}{2}$   | Bias      | 0.0860          | 0.0906 | 0.0251 | 0.3749 |
|      |                 |                 | MSE       | 0.0205          | 0.0219 | 0.0090 | 0.1646 |
| 500  | $\frac{1}{3}$   | $\frac{1}{5}$   | Bias      | 0.0449          | 0.0451 | 0.0264 | 0.3730 |
|      |                 |                 | MSE       | 0.0205          | 0.0204 | 0.0179 | 0.1870 |
| 750  | 1               | $\frac{1}{2}$   | Bias      | 0.0780          | 0.0804 | -0.1898 | 0.0044 |
|      |                 |                 | MSE       | 0.0291          | 0.0303 | 0.0406 | 0.0124 |
| 750  | 1               | $\frac{1}{5}$   | Bias      | 0.0564          | 0.0580 | -0.1800 | 0.0183 |
|      |                 |                 | MSE       | 0.0361          | 0.0368 | 0.0434 | 0.0293 |
| 750  | $\frac{1}{3}$   | $\frac{1}{2}$   | Bias      | 0.0798          | 0.0841 | 0.0214 | 0.3682 |
|      |                 |                 | MSE       | 0.0135          | 0.0143 | 0.0050 | 0.1499 |
| 750  | $\frac{1}{3}$   | $\frac{1}{5}$   | Bias      | 0.0528          | 0.0534 | 0.0221 | 0.3655 |
|      |                 |                 | MSE       | 0.0160          | 0.0161 | 0.0130 | 0.1685 |
Table 5: $\hat{\beta}_1$ Bias and MSE: $\varepsilon_i \sim t(0, \nu = 9)$

<table>
<thead>
<tr>
<th>n</th>
<th>$\sigma^2_{\varepsilon}$</th>
<th>$\sigma^2_{x_2}$</th>
<th>Ichimura Bias</th>
<th>Klein and Spady Bias</th>
<th>Probit MSE</th>
<th>Logit MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>-0.0996</td>
<td>-0.0968</td>
<td>-0.0531</td>
<td>0.4702</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0485</td>
<td>0.0488</td>
<td>0.0278</td>
<td>0.2924</td>
</tr>
<tr>
<td>250</td>
<td>1</td>
<td>$\frac{1}{5}$</td>
<td>-0.0804</td>
<td>-0.0824</td>
<td>-0.0477</td>
<td>0.4714</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0400</td>
<td>0.0420</td>
<td>0.0256</td>
<td>0.2886</td>
</tr>
<tr>
<td>250</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{2}$</td>
<td>-0.0691</td>
<td>-0.0692</td>
<td>0.4752</td>
<td>1.4116</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0135</td>
<td>0.0151</td>
<td>0.2571</td>
<td>2.0938</td>
</tr>
<tr>
<td>250</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{5}$</td>
<td>-0.0510</td>
<td>-0.0505</td>
<td>0.4862</td>
<td>1.4086</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0126</td>
<td>0.0123</td>
<td>0.2673</td>
<td>2.0831</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>-0.0715</td>
<td>-0.0729</td>
<td>-0.0779</td>
<td>0.4271</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0176</td>
<td>0.2153</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>$\frac{1}{5}$</td>
<td>-0.0501</td>
<td>-0.0511</td>
<td>0.0560</td>
<td>0.4561</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0161</td>
<td>0.0161</td>
<td>0.0150</td>
<td>0.2415</td>
</tr>
<tr>
<td>500</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{2}$</td>
<td>-0.0644</td>
<td>-0.0670</td>
<td>0.4585</td>
<td>1.3803</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0080</td>
<td>0.0084</td>
<td>0.2251</td>
<td>1.9534</td>
</tr>
<tr>
<td>500</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{5}$</td>
<td>-0.0457</td>
<td>-0.0451</td>
<td>0.4812</td>
<td>1.4002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0074</td>
<td>0.0075</td>
<td>0.2486</td>
<td>2.0160</td>
</tr>
<tr>
<td>750</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>-0.0621</td>
<td>-0.0630</td>
<td>-0.0623</td>
<td>0.4531</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0097</td>
<td>0.0096</td>
<td>0.0110</td>
<td>0.2255</td>
</tr>
<tr>
<td>750</td>
<td>1</td>
<td>$\frac{1}{5}$</td>
<td>-0.0451</td>
<td>-0.0459</td>
<td>-0.0583</td>
<td>0.4518</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0111</td>
<td>0.0112</td>
<td>0.0102</td>
<td>0.2233</td>
</tr>
<tr>
<td>750</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{2}$</td>
<td>-0.0609</td>
<td>-0.0622</td>
<td>0.4557</td>
<td>1.3764</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0058</td>
<td>0.0059</td>
<td>0.2183</td>
<td>1.9285</td>
</tr>
<tr>
<td>750</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{5}$</td>
<td>-0.0444</td>
<td>-0.0430</td>
<td>0.4882</td>
<td>1.4122</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0052</td>
<td>0.0050</td>
<td>0.2489</td>
<td>2.0292</td>
</tr>
</tbody>
</table>

