ROBUST NONPARAMETRIC ESTIMATION AND TESTING OF ECONOMETRIC MODELS

A Dissertation

by

ZHENG LI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Qi Li
Co-Chair of Committee,	Yonghong An
Committee Members,	Fernando Luco
	Ximing Wu
Head of Department,	Timothy Gronberg

May 2017

Major Subject: Economics

ABSTRACT

In the first essay, we investigate the nonlinear quantile regression with mixed discrete and continuous regressors. A local linear smoothing technique with the mixed continuous and discrete kernel function is proposed to estimate the conditional quantile regression function. Under some mild conditions, the asymptotic distribution is established for the proposed nonparametric estimators, which can be seen as the generalization of some existing theory which only handles the case of purely continuous regressors. We also study the choice of the tuning parameters in the estimation procedure which is crucial in kernel-based smoothing approach. We suggest using the cross-validation approach to choose the optimal bandwidths. A simulation study is provided to examine the finite sample behavior of the proposed method and compare it with the naive local linear quantile estimation without smoothing the discrete regressors and the nonparametric inverse-CDF (cumulative distribution function) method.

In the second essay, we propose to estimate a nonparametric regression function subject to a monotonicity restriction using the Knn (k-nearest neighbors) method. We also propose using a new convergence criterion to measure the closeness between an unconstrained and the (monotone) constrained Knn estimated curves. This method is an alternative to the monotone kernel methods. We use a bootstrap procedure for testing the validity of the monotone restriction. We apply our method to the 'Job Market Matching' data and find that the unconstrained/constrained Knn estimators work better than kernel estimators for this type of highly unevenly distributed data.

In the third essay, we propose a nonparametric methodology to test heterogeneous risk preference against asymmetric value distribution of bidders. By modeling bidders' asymmetry as unobserved heterogeneity, we first show that bid distributions conditional

on the heterogeneity are nonparametrically identified. Next, we find that the two alternative models provide distinct implications on the conditional bid distributions. Based on the estimated conditional bid distributions, we are able to distinguish the two models by formally testing the distinct model implications. The Monte Carlo experiments demonstrate the good performance of our method. In an application using the US Forest Service timber auction data, we find that the data support the model with heterogeneity in risk preference.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Qi Li (chair), Professor Yonghong An (co-chair) and Professor Fernando Luco of the Department of Economics, and Professor Ximing Wu of the Department of Agricultural Economics.

The data analyzed for Chapter 3 was provided by Professor Yonghong An. The analyses depicted in Chapter 1 and 2 were conducted in part by Professor Qi Li, Professor Degui Li and Professor Guannan Liu.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by the Department of Economics at Texas A&M University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

This dissertation includes three essays, the title of the first essay is “Nonparametric Estimation of Conditional Quantile Regression with Mixed Discrete and Continuous Data". The title of the second essay is “Nonparametric Knn Estimation with Monotone Constraints". The title of the third essay is “Nonparametric Identification and Testing of First-Price Auctions with Asymmetric Bidders".

1.1 Introduction to the First Essay

In recent years, there has been increasing interest on nonparametric estimation of the regression relationships among variables, as a nonparametric approach allows the data “speak for themselves" and thus has the ability to detect the regression structure which may be difficult to be uncovered by the traditional parametric modelling approach. Various nonparametric methods have attracted the attention of the statisticians and econometricians partly due to the wide availability of large data sets in empirical applications, see, for example, Green and Silverman (1993), Wand and Jones (1995), Fan and Gijbels (1996), Pagan and Ullah (1999), Li and Racine (2007) and Horowitz (2009). One of the most commonly-used nonparametric estimation methods is the local linear smoothing method as it is well-known that the local linear approach has advantages over the traditional Nadaraya-Watson kernel approach, such as higher asymptotic efficiency, design adaption and automatic boundary correction. We refer to the book by Fan and Gijbels (1996) for a detailed account on this subject.

Most of the literature introduced above mainly focuses on the nonparametric estimation approaches with continuous regressors. However, in practice, it is not uncommon that some of the regressors might be discrete (c.f., gender, race and religious beliefs) and the extension of the nonparametric method to handle the case of discrete regressors is

non-trivial. For example, the local linear smoothing method relies on the first-order Taylor expansion of the regression function with respect to the continuous component, and needs to be substantially generalized to handle the case of discrete regressors. A plausible method is to split the whole sample into several subgroups determined by the values of the discrete regressors, and then apply the nonparametric method separately to each subgroup. However, as mentioned by Li and Racine (2004), such splitting method may not perform well when the number of the subgroups is relatively large and the number of observations in some subgroups is small. To address this problem, they introduce a nonparametric kernel-based method with both continuous and discrete kernel functions involved, which works well for the case of mixed continuous and discrete regressors. Recent developments on this field include Racine and Li (2004), Hall et al. (2004), Li et al. (2009), Li et al. (2016), most of which investigate the nonparametric estimation of the conditional mean regression function as well as the associated econometric application.

However, it is well-known that the conditional mean may not be a good representative of the impact of the explanatory variables on the response variable. Hence, it is usually of interest to model the conditional quantile when studying the regression relationship between the dependent variable and the explanatory variables. Since the seminal paper by Koenker and Bassett Jr (1978), the quantile regression method has been widely used in many disciplines such as economics, finance, political sciences and other social science fields. The quantile method serves as a robust alternative to the mean regression method. Koenker (2005) gives an overview on various methodologies in quantile regression as well as their applications. Recent development on the nonparametric quantile estimation includes Yu and Jones (1998), Cai (2002), Yu and Lu (2004), Lin and Li (2007), Cai and Xu (2008), Li et al. (2013), Hallin et al. (2009) and Cai and Xiao (2012). In particular, Li et al. (2013) first estimate the conditional cumulative distribution function (CDF) nonparametrically which admits a mix of discrete and continuous data, and then obtain

quantile estimation by inverting the estimated conditional CDF at the desired quantiles. They choose the optimal bandwidth when estimating the nonparametric CDF, and thus avoid the direct bandwidth selection for the kernel-based quantile estimation.

In this paper, we propose a different nonparametric method to estimate the conditional quantile function via minimizing the local linear based “check function" defined in Section 2.1. The local linear smoothing technique is constructed with both the continuous and discrete kernel functions involved, and can thus handle the case of mixed continuous and discrete regressors. The asymptotic distribution is established for the proposed nonparametric estimators. We also study the choice of the tuning parameters in the local linear estimation procedure, and propose a completely data-driven cross-validation (CV) approach to directly choose the optimal bandwidths which is different from that in Li et al. (2013). The asymptotic properties of the CV bandwidth selection approach are discussed in Section 2.2, where the asymptotic optimality of the chosen bandwidths are obtained. A simulation study is presented to illustrate the finite sample behavior of the proposed method. Meanwhile, in the simulation study, we also compare our method with some existing methods such as the naive local linear quantile estimation without smoothing the discrete regressors and the nonparametric inverse-CDF method proposed by Li et al. (2013).

1.2 Introduction to the Second Essay

In regression analyses, researchers may use nonparametric estimation methods when they are not sure about the specific regression function forms for the data they want to analyze. Some widely used nonparametric estimation techniques include kernel, K-nearest-neighbor (Knn), series and wavelet methods. Undoubtedly each nonparametric estimation method has its own advantages and disadvantages, and different methods may be preferred for different empirical applications. For example, although the kernel method is the most

popular nonparametric method in practice, it has an obvious drawback when using a fixed smoothing parameter over the whole data range, the use of a constant bandwidth may result in oversmoothing in some ranges of the data support while undersmoothing in other parts of the data support,[1] this is true especially when the data is highly unevenly distributed over the data support (which is common for non-experimental data). Therefore, the estimation and inference results based on the kernel method could be unreliable in such situations. In this case, the Knn estimation method might be the preferable approach to analyze the data. Since the Knn method always uses the k nearest observations in the estimation, it effectively uses a different bandwidth when estimating the unknown regression function at each different point of the data support.

Although economic theories rarely give specific regression functional forms, they often lead to some shape restrictions on regression functions such as monotonicity, concavity and symmetry, etc. For example, based on the utility-maximization behaviors of individuals and firms, Gan and Li (2016) show that the job matching probability function is a monotonically increasing function in market size. Therefore, nonparametric regression function estimation with shape constraints is attractive and has received much attention recently among econometricians and statisticians (Delecroix and Thomas-Agnan, 2000; Henderson and Parmeter, 2009). For example, Hall and Huang (2001) consider the restricted kernel regression function estimation under the monotonicity constraint. To construct a constrained estimator, Hall and Huang (2001) adjust an unconstrained estimator by attaching different weights to different observed data points. The weights are chosen to minimize a distance measure while obeying the monotonic constraint. They show that the constrained kernel estimator is consistent and has the same rate of convergence as the unconstrained kernel estimator under some general conditions. Hall and Huang (2001)

[1] Although one can use adaptive bandwidth to avoid this problem, adaptive bandwidth method requires that one selects a different bandwidth at each different data evaluation point. It is computationally quite costly especially in large sample applications.

method can be applied to a wide range of kernel-type estimators including Nadaraya–Watson, local linear, Priestley–Chao and Gasser–Müller estimators.

Empirical applications of Hall and Huang (2001) method include Henderson et al. (2012) who estimate a monotone first-price auction function, and Xu and Phillips (2011) who suggest using Hall and Huang (2001) method to ensure monotonicity of an estimated conditional variance function. A more general approach is proposed by Du et al. (2013) who extend Hall and Huang (2001) method to more general shape restrictions and to the multivariate and multi-constraint settings. Applications of Du et al. (2013) approach can be found in Malikov et al. (2016), and Sun (2015). Specifically, Malikov et al. (2016) use Du et al. (2013) approach to impose linear homogeneity constraints onto the cost functions in studying production technologies of US retail credit unions; Sun (2015) proposes using a constrained nonparametric method to estimate an input distance function with constraints imposed on the regression function. Besides Hall and Huang (2001) method of choosing weights, Lee et al. (2014) propose a bagging constrained nonparametric estimation method. They use the local linear least square estimation method and achieve local monotonicity by imposing constraints on local coefficients. They further demonstrate that their proposed approach performs well in forecasting equity premium.

There are many research papers that promote nonparametric smoothing under shape constraints. For example, Freyberger and Horowitz (2015) propose using monotone restriction to identify an unknown regression function with endogenous discrete covariate when the discrete instruments has fewer mass points than that of the covariates. However, most of them use either the kernel-based method or the sieve-based (series) method, see Chen (2007) on an overview of using sieve method in estimating nonparametric / semiparametric models with shape restrictions. As we argued earlier when data points are highly unevenly distributed over the data support, the kernel-based estimation method has some undesirable properties when one uses a fixed bandwidth parameter. First, the uncon-

strained estimator may give unreliable estimation results especially at ranges with sparse data. Second, there may not exist a set of feasible weights that make the constraint binding for the constrained estimator. We show that this is indeed true for the 'Job Market Matching' data collected by Gan and Li (2016).

In this paper, we propose using the Knn smoothing method to estimate a nonparametric conditional mean function under the monotonicity constraint. The constrained estimation method in Hall and Huang (2001) and Du et al. (2013) can be directly applied to using the Knn estimation method. Moreover, instead of measuring the distance by a power divergence metric as proposed in Hall and Huang (2001) or the Euclidean distance between a uniform weight and the non-uniform weight proposed by Du et al. (2013), we suggest choosing the weights by minimizing an objective function that directly measures the distance between the curves with and without constraints. We apply our method to the 'Job Market Matching' data taken from Gan and Li (2016), and estimate the 'matching probabilities' (frequency of candidates finding academic jobs) as a function of the market size. This data is highly unevenly distributed, we show that both the constrained and unconstrained Knn estimators give more reasonable estimation results than those obtained by using the kernel method. We also use a bootstrap procedure discussed in Du et al. (2013) to test whether the relationship between the job matching probability and market size is monotone or not. Again, we show that using the Knn method gives more reasonable testing result than that using the kernel method.

1.3 Introduction to the Third Essay

In most of the existing studies of auction models, bidders are assumed to be symmetric, i.e., they are homogeneous except that their values are i.i.d. draws from the same value distribution. Until recently, a few empirical studies document two main sources of asymmetry among bidders: Guerre et al. (2009), and Campo (2012) show that bidders

have different levels of risk aversion, and Athey et al. (2011) find that bidders' private values are distributed according to different distributions. Nevertheless, most of the existing studies presume both the existence of asymmetry and the source of the asymmetry. It is unclear in the literature how researchers can test the existence and source (if any) of bidders' asymmetry from observed bids. We aim to develop a formal testing procedure in this paper to investigate whether asymmetry exists, and furthermore, what the source of asymmetry is. This formal testing procedure can provide solid empirical evidences of bidders' asymmetric behavior, which will shed some light on analysis of auctions' revenue as well as mechanism design questions.

We focus on two widely studied first-price sealed-bid auction models, where bidders have (1) asymmetric risk preferences (AP model), and (2) asymmetric value distributions (AVD model). The existence of equilibria is guaranteed by Athey (2001). In both models, we treat the asymmetry of bidders as unobserved and discrete heterogeneity. Thus, the observed joint distribution of bidders is a finite mixture with the components being the conditional distributions of bids and coefficients being the probability of the heterogeneity. It take several steps for us to formally test AP against AVD: First, we show that the conditional distributions of bids on the heterogeneity are nonparametrically identified and estimable from the observed distribution using the recently developed methodology in measurement error, i.e., Hu (2008). Next, we prove that the difference between two conditional distributions is strictly increasing and decreasing respectively in the quantiles for AP and AVD model. To exploit the model implications, we propose a procedure to test the monotonicity of the difference using the estimated conditional bid distributions.

In the first step, we nonparametrically recover conditional distributions of bids. Upon characterizing the asymmetry of bidders by discrete heterogeneity in both AP and AVD model, and labeling bidders with heterogeneity as "types", the existence of asymmetry is equivalent to that the number of types is greater than one. Our identification argument re-

quires that we observe the identity of each bidder, and at least three bids from each bidder. The three bids collected from each bidder enable us to recover the joint bid distribution. Assuming each bidder's type is time-invariant, the heterogeneity information is thus embedded into the joint bid distribution. It can be shown that the number of types is equal to the rank of the matrix which is constructed using the observed bid distribution. To determine the number of type, we employ the testing procedure proposed by Robin and Smith (2000), which tests the rank of a matrix in a sequential manner. The testing result gives the number of types and further implies the existence of asymmetry. The identification of the conditional bid distributions employs the recent development in measurement error, e.g., Hu (2008). The main idea of identification is to use the multiple bids of each bidder as measurements of his type. Since values are assumed to be independent across auctions for a bidder and type is time-invariant, the correlation of the multiple bids can be used to recover the conditional distributions on type. Note that the identification in this step is constructive and an estimating procedure follows directly the results of identification.

In the second step, we first investigate the testable implications based on the identified conditional distributions of bids for the two models. We prove that in both models, the difference between any two conditional bid distributions is monotone in its quantile. Specifically, the difference is increasing in AP model and decreasing in AVD model. To leverage these model implications, we obtain the distance between any two conditional bid distributions from the first step. The test of the monotonicity of this distance in its quantile is based on the method proposed by Fang and Santos (2014). The idea is that we project the estimated function (the difference as a function of its quantile) into the convex set of all monotone functions, and measure the distance between the function and its projection. This distance serves as our test statistic. A practical difficulty of the test is to construct the critical values via bootstrapping: since the operation of projection is not Hadamard differentiable, the regular procedure does not work. Thus we employ the method in Fang

and Santos (2014) to obtain the critical values by a bootstrap procedure based on delta method.

We demonstrate the performance of our approach by several numerical studies. The results suggest that both the test power and the test size behave very well for mid-size samples in both AP and AVD models. We further check the robustness of our testing procedure subject to a crucial assumption: values of a bidder are independent across auctions. For this purpose, we allow correlation of values to check the validity of our test. The results show that for mid-size samples, if the correlation of values is not large, e.g., less than 0.1 our test still performs very well.

Using the US Forest Service (USFS) Timber Auction data, we empirically test the existence and source of asymmetry of bidders. First of all, by testing the rank of an observed matrix, we find that the asymmetry exists and there are two types of bidders. For each type of bidders, we recover their type probabilities (about 0.5) and the conditional bid distributions. Then we apply the formal test to the data and we obtain a p-value 0.855 and 0.011 if the data are generated by AP and AVD model, respectively. The testing result then suggests that AP model better explains the timber auction data and this is in contrast to the findings in Athey et al. (2011) where bidders have asymmetric value distributions.

The main contribution of the paper is to show that bidders' asymmetry is testable from the observables in standard auction data. Unlike the existing literature on bidders' asymmetry, which often assumes the existence and specification of the asymmetry *ex ante* by researchers, this paper proposes a constructive method to detect the asymmetry directly from observables. We first boil down the existence of bidders' asymmetry to the rank of an estimable matrix and test the rank rigorously following Robin and Smith (2000). Similar existing studies on the rank lack a theoretical foundation, e.g., in Hu et al. (2013) the rank of a similar matrix is analyzed using condition number. A main innovation of the paper is the finding that the model implications of AP and AVD models are distinct and testable,

which enables us to test a model against the other. To my best knowledge, this is the first paper to test alternative asymmetries of bidders in first-price sealed-bid auctions. Even though that our test only applies to AP and AVD models, the methodology will shed some light on rigorous analysis of bidders' asymmetry in auctions.

Another contribution of this paper is to apply the recent development of measurement error into the analysis of first-price auctions with asymmetry. The methodology in measurement error has been applied to analyze first-price auctions (e.g., see An et al. (2010), Hu et al. (2013)). Nevertheless, this paper is among the first ones that treat bidders' asymmetry as unobserved heterogeneity and conduct the analysis using the method in measurement error. In addition, this paper employs a newly developed monotonicity test to deal with auction data. Specifically, our test statistic involves non-differentiable functions and a new bootstrap method has to be implemented to obtain the critical values. Also, our approach is fully nonparametric and does not depend on any parametric assumptions. To the best of our knowledge, ours is the first paper to provide empirical evidence on the existence of bidders' asymmetry by using a fully nonparametric approach.

This paper is closely related to the literature of first-price auctions with asymmetric bidders, e.g., Athey et al. (2011) and Campo (2012). Athey et al. (2011) investigate the same USFS timber auctions as we do. They assume that the asymmetry is indicated by some observed characteristics of bidders (e.g. the scale of the company), and further assume the asymmetry lies in value distributions. Campo (2012) uses the construction contracts from Los Angeles City Hall. She documents that firms' heterogeneous risk preferences are affected by its financial situation. These connections between bidders' characteristics and asymmetry are plausible and yet vague. Our approach does not rely on the presumed relationship between bidders' characteristics and asymmetry. Instead, our formal test procedures provide a new and convincing evidence for the existence of asymmetry among bidders. Especially, for the USFS timber auction data, we document that

the source of the asymmetry is in the risk preferences, other than the value distributions indicated by Athey et al. (2011).

This paper is further related to some recent studies of auctions with heterogeneity. An (2010) considers auctions with bidders being different levels of cognitive ability. Hu et al. (2013) focus on auctions with unobserved and nonseparable heterogeneity. Our approach of identification using the results in measurement error is related to Li and Vuong (1998), Li et al. (2000), Krasnokutskaya (2011) and An et al. (2010). Our monotonicity test leverages the latest development of the inference on non-differentiable but directionally differentiable functions (Fang and Santos, 2014; Hong and Li, 2014). This is related to the literature of nonparametric testing of shape constraints on functions (Hall et al., 2001; Hall and Huang, 2001; Du et al., 2013; Li et al., 2016).

2. NONPARAMETRIC ESTIMATION OF CONDITIONAL QUANTILE REGRESSION WITH MIXED DISCRETE AND CONTINUOUS DATA

2.1 Local Linear Quantile Regression

In this section, we describe the nonparametric estimation method for the conditional quantile regression function, and then give the asymptotic theory of the resulting estimator. Suppose that $(Y_i, X_i', Z_i')'$, $i = 1, \cdots, n$, are the observations independently drawn from an identical distribution, where Y_i is univariate, X_i is a d_1-dimensional continuous vector and Z_i is a d_2-dimensional discrete vector. For notational and expositional simplicity, we will only consider the case that X_i and Z_i are both scalars. That is, we let $d_1 = d_2 = 1$ throughout the paper. Extension of the methodology and theoretical property to the case of $d_1 > 1$ and $d_2 > 1$ is straightforward.

For fixed x_0 and z_0, denote by $\mathbb{F}(y|x_0, z_0)$, $y \in \mathsf{R}$, the conditional CDF of the response variable Y_i (evaluated at y) given the regressors $X_i = x_0$ and $Z_i = z_0$. Let $\mathbb{Q}_\tau(x_0, z_0)$, $0 < \tau < 1$, be the conditional τ-quantile regression function of Y_i given $X_i = x_0$ and $Z_i = z_0$, which is defined as

$$\mathbb{Q}_\tau(x_0, z_0) = \inf\big\{y \in \mathsf{R} : \ \mathbb{F}(y|x_0, z_0) \geq \tau\big\} \tag{2.1}$$

or

$$\mathbb{Q}_\tau(x_0, z_0) = \arg\min_{a \in \mathsf{R}} \mathbb{E}\big[\rho_\tau(Y_i - a)\big|X_i = x_0, Z_i = z_0\big], \tag{2.2}$$

where $\rho_\tau(\cdot)$ is the check (loss) function

$$\rho_\tau(y) = y(\tau - I_{\{y<0\}})$$

with I_A being the indicator function of the set A.