23
Table 6: $\hat{\beta}_2$ Bias and MSE: $\varepsilon_i \sim t(0, \nu = 9)$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\sigma^2_\varepsilon$</th>
<th>$\sigma^2_{x_2}$</th>
<th>Ichimura</th>
<th>Klein and Spady</th>
<th>Probit</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>1</td>
<td>$1/2$</td>
<td>Bias</td>
<td>0.0816</td>
<td>0.0768</td>
<td>-0.0394</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0426</td>
<td>0.0421</td>
<td>0.0175</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$1/5$</td>
<td>Bias</td>
<td>0.0522</td>
<td>0.0531</td>
<td>-0.0269</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0472</td>
<td>0.0476</td>
<td>0.0361</td>
</tr>
<tr>
<td>250</td>
<td>$1/3$</td>
<td>$1/2$</td>
<td>Bias</td>
<td>0.0870</td>
<td>0.0860</td>
<td>0.2663</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0192</td>
<td>0.0188</td>
<td>0.0891</td>
</tr>
<tr>
<td></td>
<td>$1/3$</td>
<td>$1/5$</td>
<td>Bias</td>
<td>0.0534</td>
<td>0.0526</td>
<td>0.2671</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0225</td>
<td>0.0226</td>
<td>0.1145</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>$1/2$</td>
<td>Bias</td>
<td>0.0868</td>
<td>0.0888</td>
<td>-0.0394</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0214</td>
<td>0.0218</td>
<td>0.0087</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$1/5$</td>
<td>Bias</td>
<td>0.0449</td>
<td>0.0466</td>
<td>-0.0398</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0258</td>
<td>0.0258</td>
<td>0.0176</td>
</tr>
<tr>
<td>500</td>
<td>$1/3$</td>
<td>$1/2$</td>
<td>Bias</td>
<td>0.0892</td>
<td>0.0927</td>
<td>0.2558</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0144</td>
<td>0.0149</td>
<td>0.0757</td>
</tr>
<tr>
<td></td>
<td>$1/3$</td>
<td>$1/5$</td>
<td>Bias</td>
<td>0.0572</td>
<td>0.0562</td>
<td>0.2582</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0146</td>
<td>0.0136</td>
<td>0.0914</td>
</tr>
<tr>
<td>750</td>
<td>1</td>
<td>$1/2$</td>
<td>Bias</td>
<td>0.0817</td>
<td>0.0832</td>
<td>-0.0426</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0163</td>
<td>0.0163</td>
<td>0.0078</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$1/5$</td>
<td>Bias</td>
<td>0.0462</td>
<td>0.0475</td>
<td>-0.0469</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0207</td>
<td>0.0208</td>
<td>0.0157</td>
</tr>
<tr>
<td>750</td>
<td>$1/3$</td>
<td>$1/2$</td>
<td>Bias</td>
<td>0.0881</td>
<td>0.0901</td>
<td>0.2514</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0116</td>
<td>0.0117</td>
<td>0.0701</td>
</tr>
<tr>
<td></td>
<td>$1/3$</td>
<td>$1/5$</td>
<td>Bias</td>
<td>0.0611</td>
<td>0.0591</td>
<td>0.2666</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE</td>
<td>0.0107</td>
<td>0.0104</td>
<td>0.0867</td>
</tr>
</tbody>
</table>
Appendix 2: Figures

Figure 1: Error Distributions:

(a) Cumulative Distribution Functions

(b) Probability Density Functions
Figure 2: Select Estimated Densities from Ichimura SLS:

(a) \( n = 250, \sigma^2_\varepsilon = 1, \sigma^2_x = \frac{1}{5}, \varepsilon_i \sim N(0, 1) \)

(b) \( n = 750, \sigma^2_\varepsilon = 1, \sigma^2_x = \frac{1}{5}, \varepsilon_i \sim N(0, 1) \)

(c) \( n = 750, \sigma^2_\varepsilon = 1, \sigma^2_x = \frac{1}{5}, \varepsilon_i \sim L(0, \frac{\pi^2}{3}) \)

(d) \( n = 750, \sigma^2_\varepsilon = 1, \sigma^2_x = \frac{1}{5}, \varepsilon_i \sim t(0, \nu = 9) \)

Notes:
Kernel density estimate using parameter estimates compared to true probability density function evaluated with parameter estimates. Bandwidth selected to illustrate smoothness in density estimates.
Figure 3: Select Estimated Densities from Klein and Spady SML:

(a) $n = 250, \sigma^2_\varepsilon = 1, \sigma^2_{x_2} = 1/5, \varepsilon_i \sim N(0, 1)$

(b) $n = 750, \sigma^2_\varepsilon = 1, \sigma^2_{x_2} = 1/5, \varepsilon_i \sim N(0, 1)$

(c) $n = 750, \sigma^2_\varepsilon = 1, \sigma^2_{x_2} = 1/5, \varepsilon_i \sim L(0, \pi^2/3)$

(d) $n = 750, \sigma^2_\varepsilon = 1, \sigma^2_{x_2} = 1/5, \varepsilon_i \sim t(0, \nu = 9)$

Notes:
Kernel density estimate using parameter estimates compared to true probability density function evaluated with parameter estimates. Bandwidth selected to illustrate smoothness in density estimates.
Figure 4: Select Estimated Distributions from Ichimura SLS:

(a) $n = 250, \sigma_{\varepsilon}^2 = 1, \sigma_{x_2}^2 = 1/5, \varepsilon_i \sim N(0, 1)$

(b) $n = 750, \sigma_{\varepsilon}^2 = 1, \sigma_{x_2}^2 = 1/5, \varepsilon_i \sim N(0, 1)$

(c) $n = 750, \sigma_{\varepsilon}^2 = 1, \sigma_{x_2}^2 = 1/5, \varepsilon_i \sim L(0, \pi^2/3)$

(d) $n = 750, \sigma_{\varepsilon}^2 = 1, \sigma_{x_2}^2 = 1/5, \varepsilon_i \sim t(0, \nu = 9)$

Notes:
Leave-one-out NW estimator using parameter estimates compared to true cumulative distribution function evaluated with parameter estimates. Bandwidth selected to illustrate smoothness in distribution estimates.
Figure 5: Select Estimated Distributions from Klein and Spady SML:

(a) \( n = 250, \sigma^2 = \frac{1}{3}, \sigma^2_{x_2} = \frac{1}{5}, \varepsilon_i \sim N(0, 1) \)

(b) \( n = 250, \sigma^2 = \frac{1}{3}, \sigma^2_{x_2} = \frac{1}{5}, \varepsilon_i \sim N(0, 1) \)

(c) \( n = 750, \sigma^2 = \frac{1}{3}, \sigma^2_{x_2} = \frac{1}{5}, \varepsilon_i \sim L(0, \pi^2/3) \)

(d) \( n = 750, \sigma^2 = \frac{1}{3}, \sigma^2_{x_2} = \frac{1}{5}, \varepsilon_i \sim t(0, \nu = 9) \)

Notes:
Leave-one-out NW estimator using parameter estimates compared to true cumulative distribution function evaluated with parameter estimates. Bandwidth selected to illustrate smoothness in distribution estimates.
Appendix 3: Select MATLAB Code