We next apply the local linear smoothing approach to estimate the conditional τ-quantile regression function based on the definition (2.2). It is well-known that the local linear approach has various advantages over the Nadaraya-Watson kernel approach (c.f., Fan and Gijbels, 1996). However, the mixture of categorical and continuous data in the regressors makes our estimation methodology more complicated than that in existing literature which only considers continuous regressors. To address this issue, two types of kernel-weights are needed to handle the mixed continuous and discrete data locally. For the continuous regressor, we use a kernel defined by

$$\mathbb{K}_h(X_i - x_0) = \frac{1}{h}\mathbb{K}\big(\frac{X_i - x_0}{h}\big), \tag{2.3}$$

where h is the bandwidth for continuous regressor and $\mathbb{K}(\cdot)$ is a univariate kernel function. For the discrete regressor, we use the following discrete kernel

$$\Lambda_\lambda(Z_i, z_0) = \lambda^{I_{\{Z_i \neq z_0\}}}, \tag{2.4}$$

where $\lambda \in [0, 1]$ can be seen as the bandwidth for the discrete regressor. Then, the local linear estimates of $\mathbb{Q}_\tau(x_0, z_0)$ and its derivative (with respect to the continuous component) $\mathbb{Q}_\tau^{(1)}(x_0, z_0)$ are obtained by minimizing the weighted loss function

$$\mathcal{L}_n(\alpha, \beta; x_0, z_0) = \sum_{j=1}^{n} \rho_\tau\big[Y_j - \alpha - \beta(X_j - x_0)\big]\mathbb{K}_h(X_j - x_0)\Lambda_\lambda(Z_j, z_0) \tag{2.5}$$

with respect to α and β. We denote the minimizers by $\widehat{\mathbb{Q}}_\tau(x_0, z_0)$ and $\widehat{\mathbb{Q}}_\tau^{(1)}(x_0, z_0)$, respectively. When the tuning parameter in the discrete kernel, λ, is chosen as zero, the above estimator reduces to the local linear quantile estimators by the traditional approach

which splits the whole sample into several groups (or sub-samples) according to the value of the discrete regressor. Thus, there would be non-smoothing over the discrete regressor. However, as pointed out by Li and Racine (2004), such a splitting method may increase the estimation variance. In particular, it is well-known that if the sample size in splitted subgroups are too small, one cannot hope to get sensible results with the splitting local linear quantile estimation method. On the other hand, when λ is chosen as one, the discrete regressor would not have any influence on the response variable and such a discrete regressor can be regarded as an irrelevant regressor (Hall, Li and Racine, 2007). Based on the above discussion, throughout this paper, we let $0 \leq \lambda \leq 1$. Section 2.2 below will introduce a data-driven method to choose appropriate bandwidths h and λ.

Before giving the asymptotic distribution theory for $\widehat{\mathbb{Q}}_\tau(x_0, z_0)$ and $\widehat{\mathbb{Q}}_\tau^{(1)}(x_0, z_0)$, we give some notations. For expositional simplicity, we only consider the case that the univariate discrete variable takes two values (z_0 or z_1) as the extension to the more general case is straightforward. Let $\mu_k = \int u^k \mathbb{K}(u)du$ and $\nu_k = \int u^k \mathbb{K}^2(u)du$ for $k = 0, 1, 2, \cdots$. Let $f_e(\cdot|x, z)$ be the conditional density function of $e_i \equiv Y_i - \mathbb{Q}_\tau(X_i, Z_i)$ for given $X_i = x$ and $Z_i = z$ and assume that $\mathbb{P}(e_i \leq 0 | X_i = x_0, Z_i = z_0) = \tau$. Let p_1 be the probability of $Z_i = z_0$ and $f(x|z)$ be the conditional density function of X_i for given $Z_i = z$. Let $\mathbb{Q}_\tau^{(k)}(x_0, z_0)$ be the k-th order derivative function of $\mathbb{Q}_\tau(\cdot, \cdot)$ (with respect to the continuous component) at point (x_0, z_0), $k \geq 1$. Define

$$S(x_0, z_0) = \begin{bmatrix} S_1(x_0, z_0) & S_2(x_0, z_0) \\ S_2(x_0, z_0) & S_3(x_0, z_0) \end{bmatrix} \text{ with } S_k(x_0, z_0) = p_1 \mu_{k-1} f_e(0|x_0, z_0) f(x_0|z_0),$$

$\Omega(x_0, z_0) = \tau(1-\tau) p_1 f(x_0|z_0) \text{diag}(\nu_0, \nu_2)$ and $\overline{b}(x_0, z_0) = \big[\overline{b}_1(x_0, z_0), \overline{b}_2(x_0, z_0)\big]'$, where

$$\overline{b}_1(x_0, z_0) = \frac{1}{2} h^2 \mathbb{Q}_\tau^{(2)}(x_0, z_0)\mu_2 + \lambda \frac{(1-p_1) f(x_0|z_1)}{p_1 f(x_0|z_0)} \big[\mathbb{Q}_\tau(x_0, z_1) - \mathbb{Q}_\tau(x_0, z_0)\big]$$

and $\overline{b}_2(x_0, z_0) = o(h^2 + \lambda)$. We establish the asymptotic distribution theory for $\widehat{\mathbb{Q}}_\tau(x_0, z_0)$ and $\widehat{\mathbb{Q}}_\tau^{(1)}(x_0, z_0)$ in the following theorem.

THEOREM 2.1. *Suppose that Assumptions 1–3 and 4(i) in Appendix A are satisfied. Then, we have*

$$\sqrt{nh}\begin{bmatrix} \widehat{\mathbb{Q}}_\tau(x_0, z_0) - \mathbb{Q}_\tau(x_0, z_0) - \overline{b}_1(x_0, z_0) \\ h\widehat{\mathbb{Q}}_\tau^{(1)}(x_0, z_0) - h\mathbb{Q}_\tau(x_0, z_0) - \overline{b}_2(x_0, z_0) \end{bmatrix} \xrightarrow{d} \mathbb{N}\big[\mathbf{0}, V_*(x_0, z_0)\big], \tag{2.6}$$

where

$$V_*(x_0, z_0) = S^{-1}(x_0, z_0)\Omega(x_0, z_0)S^{-1}(x_0, z_0) = \frac{\tau(1-\tau)}{f_e^2(0|x_0, z_0)p_1 f(x_0|z_0)}\begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2/\mu_2^2 \end{pmatrix}.$$

The above theorem can be seen as the extension of the corresponding results from the continuous regressors case (c.f., Yu and Jones, 1998; and Hallin, Lu and Yu, 2009) to the mixed continuous and categorical regressors case. The point-wise convergence rate ($\sqrt{nh}$-rate) and the form of the asymptotic variance in (2.6) indicate that the discrete kernel $\Lambda_\lambda(Z_i, z_0)$ does not contribute to the asymptotic variances of the estimators $\widehat{\mathbb{Q}}_\tau(x_0, z_0)$ and $\widehat{\mathbb{Q}}_\tau^{(1)}(x_0, z_0)$. However, the involvement of the discrete kernel function would influence the form of the asymptotic bias, which can be seen from the second term of $\overline{b}_1(x_0, z_0)$,

$$\lambda\frac{(1-p_1)f(x_0|z_1)}{p_1 f(x_0|z_0)}\big[\mathbb{Q}_\tau(x_0, z_1) - \mathbb{Q}_\tau(x_0, z_0)\big].$$

This finding is similar to that obtained by Li, Lin and Racine (2013).

2.2 Choice of the Tuning Parameters

In this section, we study the bandwidth selection problem, which is of crucial importance in nonparametric local linear smoothings. A completely data driven cross validation

(CV) based method will be introduced to choose appropriate tuning parameters h and λ. The CV-based bandwidth selection criterion has been extensively studied in the context of local kernel conditional mean regression estimation with continuous regressors (c.f., Rice, 1984; Härdle and Vieu, 1992; Hall, Lahiri and Polzehl, 1995; Xia and Li, 2002; and Leung, 2005). In recent years, there has been also increasing interest in the CV bandwidth selection approach for the case with mixed continuous and categorical regressors (c.f., Li and Racine, 2004; Racine and Li, 2004; and Li, Simar and Zelenyuk, 2014). However, most of the existing literature focuses on the bandwidth selection in the kernel-based estimation in the context of *conditional mean* regression. Extension of the CV bandwidth selection method to the conditional quantile regression is non-trivial and the derivation of the asymptotic theory in quite challenging as there is no closed form expression for the local linear quantile estimation. There is no theoretical result on investigating the bandwidth selection issue in the local linear quantile regression with mixed continuous and discrete data. In fact, to the best of our knowledge, there is no theoretical result on using completely data driven CV method to select smoothing parameters even with only continuous covariates. Our paper aims to fill this gap.

Let $\widehat{\mathbb{Q}}_{(-i)}(X_i, Z_i; h, \lambda)$ be the leave-one-out local linear estimated value of $\mathbb{Q}_\tau(X_i, Z_i)$ with bandwidths h and λ, which can be obtained by minimizing (2.5) with (x_0, z_0) being replaced by (X_i, Z_i) and $\sum_{j=1}^n$ being replaced by $\sum_{j=1, j\neq i}^n$. Define the CV-based loss function as

$$\mathbb{CV}(h, \lambda) = \sum_{i=1}^{n} \rho_\tau\big[Y_i - \widehat{\mathbb{Q}}_{(-i)}(X_i, Z_i; h, \lambda)\big] M(X_i, Z_i), \tag{2.7}$$

where $M(X_i, Z_i)$ is a weight function trimming out boundary observations. Then, the optimal bandwidths can be chosen as $\widehat{h}$ and $\widehat{\lambda}$, which minimize $\mathbb{CV}(h, \lambda)$ defined in (2.7).

We next study the asymptotic property of the CV bandwidth selection method. Define

$$b(X_i, z_0; h, \lambda) = \frac{1}{2}h^2\mathbb{Q}_\tau^{(2)}(X_i, z_0)\mu_2 + \lambda\frac{(1-p_1)f(X_i|z_1)}{p_1 f(X_i|z_0)}\big[\mathbb{Q}_\tau(X_i, z_1) - \mathbb{Q}_\tau(X_i, z_0)\big],$$

$$b(X_i, z_1; h, \lambda) = \frac{1}{2}h^2\mathbb{Q}_\tau^{(2)}(X_i, z_1)\mu_2 + \lambda\frac{p_1 f(X_i|z_0)}{(1-p_1)f(X_i|z_1)}\big[\mathbb{Q}_\tau(X_i, z_0) - \mathbb{Q}_\tau(X_i, z_1)\big],$$

and

$$\sigma^2(X_i, z_0; h) = \frac{1}{nh}\cdot\frac{\tau(1-\tau)\nu_0}{f_e^2(0|X_i, z_0)p_1 f(X_i|z_0)},$$

$$\sigma^2(X_i, z_1; h) = \frac{1}{nh}\cdot\frac{\tau(1-\tau)\nu_0}{f_e^2(0|X_i, z_1)(1-p_1) f(X_i|z_1)}.$$

The following theorem gives the asymptotic expansion of $\mathbb{CV}(h, \lambda)$, which is critical to derive the asymptotic optimality of $\widehat{h}$ and $\widehat{\lambda}$.

THEOREM 2.2. *Suppose that the conditions of Theorem 2.1 and Assumption 4(ii) in Appendix A are satisfied. Then, we have*

$$\mathbb{CV}(h, \lambda) = \mathbb{CV}_1 + \frac{1}{2}\sum_{i=1}^{n}\big[b^2(X_i, Z_i; h, \lambda) + \sigma^2(X_i, Z_i; h)\big]M(X_i, Z_i)f_e(0|X_i, Z_i) + s.o., \tag{2.8}$$

where $\mathbb{CV}_1 \equiv \sum_{i=1}^{n}\rho_\tau(e_i)M(X_i, Z_i)$ *is unrelated to the tuning parameters and* $s.o.$ *represents some terms with smaller (asymptotic) probability order.*

Define

$$\mathbb{MSE}(h, \lambda) = \sum_{i=1}^{n}\big[\mathbb{Q}_\tau(X_i, Z_i) - \widehat{\mathbb{Q}}_{(-i)}(X_i, Z_i; h, \lambda)\big]^2 M(X_i, Z_i)f_e(0|X_i, Z_i). \tag{2.9}$$

If we further assume that e_i is independent of X_i and Z_i, $f_e(0|X_i, Z_i)$ would reduce to $f_e(0)$ the density function of e_i at point zero, which indicates that $f_e(0|X_i, Z_i)$ can be

removed in (2.9). Following the proof of Theorem 2.2, one can show that

$$\mathbb{MSE}(h,\lambda)=\sum_{i=1}^{n}\left[b^2(X_i,Z_i;h,\lambda)+\sigma^2(X_i,Z_i;h)\right]M(X_i,Z_i)f_e(0|X_i,Z_i)+s.o. \quad (2.10)$$

Letting h_* and λ_* be the minimizers to $\mathbb{MSE}(h,\lambda)$, by Theorem 2.2 above and standard argument (c.f., proof of Theorem 3.1 in Li and Racine, 2004), we have

$$\frac{\widehat{h}-h_*}{h_*}=o_P(1),\ \ \frac{\widehat{\lambda}-\lambda_*}{\lambda_*}=o_P(1), \quad (2.11)$$

which shows the asymptotic optimality of $\widehat{h}$ and $\widehat{\lambda}$.

2.3 A Simulation Study

In this section, we use simulations to examine the finite sample performance of the check function based local linear conditional quantile function estimation, where the bandwidths are chosen by the CV method. Furthermore, we compare our method with the traditional check function based local linear quantile estimation but only smoothing the continuous covariate (it thus splits the sample into cells according to different values of the discrete covariate), and the nonparametric inverse-CDF estimation, where the bandwidths for nonparametric CDF estimation are chosen by CV method (Li, Lin and Racine, 2013).

We consider the following two data generating processes:

$$\textbf{DGP 1}: \qquad Y_i=1+\sin(X_i)+\frac{1}{5}Z_i+|X_i|\,u_i,\ \ i=1,\cdots,n,$$

where $X_i\sim\mathbb{N}(0,1)$ and $Z_i\sim\mathbb{B}(3,0.5)$ (sum of 3 Bernoulli trials with success probability 0.5 for each trial), i.e., $Z_i\in\{0,1,2,3\}$ with $\mathbb{P}(Z_i=0)=0.5^3=1/8$, $\mathbb{P}(Z_i=1)=$

$3(0.5)^3 = 3/8$, $\mathbb{P}(Z_i = 2) = 3(0.5^3) = 3/8$ and $\mathbb{P}(Z_i = 3) = 0.5^3 = 1/8$; and

$$\textbf{DGP 2}: \qquad Y_i = \log(X_i + 1) + Z_{1i} + (X_i + 1)u_i, \;\; i = 1, \cdots, n,$$

where $X_i \sim \mathbb{U}[0, 1]$ and $Z_i \sim \mathbb{B}(3, 0.3)$, i.e., $Z_i \in \{0, 1, 2, 3\}$ with $\mathbb{P}(Z_1 = 0) = 0.7^3$, $\mathbb{P}(Z_1 = 1) = 3(0.3)(0.7)^2$, $\mathbb{P}(Z_1 = 2) = 3(0.3)^2(0.7)$ and $\mathbb{P}(Z_1 = 3) = 0.3^3$. We consider two distributions for the error term u_i: the student's t-distribution with 5 degrees of freedom denoted by $t(5)$, and the Laplace distribution which is denoted by $L(0, 1)$. The quantiles we consider in the simulation are $\tau = 0.10, 0.25, 0.50, 0.75, 0.90$. The sample sizes are $n = 100$ and $n = 200$, and the number of replications is 500.

Tables 2.1 and 2.2 (which correspond to the t-distribution and Laplace distribution for the error term, respectively) report the simulation results of the average MSE under DGP 1. For simplicity, in the tables, "method 1" stands for the naive local linear based check function method without smoothing over the discrete regressor, "method 2" stands for the local linear based check function method with bandwidths selected by the CV method, and "method 3" stands for the nonparametric inverse-CDF method proposed by Li, Lin and Racine (2013). From the two tables, we can see that our method which smoothes both continuous and discrete variables performs better than the naive method which only smoothes the continuous variable but does not smooth the discrete variables. This is similar to the findings by the existing literature in the context of conditional mean regression such as Hall, Li and Racine (2007). The main reason is that smoothing discrete variable can borrow data from the neighborhood to significantly reduce variance, while only introducing mild biases. As a result, the finite sample mean squared errors can be reduced. Meanwhile, we can also find from the two tables that our check function method performs better than the inverse-CDF method in particular at extreme quantiles.

Tables 2.3 and 2.4 (which correspond to the t-distribution and Laplace distribution for

the error term, respectively) report the simulation results of the average MSE under the DGP 2. We have the similar findings to those in Tables 2.1 and 2.2.

In addition to the smaller average MSE, our local linear based check function method also has the advantage of less computational time than the inverse-CDF method. Specifically, the method proposed in this paper involves $O(n^2)$ computations, while the inverse-CDF method requires $O(n^3)$ computations. As a result, our method takes 2.18 and 7.72 seconds per replication when $n = 100$ and $n = 200$, respectively, while the inverse-CDF method takes 2.30 and 14.74 seconds per replication. Such advantage in computational time would be more obvious as the sample size increases.

Table 2.1: Average MSE: DGP 1 with Distribution 1

Method	n/τ	0.10	0.25	0.50	0.75	0.90
Method 1	100	1.07	0.49	0.39	0.49	1.07
Method 1	200	0.76	0.35	0.27	0.33	0.72
Method 2	100	0.85	0.44	0.36	0.43	0.84
Method 2	200	0.62	0.32	0.28	0.33	0.59
Method 3	100	0.87	0.47	0.37	0.46	0.87
Method 3	200	0.68	0.37	0.30	0.36	0.66

Table 2.2: Average MSE: DGP 1 with Distribution 2

Method	n/τ	0.10	0.25	0.50	0.75	0.90
Method 1	100	1.33	0.59	0.41	0.58	1.34
Method 1	200	0.93	0.38	0.28	0.39	0.96
Method 2	100	1.04	0.46	0.37	0.47	1.03
Method 2	200	0.77	0.36	0.28	0.37	0.79
Method 3	100	1.08	0.53	0.40	0.57	1.07
Method 3	200	0.86	0.42	0.31	0.44	0.88

Table 2.3: Average MSE: DGP 2 with Distribution 1

Method	n/τ	0.10	0.25	0.50	0.75	0.90
Method 1	100	0.58	0.24	0.21	0.35	0.79
Method 1	200	0.30	0.12	0.11	0.22	0.52
Method 2	100	0.47	0.20	0.19	0.34	0.68
Method 2	200	0.29	0.12	0.12	0.21	0.48
Method 3	100	0.51	0.23	0.20	0.33	0.71
Method 3	200	0.30	0.14	0.12	0.19	0.45

Table 2.4: Average MSE: DGP 2 with Distribution 2

Method	n/τ	0.10	0.25	0.50	0.75	0.90
Method 1	100	0.87	0.29	0.18	0.42	1.13
Method 1	200	0.48	0.15	0.10	0.25	0.74
Method 2	100	0.68	0.25	0.16	0.35	0.89
Method 2	200	0.40	0.14	0.11	0.24	0.63
Method 3	100	0.75	0.28	0.18	0.37	0.97
Method 3	200	0.45	0.16	0.10	0.22	0.63

3.1 The Regression Model

3.1.1 Unconstrained and Constrained Kernel Estimators

Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ denote a sample of pairs of explanatory and dependent variables. We assume that $\{X_i, Y_i\}_{i=1}^n$ are independent and identically distributed. Consider a nonparametric regression model of the form

$$Y_i = g(X_i) + u_i, \qquad i = 1, ...n, \tag{3.1}$$

where $g(x) = E(Y|X = x)$ and the functional form of $g(x)$ is not specified.

Conventional unconstrained kernel estimators of $g(x)$ can be expressed as

$$\widetilde{g}(x) = \frac{1}{n}\sum_{i=1}^{n} A_i(x)Y_i, \tag{3.2}$$

where $A_i(x)$ is a weight function which depends only on x and observations X_is that are close to x. The forms of $A_i(x)$ and $\widetilde{g}(x)$ can be very general and include the local constant (Nadaraya (1965), Watson (1964)) estimator, the local polynomial estimator (Fan (1992)), the Priestley-Chao estimator (Priestley and Chao (1972)) and the Gasser–Müller estimator (Gasser and Müller (1979)), among others. For example, for the local constant estimator, we have

$$A_i = \frac{k(\frac{X_i - x}{h})}{\frac{1}{n}\sum_j k(\frac{X_j - x}{h})},$$

where $k(\cdot)$ is a bounded non-negative kernel function that satisfies

$$\begin{aligned} \int k(v)dv &= 1, \\ k(v) &= k(-v), \\ \int v^2 k(v)dv &= \mu_2, \end{aligned}$$

where μ_2 is a positive constant.

Some commonly used kernel functions include the Gaussian and Epanechnikov weight functions. The smoothing parameter h is usually selected using some data-driven methods such as the leave-one-out least squares cross-validation method, i.e.,

$$\widehat{h}_{LS-CV} = \arg\min_h CV(h) = \arg\min_h \frac{1}{n}\sum_{i=1}^{n}[Y_i - \widetilde{g}_{-i}(X_i)]^2 M(X_i),$$

where $\widetilde{g}_{-i}(X_i)$ is the leave-one-out kernel estimator of $g(X_i)$ and $M(X_i)$ is a compactly supported weight (trimming) function that avoids a zero value at the denominator of $CV(h)$ or large estimation bias at the boundary region.