Monte Carlo Simulation

```matlab
% MONTE CARLO SIMULATION using parametric probit and logit as well as
% semiparametric estimators proposed in Ichimura(1993) and Klein and Spady
% (1993)

% First column: sample size (n = 250, 500, 750)
% Second column: error distribution (1=std norm, 2=logistic, 3=student t)
% Third column: error variance parameter (1/3, 1)
% Fourth Column: x^2 variance parameter (1/2, 1/5)

par_ex = [250 1 1 1/2 ; % experiment matrix
          250 1 1/3 1/2 ;
          250 1 1/3 1/5 ;
          250 2 1 1/2 ;
          250 2 1 1/5 ;
          250 2 1/3 1/2 ;
          250 2 1/3 1/5 ;
          250 3 1 1/2 ;
          250 3 1 1/5 ;
          250 3 1/3 1/2 ;
          250 3 1/3 1/5 ;
          500 1 1 1/2 ;
          500 1 1 1/5 ;
          500 1 1/3 1/2 ;
          500 1 1/3 1/5 ;
          500 2 1 1/2 ;
          500 2 1 1/5 ;
          500 2 1/3 1/2 ;
          500 2 1/3 1/5 ;
          500 3 1 1/2 ;
          500 3 1 1/5 ;
          500 3 1/3 1/2 ;
          500 3 1/3 1/5 ;
          750 1 1 1/2 ;
          750 1 1 1/5 ;
          750 1 1/3 1/2 ;
          750 1 1/3 1/5 ;
          750 2 1 1/2 ;
          750 2 1 1/5 ;
          750 2 1/3 1/2 ;
          750 2 1/3 1/5 ;
          ];

par_ex = par_ex(1:40, :);
```

Special thanks to Professor Martins-Filho for indispensable guidance and assistance in developing these codes.
num_iter = 500; % Number of iterations of each experiment
pre.m_vol = zeros(num_iter, 16); % Size of matrix for first output
metrics = zeros(num_iter, 16); % Size of matrix for final .txt output

for i=1:size(par_ex, 1) % experiment loop
    fname = sprintf('results%d.txt', i); % format data into string
tic;
    for d=1:num_iter % iteration loop
        % DATA GENERATION------------------
        beta_1 = sqrt(3)/2; % true parameters
        beta_2 = 0.5;
        beta_true = [beta_1; beta_2];
        n = par_ex(i,1); % sample size
        esigma = sqrt(par_ex(i,3)); % variance parameter
        xsigma = sqrt(par_ex(i,4)); % variance parameter
        x_1 = rand(n, 1); % uniformly distributed [0, 1]
        x_2 = randn(n, 1); % standard normally distributed
        x = [x_1 x_2]; % x matrix
        z = x_1*beta_1 + x_2*beta_2; % index
        if par_ex(i,2) == 1 % Normal
            e = esigma*randn(n, 1); % normally distributed error term
            ystar = z + e; % latent
            y = (ystar >= 0); % threshold
        elseif par_ex(i,2) == 2 % Logistic
            u = rand(n, 1);
            l = log(u) - log(1-u); % logistically distributed random numbers
            e = esigma*l; % logistically distributed error term
            ystar = z + e; % latent
            y = (ystar >= 0); % threshold
        else % t
            v = 5; % degrees of freedom
            e = esigma*trnd(v, [n, 1]); % student t distributed error term
            ystar = z + e; % latent
            y = (ystar >= 0); % threshold
        end
        % ESTIMATION-----------------------------
    end
    metrics(d, :)= y; % store final output
end
h=1; % initial bandwidth

% Ichimura estimator least squares function
ls = @(theta)sum( (dens_ichimura(x,theta)>=0).*((y-gbeta_ichimura(y,x,theta)).^2 ) );

% Klein and Spady estimator log-likelihood function
log_likelihood = @(theta)−sum( ( y.*log(gbeta_ichimura(y,x,theta)) + (1−y).*log((1−gbeta_ichimura(y,x,theta)))));

b_dim = size(x,2); % columns of x matrix
lp = −ones(b_dim,1);
up = ones(b_dim,1);
lb = [lp;0.3]; % lower bound for bandwidth
ub = [up;Inf]; % upper bound for bandwidth

theta_initial_i = [beta_1; beta_2; h]; % normalized initial estimates
theta_initial_ks = [beta_1; beta_2; h]; % with initial bandwidth added

options = optimset('GradConstr','on');

% Ichimura estimation
[theta_ichimura,fval_i,exitflag_i,output_i,lambda_i,grad_i,hessian_i] = fmincon(ls,theta_initial_i,[],[],[],[],lb,ub,@norm_cons,options);
ichimura_beta = theta_ichimura(1:b_dim);