In many economics applications, the functional relationship of a response variable and the explanatory variables is believed to satisfy some shape restrictions such as monotonicity and/or concavity. In such situations if the unconstrained estimator does not satisfy the constraint, one may want to use some estimation methods with constraints imposed to re-estimate the model. Hall and Huang (2001) and more recently, Du, Parmeter and Racine (2013) suggest methods for 'monotonizing' a kernel estimator. Hall and Huang (2001) propose a constrained kernel estimator of $g(x)$ based on the following weighted estimator of Y's.

$$\widehat{g}(x|p) = \sum_{i=1}^{n} p_i A_i(x) Y_i, \tag{3.3}$$

where $p = (p_1, ..., p_n)$ is a n-vector of weights attached to the data set $\{X_i\}_{i=1}^n$. The unconstrained estimator $\widetilde{g}(x)$ can be seen as a special case of (3) if we take a uniform weight $p = p_u = (1/n, ...1/n)$. When $p_i \neq 1/n$ for some i for the constrained estimator, we say that the constraint is binding.

To impose the monotone constraint while making the least possible change to the unconstrained estimated curve, one should select $p = \widehat{p}$ to minimize some distance function, say $D(p)$, subject to that the estimated derivative function is non-negative (non-positive) if $g(\cdot)$ is believed to be a monotone increasing (decreasing) function, i.e., one imposes that

$$\widehat{g}'(x|p) \geq 0 \qquad \text{(or } \widehat{g}'(x|p) \leq 0),$$

where $\widehat{g}'(x|p) = \partial \widehat{g}(x|p)/\partial x$ is the partial derivative of $\widehat{g}(x|p)$ with respect to x.

Following Hall and Huang (2001) we use $D(p)$ to denote a distance measure between a uniform weight $p_u = (1/n, ..., 1/n)$ and a non-uniform weight $p = (p_1, ..., p_n)$. The measure $D(p)$ should have the property that $D(p) \geq 0$ and $D(p_u) = 0$. Several different forms of distance function $D(p)$ are proposed in the literature. Hall and Huang (2001) consider probability weights and the power divergence distance measure (Cressie and Read 1984):

$$D_{HH,\rho}(p) = \frac{1}{\rho(1-\rho)} \left\{ n - \sum_{i=1}^{n} (np_i)^{\rho} \right\}, \tag{3.4}$$

for $-\infty < \rho < \infty$ and $\rho \neq 0, 1$, and

$$D_{HH,0}(p) = -\sum_{i=1}^{n} \log(np_i), \qquad D_{HH,1}(p) = \sum_{i=1}^{n} p_i \log(np_i).$$

with $\sum_{i=1}^{n} p_i = 1$ and $\min_{1 \leq i \leq n} p_i \geq 0$.

Du, Parmeter and Racine (2013) suggest using a different distance function $D_{DPR}(p)$

based on the L_2-distance (square) between p and the uniform weight p_u:

$$D_{DPR}(p) = (p - p_u)'(p - p_u). \tag{3.5}$$

Since Du, Parmeter and Racine (2013) do not assume p have to be probability weights, they allow for both positive and negative weights and without requiring that $\sum_{i=1}^{n} p_i = 1$.[1]

In this paper, instead of minimizing the distance between p and the uniform weights p_u, we propose using a new distance measure that directly measures the closeness between the constrained and unconstrained curves. We suggest selecting $p = \widehat{p}$ by minimizing an alternative $D(p)$, denoted as $D_{curve}(p)$, and is defined by

$$D_{curve}(p) = \sum_{i=1}^{n} (\widehat{g}(X_i|p) - \widetilde{g}(X_i))^2. \tag{3.6}$$

As in Du, Parmeter and Racine (2013), we do not require that either $\min_{1 \le i \le n} p_i \ge 0$, nor do we require that $\sum_{i=1}^{n} p_i = 1$.

One problem with the kernel method is that it effectively uses data points falling inside the interval $[x - h, x + h]$ (with a fixed interval length) when estimating $g(x)$. Since the bandwidth h is fixed, more observations are used to estimate $g(x)$ at the high density point (when the density function $f(x)$ is large), while much fewer observations are used at points when $f(x)$ is small. The constrained kernel estimator has the same problem. Therefore, in practice, sometimes the kernel method does not work well.

Let y_D be an $n \times n$ diagonal matrix with y_i in the i^{th} diagonal. Denote A as an $n \times n$ matrix with $A_{ij} = A_i(x_j)$. Then our objective function can be written as $(Ay_Dp - Ay_Dp_0)'(Ay_Dp - Ay_Dp_0)$. Hence, both D_{DPR} and D_{curve} are quadratic in p and the monotone constraints are linear in p. We are able to use quadratic programming in the

[1] We thank a referee pointing out that Du, Parmeter and Racine (2013) does not require $\sum_{i=1}^{n} p_i = 1$ and that this condition was listed in the paper is in fact a typo.

optimization. Quadratic programming is highly efficient in computation, which can solve large optimization problems in a short time. On the other hand, quadratic programming does not apply to HH method and we need to employ some alternative algorithms such as sequential quadratic programming to solve the optimization problem. Those alternative algorithms have much lower computational efficiency than quadratic programming, see Henderson and Parmeter (2015, Chapter 12) for a detailed discussion on this. Therefore, for the computational efficiency ranking, our proposed curve method is as efficient as the DPR method and both of them are much more efficient than the HH method.

3.1.2 The Knn Method

In order to have enough observed data in estimating the regression function at each evaluation point, including the low density range, one can use the Knn estimation method. The constrained Knn estimator can be constructed in a similar way as the constrained kernel estimator.

We use R_x to denote the Knn distance to x which is defined as

$$R_x = \text{ the Euclidean distance between } x \text{ and its } k^{th} \text{ nearest neighbor among } \{X_i\}_{i=1}^n.$$

We still use equation (3) to construct the constrained estimator, while the forms of $A_i(x)$ need to be changed since the bandwidth is now stochastic. For the local constant method, we have

$$A_i = \frac{K(\frac{X_i - x}{R_x})}{\frac{1}{n}\sum_j K(\frac{X_j - x}{R_x})},$$

where $K(\cdot)$ is a kernel function such as the Gaussian or Epanechnikov weight functions. We see that the difference between a kernel and a Knn estimator is that the non-random bandwidth h in the kernel estimator is replaced by a random bandwidth R_x.[2]

[2]The asymptotic analysis for the Knn method is more complex than that for the kernel method, see Mack and Rosenblatt (1979), Mack (1981), Ouyang, Li and Li (2006), Fan and Liu (2015), among others.

For the Knn estimation method, Li (1987) established the optimality property of using the the least-squares cross-validation (LS-CV) method to select k. Therefore, we will use the LS-CV method to select k:

$$\widehat{k}_{LS-CV} = \arg\min_{k} CV(k) = \arg\min_{k} \frac{1}{n}\sum_{i=1}^{n}[Y_i - \widetilde{g}_{-i}(X_i)]^2 M(X_i)$$

where $\widetilde{g}_{-i}(X_i)$ is the leave-one-out Knn estimator of $g(X_i)$ and $M(\cdot)$ is a trimming weight function.

3.2 Simulation Results

3.2.1 The Data Generating Process

This section provides simulation results. We consider the following data generating processes (DGPs):

$$DGP1 : Y_i = ln(X_i) + u_i, \qquad i = 1, ..., n,$$

$$DGP2 : Y_i = ln(Z_i) + u_i, \qquad i = 1, ..., n,$$

$$DGP3 : Y_i = sin(X_i) - ln(X_i) + u_i, \qquad i = 1, ..., n,$$

$$DGP4 : Y_i = sin(Z_i) - ln(Z_i) + u_i, \qquad i = 1, ..., n,$$

where $u_i \sim N(0, \sigma_u^2)$ with $\sigma_u = 1/2$. The regressor X is unevenly distributed in its support and is generated via:

$$\begin{aligned} X_i &\sim \text{Uniform}[\pi, 1.5\pi] \text{ for } i = 1, ..., 0.8n, \\ X_i &\sim \text{Uniform}[1.5\pi, 3\pi] \text{ for } i = 0.8n + 1, ..., n. \end{aligned}$$

The regressor Z is evenly distributed with Uniform[π,3π]. $\{u_i\}_{i=1}^n$, $\{Z_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$ are independent with each other.

We use DGP1 and DGP2 to study the performance of kernel and Knn methods with unevenly (DGP1) and evenly (DGP2) distributed data. DGP1 and DGP2 also serve to examine size performance of our test for monotonicity and DGP3 and DGP4 serve to study the power performance of our monotonicity test.

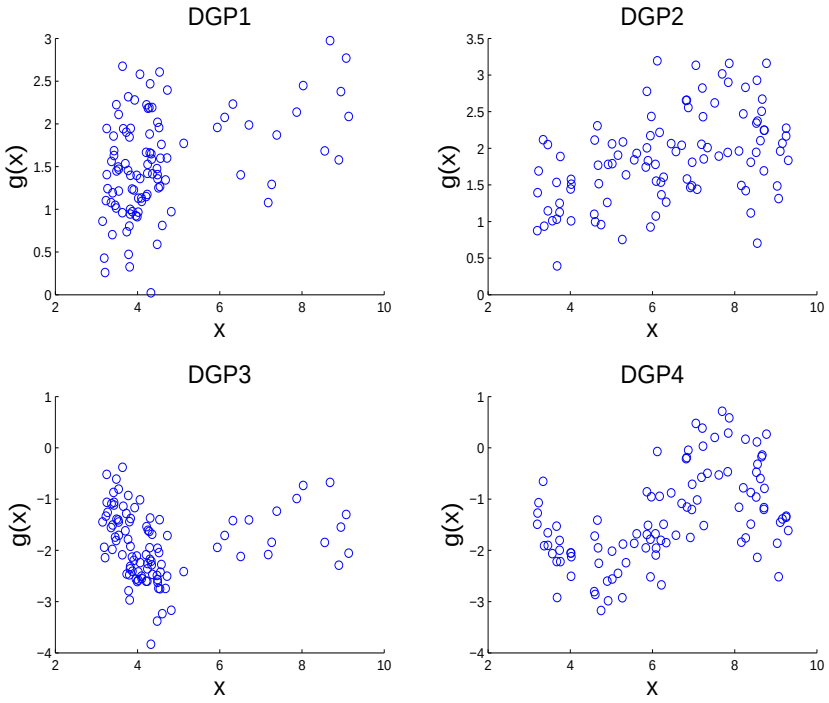


Figure 3.1: Scatter of Data Generating Processes

Figure 3.1 shows the data points generated from DGP 1 to DGP4. We can see that points generated by DGP1 and DGP3 are highly unevenly distributed while DGP2 and DGP4 generate evenly distributed data. With sparse data at the tail of the distribution, the kernel approach may not be a suitable method (when using a fixed bandwidth) for

analyzing such type a data. Therefore, we will mainly use the Knn method to estimate the regression model and to test the null hypothesis of monotonicity of $g(\cdot)$. To compare the Knn estimation/testing results with those of the kernel method, we also include a constrained kernel estimator based on our new distance measure (we term it as the kernel-curve estimator or simply kernel estimator).