% Klein and Spady estimation
[theta_ks,fval,exitflag,output,lambda,grad,hessian] = fmincon(log_likelihood,theta_initial_ks,[],[],[],[],lb,ub,
                              @norm_cons,options);
ks_beta = theta_ks(1:b_dim);

% Probit estimation
probit(y, x, 1000, .0000001);
probit_beta = ans.beta;

% Logit estimation
logit(y, x, 1000, .0000001);
logit_beta = ans.beta;

% Bias Calculation
bias_ichimura = (ichimura_beta − beta_true)';
bias_ks = (ks_beta − beta_true)';
bias_probit = (probit_beta − beta_true)';
bias_logit = (logit_beta − beta_true)';

% Squared Error Calculation
se_ichimura = ((ichimura_beta − beta_true).^2)';
se_ks = ((ks_beta − beta_true).^2)';
se_probit = ((probit_beta − beta_true).^2)';
se_logit = ((logit_beta - beta_true).^2)';

% Initial results output
pre_m_vol(d,:) = [ bias_ichimura bias_ks se_ichimura se_ks bias_probit
                   bias_logit se_probit se_logit];

toc
end

% PERFORMANCE METRICS

% Bias
BIAS_ichimura_1 = (1/num_iter)*sum(pre_m_vol(:,1));
BIAS_ichimura_2 = (1/num_iter)*sum(pre_m_vol(:,2));
BIAS_ks_1 = (1/num_iter)*sum(pre_m_vol(:,3));
BIAS_ks_2 = (1/num_iter)*sum(pre_m_vol(:,4));
BIAS_probit_1 = (1/num_iter)*sum(pre_m_vol(:,9));
BIAS_probit_2 = (1/num_iter)*sum(pre_m_vol(:,10));
BIAS_logit_1 = (1/num_iter)*sum(pre_m_vol(:,11));
BIAS_logit_2 = (1/num_iter)*sum(pre_m_vol(:,12));

% Mean Squared Error
MSE_ichimura_1 = (1/num_iter)*sum(pre_m_vol(:,5));
MSE_ichimura_2 = (1/num_iter)*sum(pre_m_vol(:,6));
MSE_ks_1 = (1/num_iter)*sum(pre_m_vol(:,7));
MSE_ks_2 = (1/num_iter)*sum(pre_m_vol(:,8));
MSE_probit_1 = (1/num_iter)*sum(pre_m_vol(:,13));
MSE_probit_2 = (1/num_iter)*sum(pre_m_vol(:,14));
MSE_logit_1 = (1/num_iter)*sum(pre_m_vol(:,15));
MSE_logit_2 = (1/num_iter)*sum(pre_m_vol(:,16));

% Matrix to output
metrics = [ BIAS_ichimura_1 BIAS_ichimura_2 BIAS_ks_1 BIAS_ks_2
            MSE_ichimura_1 MSE_ichimura_2 MSE_ks_1 MSE_ks_2 BIAS_probit_1
            BIAS_probit_2 BIAS_logit_1 BIAS_logit_2 MSE_probit_1
            MSE_logit_1 MSE_logit_2]; %Final output matrix

%Saves the .txt file
save(fname,'metrics','-ASCII') %Saves the .txt file
end
Leave-one-out Nadaraya Watson Kernel Estimator

% Leave-one-out Nadaraya Watson Kernel Estimator
% OUTPUT:
% g: a vector of values for the estimated g function at the sample points
% for fixed theta, vector of regressand y and regressor matrix x
% INPUT:
% y: is a vector of regressands
% x: is a matrix of regressors
% theta: is a parameter vector with last position being the bandwidth

function g = gbeta_ichimura(y,x,theta)
col_x=size(x,2);
z = x*theta(1:col_x); % index
n=size(y,1); % sample size

% Leave-one-out Nadaraya Watson estimator
a=(1/theta(col_x+1))*(gsubtract(z,z')); % weights
k=(1/sqrt(2*pi))*exp(-0.5*a.^2); % gaussian kernel
kj=k-k(1,1)*eye(n); % zero diagonal
g=(kj'*y)./(kj'*ones(n,1)); % vector of density estimates at sample points
References