We estimate $\widehat{g}(x|p)$ using the local constant estimation method with a Gaussian kernel as the weight function and least-squares cross-validation for k selection

$$\widehat{k}_{LS-CV} = \arg\min_{k} CV(k) = \arg\min_{k} \frac{1}{n}\sum_{i=1}^{n}[Y_i - \widetilde{g}_{-i}(X_i)]^2 M(X_i).$$

We consider three forms of distance measure in our experiments: Hall and Huang's distance function $D_{HH,\rho}(p)$, Du, Parmeter and Racine's distance function $D_{DPR}(p)$ and our new distance function $D_{curve}(p)$. For Hall and Huang's distance function $D_{HH,\rho}(p)$, we set the parameter $\rho = 0.5$.

All simulation results are based on 1000 replications, with results reported for sample sizes of $n = 50$, 100 and 200.

3.2.2 The Performance in Estimation

First we compare the MSEs of five estimation methods: kernel-curve-local-constant (Kernel-LC, we omit 'curve' in our short hand notation to save spaces in tables), the Knn-curve-local-constant (Knn-LC), the Knn-curve-local-cubic (Knn-L-Cubic),[3] the Knn-HH and Knn-DPR estimators. Specifically, we compute the sample mean square errors as follows

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{g}(X_i) - g(X_i)\right]^2,$$

[3]We would like to thank a referee for pointing out that local-polynomial estimates would serve better as criterion in the curve method.

where $n \in \{50,\ 100,\ 200\}$ is the sample size, $M = 1000$ is the number of simulations and $\hat{g}(X_i)$ is one of the five estimators: kernel-curve-local-constant, Knn-curve-local-constant, knn-curve-local-cubic, Knn-HH and Knn-DPR.

Table 3.1 reports the results when data is unevenly distributed (DGP1) while Table 3.2 reports the results when data is evenly distributed (DGP2). The MSEs are reported for the 25%, 50% and 75% quantiles. First we focus on the unevenly distribute data case of DGP1. From Table 3.1 we see that the the constrained Knn local cubic estimator performs the best, followed by the the constrained Knn local constant estimator.

Next, we examine the estimation results for evenly distributed data (DGP2). From Table 3.2 we observe that when $n = 50$, the constrained kernel local constant based estimator has the smallest estimation MSE. However, when $n = 100$ and $n = 200$, the Knn-HH performs the best. When sample size is large ($n = 200$), the MSEs of different estimation methods become close to each other.

Table 3.3 reports the proportion of points violating the monotonicity constraint under DGP1 and DGP2 with unconstrained kernel and Knn methods. Let $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ be the order statistics, we say that $g(\cdot)$ violates the monotone (increasing) condition at $x_{(i)}$ if $g(x_{(i)}) < g(x_{(i-1)})$. From Table 3.3 we observe that for both unevenly and evenly distributed data cases, the Knn method has more points violating the monotone constraint than the kernel method. This result suggests that the better performance of the Knn method (when data are unevenly distributed) is not due to having less points violating the monotone constraint, rather it is due to its more accurate estimation of the unknown regression function. We observe the same phenomenon in our empirical application using the 'Job Matching data'.

Table 3.1: MSE-DGP1 (Unevenly Distributed)

Quantile	Kernel-LC	Knn-LC	Knn-L-Cubic	Knn-HH	Knn-DPR
			$n = 50$		
0.25	0.0124	0.0117	0.0110	0.0129	0.0148
0.50	0.0194	0.0182	0.0172	0.0217	0.0294
0.75	0.0285	0.0279	0.0273	0.0384	0.0579
			$n = 100$		
0.25	0.0076	0.0072	0.0068	0.0071	0.0092
0.50	0.0121	0.0107	0.0100	0.0116	0.0156
0.75	0.0164	0.0154	0.0147	0.0187	0.0272
			$n = 200$		
0.25	0.0046	0.0047	0.0042	0.0049	0.0053
0.50	0.0072	0.0064	0.0061	0.0077	0.0083
0.75	0.0101	0.0093	0.0087	0.0128	0.0130

Table 3.2: MSE-DGP2 (Evenly Distributed)

Quantile	Kernel-LC	Knn-LC	Knn-L-Cubic	Knn-HH	Knn-DPR
			$n = 50$		
0.25	0.0084	0.0091	0.0099	0.0088	0.0089
0.50	0.0143	0.0156	0.0169	0.0164	0.0158
0.75	0.0238	0.0247	0.0262	0.0311	0.0292
			$n = 100$		
0.25	0.0058	0.0057	0.0057	0.0052	0.0053
0.50	0.0088	0.0091	0.0097	0.0081	0.0087
0.75	0.0138	0.0141	0.0148	0.0130	0.0141
			$n = 200$		
0.25	0.0033	0.0033	0.0033	0.0028	0.0029
0.50	0.0052	0.0051	0.0051	0.0045	0.0048
0.75	0.0076	0.0075	0.0076	0.0070	0.0072

3.2.3 The Performance in Testing

3.2.3.1 The Bootstrap Procedure

To test the monotonic relationship between a dependent variable and an explanatory variable (say, the relationship of job matching probability and market size of Gan and Li

Table 3.3: Propotion of Points Violating Constraints

	Kernel-LC	Knn-LC	Kernel-LC	Knn-LC
n	DGP1		DGP2	
50	0.071	0.342	0.095	0.216
100	0.081	0.391	0.092	0.195
200	0.074	0.383	0.082	0.169

(2016)), following Du, Parmeter and Racine (2013) we use a bootstrap testing procedure to obtain critical values. We note that Du, Parmeter and Racine (2013) has shown that their test statistic $D_{DPR}(\widehat{p})$ has a limiting χ^2-distribution under the null hypothesis that the regression function is a monotone function. We leave the theoretical investigation of the limiting distribution of our $D_{curve}(\widehat{p})$ to a possible future research topic. Below we describe the bootstrap procedure of our test.

First, We use the Knn method and one of the three distance functions $D_{HH,\rho}(p)$, $D_{DPR}(p)$ and $D_{curve}(p)$ to estimate the constrained regression function $\widehat{g}(x|p)$ using the Knn method based on the observations $\{Y_i, X_i\}_{i=1}^n$, and get $\widehat{p}$ and $D(\widehat{p})$, where $D(\widehat{p})$ denotes $D_{HH,\rho}(\widehat{p})$, or $D_{DPR}(\widehat{p})$ or $D_{curve}(\widehat{p})$. Then we use $D(\widehat{p})$ as the test statistic and reject the null hypothesis if $D(\widehat{p})$ is too large. The resampling approach involves generating resamples for Y_i^* using iid residual resampling. Note that the bootstrap sample $(Y_i^*, X_i)_{i=1}^n$ must satisfy a monotone relationship, i.e., one needs to impose the null hypothesis when generating Y_i^*. The bootstrap steps are as follows:

(i) Estimate $\widehat{g}(x|p)$ under the monotone constraint and obtain residuals: $\widehat{u}_i = Y_i - \widehat{g}(X_i|\widehat{p})$ for $i = 1, ..., n$.

(ii) For each i, generate $Y_i^* = \widehat{g}(X_i|\widehat{p}) + \widehat{u}_i^*$, where $\widehat{u}_i^*$ is drawn randomly from $\{\widehat{u}_1 - \frac{1}{n}\sum_{j=1}^n \widehat{u}_j, ..., \widehat{u}_n - \frac{1}{n}\sum_{j=1}^n \widehat{u}_j\}$.

(iii) Use the bootstrap sample $\{Y_i^*, X_i\}$ to recompute $\widehat{g}(x|p)$, denoted $\widehat{g}(x|p^*)$, and obtain $D(p^*)$.

(iv) Repeat steps (ii) and (iii) $B-1$ times and then construct the empirical distribution of the B bootstrap statistics (by adding the original statistic to the $B-1$ bootstrap statistics), $\{D(p_j^*)\}_{j=1}^B$. We reject the null hypothesis if $D(\widehat{p}) > D(p^*_{(\alpha B)})$, where $D(p^*_{(\alpha B)})$ is the upper α-percentile of $\{D(p_j^*)\}_{j=1}^B$.

We use Monte Carlo simulations to examine the finite sample performance of the bootstrap testing procedure. The number of simulations is 1000. We follow the bootstrap testing procedure stated above and take $B-1=199$.[4]

Remark: When $D(\hat{p})=0$, the bootstrap is degenerated and the decision must be not rejected. When assessing the test size and test power, this degeneration of bootstrap will lead to a downward bias of rejection rate. To correct this bias, one can simply drop out the cases when $D(\hat{p})=0$ in simulations.[5]

3.2.3.2 The Estimated Sizes

We first simulate the data under the null hypothesis of a monotone regression function, i.e., a monotone increasing function: $g(X_i)=\ln(X_i)$. The simulation results are provided for different estimation methods (Kernel and Knn) and for different forms of distance measures (curve, HH and DPR): Kernel-LC, Knn-LC, Knn-HH and Knn-DPR. Tables 3.4 reports empirical rejection rates of DGP1 (unevenly) for the bootstrap tests with nominal significance levels (α) of 1%, 5% and 10%, using $D_{HH,0.5}(p)$, $D_{DPR}(p)$ and $D_{curve}(p)$ as the distance measures, respectively. The simulation results present two noteworthy points. First, no matter which distance function among $D_{HH,0.5}(p)$, $D_{DPR}(p)$ and $D_{curve}(p)$ are used, the rejection rates are close to their nominal values, even for a small sample size, e.g., $n=50$. Secondly, when comparing the sizes of the four tests with different distance

[4]In the simulation part, since the null hypothesis is a monotonic increasing function, we estimate all $\widehat{g}(x|p)$ and $\widehat{g}(x|p^*)$ by selecting $p=\widehat{p}$ or p^* to minimize some distance functions $D(p)$, while imposing that derivative function is always non-negative.

[5]We would like to thank Jeffrey Racine for his helpful suggestion about correcting bias in estimating test size.

measures, we can see overall that the test using our new distance function $D_{curve}(p)$ shows the closest rejection rates to the nominal significance levels (α), although all four tests give reliable rejection rates. Table 3.5 shows the empirical rejection rates of DGP2 (evenly distributed data) for the same settings of Table 3.4. All methods we discuss here give normal rejection rates. We next examine performances of the tests under the alternative hypothesis, that is, the power of the tests.

Table 3.4: Estimated Test Size (DGP1)

α	Kernel-LC	Knn-LC	Knn-HH	Knn-DPR
		$n = 50$		
0.01	0.002	0.010	0.006	0.007
0.05	0.058	0.050	0.037	0.040
0.10	0.117	0.117	0.085	0.079
		$n = 100$		
0.01	0.009	0.011	0.006	0.009
0.05	0.048	0.053	0.040	0.035
0.10	0.099	0.105	0.099	0.092
		$n = 200$		
0.01	0.008	0.012	0.008	0.009
0.05	0.041	0.050	0.033	0.036
0.10	0.100	0.097	0.080	0.090

3.2.3.3 The Estimated Powers

We consider the powers of the bootstrap tests against the alternative hypothesis DGP3, i.e., $g(X_i) = \sin(X_i) - ln(X_i)$, which is not monotonic on the interval $[\pi, 3\pi]$. The simulation results for the tests with different forms of distance measures and different regression models are reported in Tables 3.6. As expected, all four tests show reasonable power and the power increases in both sample size and nominal significance levels (α). Comparing the four tests with different distance measures and different regression models, we find

Table 3.5: Estimated Test Size (DGP2)

α	Kernel-LC	Knn-LC	Knn-HH	Knn-DPR
		$n = 50$		
0.01	0.016	0.011	0.011	0.011
0.05	0.062	0.058	0.055	0.061
0.10	0.109	0.103	0.120	0.117
		$n = 100$		
0.01	0.007	0.005	0.010	0.007
0.05	0.055	0.037	0.057	0.057
0.10	0.117	0.085	0.135	0.133
		$n = 200$		
0.01	0.007	0.009	0.012	0.010
0.05	0.049	0.043	0.055	0.057
0.10	0.100	0.078	0.114	0.111

the tests of Knn-HH and Knn-DPR show the weaker power than the tests of Kernel-LC and Knn-LC. This is due to the advantage of our new measure of distance. If we compare Kernel-LC and Knn-LC, we can see that Knn-LC outperforms Kernel-LC, especially when $\alpha = 0.01$. These results suggest that Knn method has advantage against kernel method when data is unevenly distributed.

In addition, we examine the case when the data is evenly distributed. Table 3.7 reports the empirical rejection rates with DGP4 with the same settings of Table 3.6. In this case, the advantage of Knn method vanishes as expected, but the advantage of curve method still plays its role. Speaking of numbers, Kernel-LC has the highest test power followed by Knn-LC, while Knn-HH and Knn-DPR have lower test power.

3.3 An Empirical Study

3.3.1 The Data

In this section we use the 'Job Market Matching' data taken from Gan and Li (2016) to illustrate that when a data is highly unevenly distributed, the kernel method may not be

Table 3.6: Estimated Test Power (DGP3)

α	Kernel-LC	Knn-LC	Knn-HH	Knn-DPR
	$n = 50$			
0.01	0.702	0.865	0.207	0.225
0.05	0.940	0.966	0.522	0.543
0.10	0.988	0.989	0.678	0.728
	$n = 100$			
0.01	0.885	0.962	0.635	0.630
0.05	0.987	0.993	0.878	0.894
0.10	0.997	0.996	0.941	0.955
	$n = 200$			
0.01	0.970	0.991	0.925	0.936
0.05	0.994	1.000	0.990	0.992
0.10	1.000	1.000	0.995	0.998

Table 3.7: Estimated Test Power (DGP4)

α	Kernel-LC	Knn-LC	Knn-HH	Knn-DPR
	$n = 50$			
0.01	0.823	0.615	0.138	0.189
0.05	0.966	0.832	0.409	0.515
0.10	0.993	0.932	0.612	0.700
	$n = 100$			
0.01	0.994	0.923	0.510	0.586
0.05	0.999	0.979	0.819	0.871
0.10	1.000	0.989	0.924	0.943
	$n = 200$			
0.01	1.000	0.990	0.926	0.941
0.05	1.000	0.997	0.989	0.991
0.10	1.000	0.998	0.995	0.998

the appropriate approach to analyze such a data. Gan and Li collect data from 1999-2000 and 2000-2001 job openings/candidates market for economists and organize new Ph.D economists' market by field. This data set contains three variables described as follows:

Number of job openings (V): the total number of job opening in the top 50 U.S.

universities by each field in 1999 and 2000. The definition of 'field' can be found in the "Classification System of Journal Articles" by the *Journal of Economic Literature*. In particular, they label the field with a capital letter and a numeral. For example, E0 means 'Macroeconomics and Monetary Economics'.

Number of filled positions (M): the number of filled positions describes how many academic job openings in 1999 and 2000 for the top 50 economic departments in the U.S. were finally filled for each field in the following years (in 2000 and 2001).

Number of candidates (U): the number of candidates is obtained by searching the links of job candidates in each of the top 50 economic departments in the U.S. The first field listed in each candidate's CV or his/her brief research statement is used to determine the field of each candidate.

We are interested in the relationship between the matching probability (y) (percentage of candidates finding academic jobs) and the market size (x). Therefore, following Gan and Li (2016) we use $Y_{it} = M_{it}/V_{it}$ as the dependent variable, where M_{it} is the number of positions filled in field i at year t ($t = 1, 2$ correspond to 1999-2000, 2000-2001), V_{it} is the number of openings for field i and time t, the explanatory variable is defined as $X_{it} = (U_{it}^2 + V_{it}^2)^{1/2}$, where U_{it} is the total number of job candidates in field i at year t. X_{it} describe the thickness for field i at year t. A large value of X_{it} means a large market size for field i at time t. Gan and Li (2016) show that the theoretical matching probability should be an increasing function in market size. The sample size for this data set is $n = 128$. The data points are plotted in Figure 3.2.

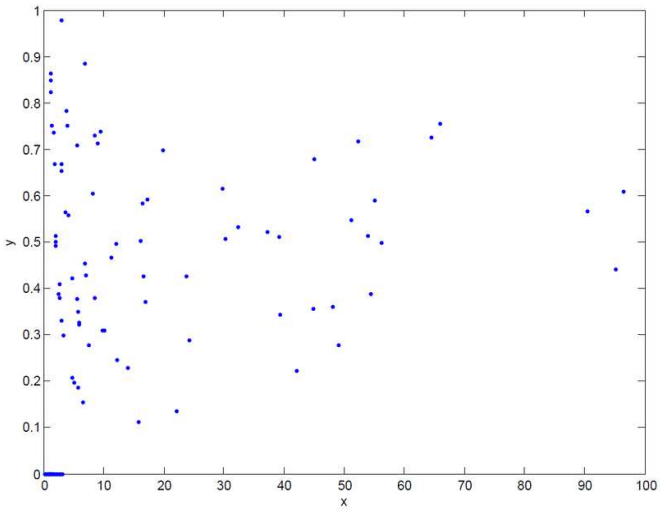


Figure 3.2: Job Matching Data

3.3.2 The NLS Estimation

Gan and Li (2016) prove that the probabilities of matches in a thin market are significantly lower than those in a thick market. In a regression framework we have

$$Y_{it} = g(X_{it}) + u_{it} \tag{3.7}$$

then $g(x)$ should be an (monotone) increasing function in x. Similar to Gan and Li (2016) we propose using a simple parametric model to study the relationship between job matching probability Y and market size X:

$$Y_{it} = \alpha_1 + \alpha_2 / X_{it}^{\alpha_3} + u_{it}. \tag{3.8}$$

The monotonic increasing relationship between Y and X implies that $\alpha_2 < 0$ and $\alpha_3 > 0$. Also, the non-negativity of matching probability requires that $\alpha_1 > 0$. Using 1999

and 2000 economist job market data to estimate model (8) gives the following results:

$$\hat{\alpha}_1 = 0.683, \quad t\text{-stat} = 3.72, \quad p\text{-value} = 0.000;$$
$$\hat{\alpha}_2 = -0.507, \; t\text{-stat} = -2.55, \; p\text{-value} = 0.012;$$
$$\hat{\alpha}_3 = 0.297, \quad t\text{-stat} = 2.03, \quad p\text{-value} = 0.045.$$

We see that all the estimated coefficients have the predicted signs and they are all statistically significantly different from zero at the 5% level. We plot in Figure 3.3 the parametric model fitted curve: $\hat{\alpha}_1 + \hat{\alpha}_2 / X_{it}^{\hat{\alpha}_3}$ with the above reported estimated coefficients.

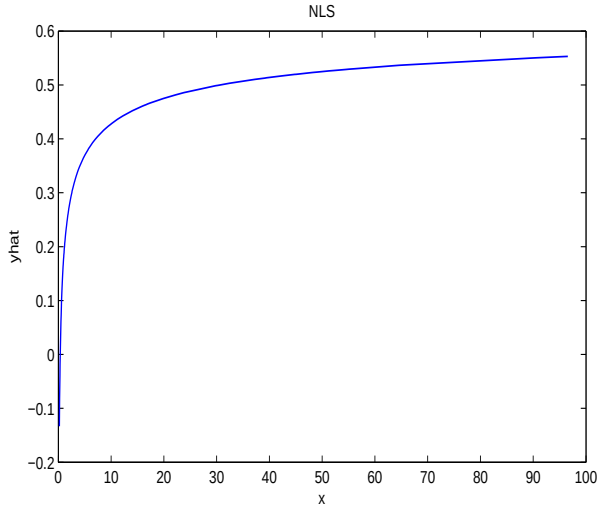


Figure 3.3: NLS Estimation Result

Figure 3.3 confirms that job matching probability is an increasing function in market size. However, model (8) imposes a strong parametric regression functional form. To examine whether the finding that matching probability is a monotone increasing function of

market size is not influenced by the specific chosen functional form, we turn to using nonparametric methods to analyze this data and examine whether the monotone relationship between job matching probability and market size is supported by nonparametric estimation results that are robust to regression functional form specifications.

3.3.3 The Kernel Estimation

Below we first use unconstrained and constrained kernel methods to estimate the regression function $g(\cdot)$. For smoothing parameter selection, one of the most commonly used method is the least-squares cross-validation method, see Härdle et al. (1988, 1992) for the optimality property of using the least-squares cross-validation method to select smoothing parameter h. Therefore, we use the local constant estimation method with the least-squares cross-validation for bandwidth selection, i.e., we select the smoothing parameter h via

$$\widehat{h}_{LS-CV} = \arg\min CV(h) = \frac{1}{n}\sum_{t=1}^{2}\sum_{i=1}^{n_t}[Y_{it} - \widetilde{g}_{-(it)}(X_{it})]^2,$$

where $\widetilde{g}_{-(it)}(X_{it})$ is the leave-one-out kernel estimator of $g(X_{it})$, $n = n_1 + n_2$ and n_t is the number of observations at time period t ($t = 1, 2$). When applying a Gaussian kernel, the cross-validated bandwidth $\widehat{h}_{LS-CV} = 0.8389$. While for constrained estimators $\widehat{h}_{LS-CV}$ is too small to find feasible solutions satisfying the constraints, we have to increase the bandwidths in order to obtain constrained estimated curves that satisfy the monotone restriction. The unconstrained and constrained kernel estimation results using different distance measures are presented in Figures 3.4, 3.5 and 3.6.

The smoothing parameter is selected by the least squares cross validation method. Figures 3.4 - 3.6 show the unconstrained kernel estimated curve (without imposing the monotone constraint) is too wiggly and does not seem to describe a reasonable relation-

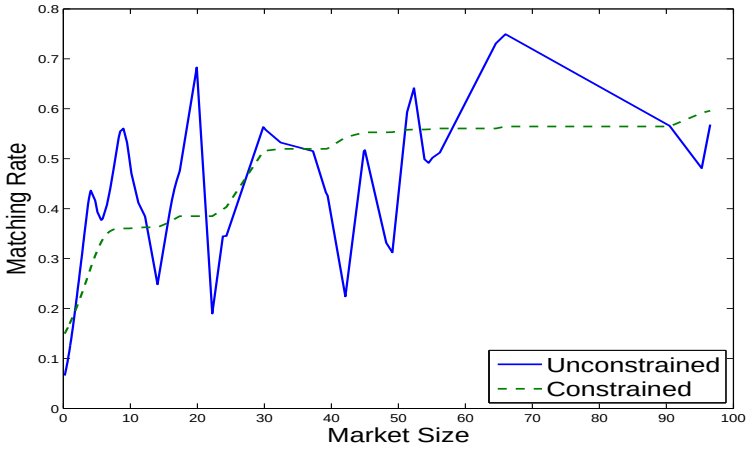


Figure 3.4: Kernel Estimation: $D_{HH}(p)$

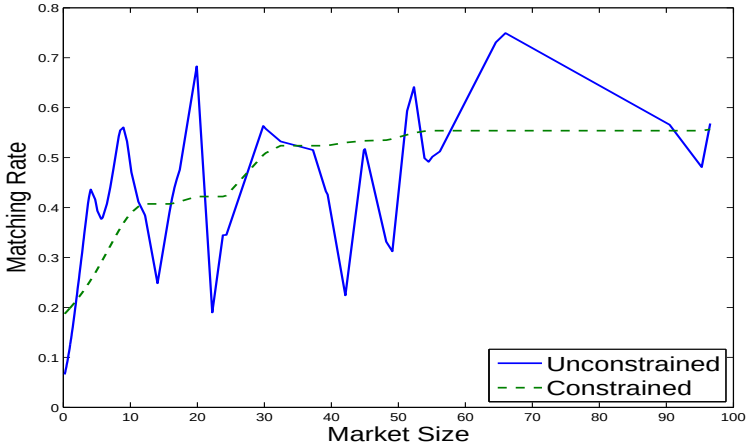


Figure 3.5: Kernel Estimation: $D_{DPR}(p)$

ship between the matching probability and the market size. This data is highly unevenly distributed. Most of X_i's (103 of 128) are within 20, while only 5 points lie between 60 and 96. When fixing the bandwidth, there is not enough data to use at the tail of the distribution. Therefore, the Knn method may be more suitable to this kind of data application. We next present the Knn unconstrained and constrained estimation results at the next section.

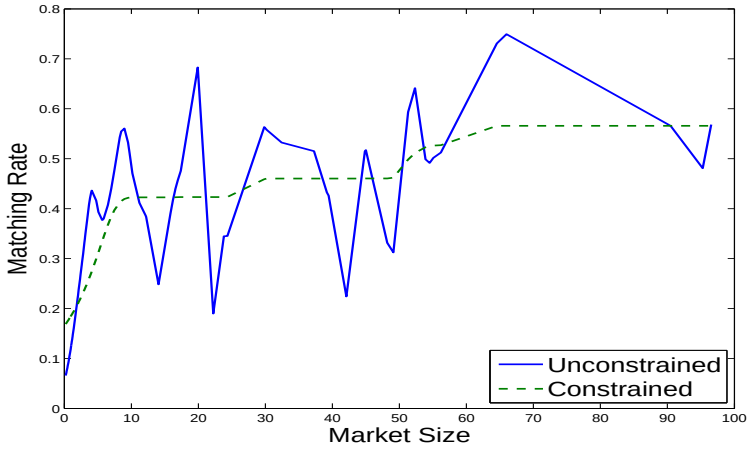


Figure 3.6: Kernel Estimation: $D_{curve}(p)$ with Local-Constant

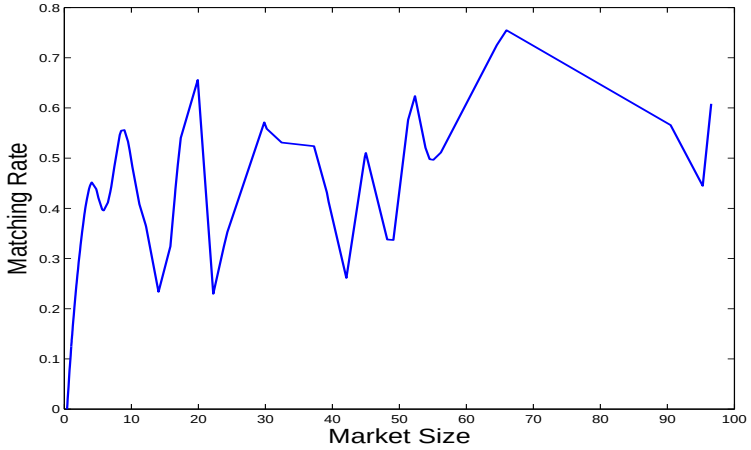


Figure 3.7: Unconstrained Kernel Estimation: Local-Cubic

We also consider the kernel estimation of local cubic and the result is plotted in Figure 3.7. We can see that with this empirical data set, local cubic estimated curve does not deviate much from that of the local constant method. The two curves are actually very close to each other. To save space, we will only report local constant estimation results in the remaining part of this paper.

3.3.4 The Knn Estimation

This section reports Knn method estimation results for the 'Job Market Matching' data. We use a Gaussian kernel and use the LS-CV method to choose K. The LS-CV selects $\widehat{k}_{LS-CV} = 9$. When conducting local constant estimation with $\widehat{k}_{LS-CV} = 9$, the unconstrained and constrained Knn estimated curves using different distance measures are presented in Figures 3.8, 3.9 and 3.10. We would like to emphasize the importance of imposing monotonicity constraints with this data set. From figure 3.8 we observe that the unconstrained Knn estimated curve has two peaks around market size $x = 5$ and $x = 10$. Based on the theoretical result of Gan and Li (2016), we believe that the constrained curve more accurately reflects the true regression functional form, the matching probability at market size $x = 5$ ($x = 10$) should be similar to that of x close to 5 (10) rather than significantly higher than that with similar market size. The observed peaks are likely due to sampling errors.

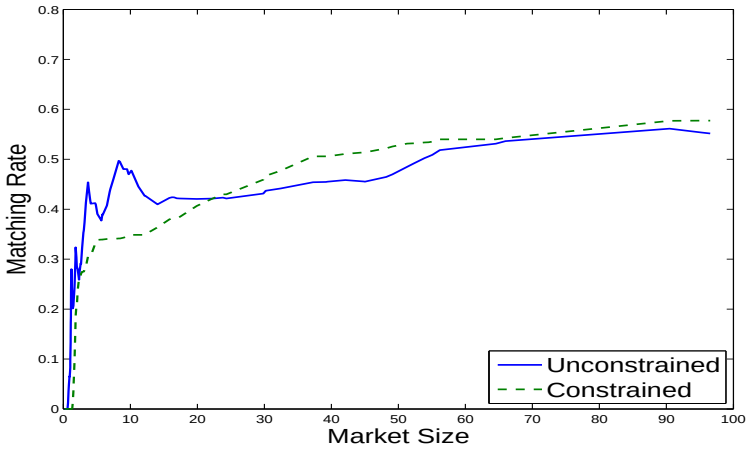


Figure 3.8: Knn Estimation: $D_{HH}(p)$

Comparing Figures 3.8 - 3.10 with Figures 3.4 - 3.6, we see that the Knn method gives

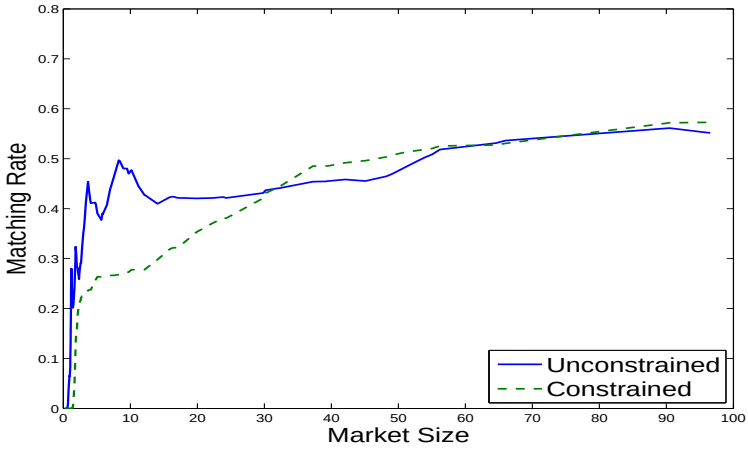


Figure 3.9: Knn Estimation: $D_{DPR}(p)$

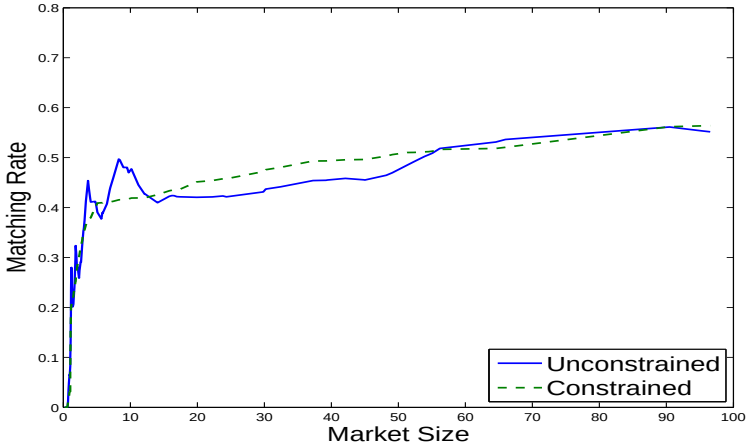


Figure 3.10: Knn Estimation: $D_{curve}(p)$

more reasonable estimated curves than those obtained by using the kernel method. Moreover, when using our proposed distance measure (equation (3.6)), the constrained curve is quite close to the unconstrained curve. We compute the empirical L_2 distance of the three constrained Knn curves to the unconstrained Knn curve via $L_2 = \{n^{-1}\sum_{i=1}^{n}[\hat{g}(X_i) - \tilde{g}(X_i)]^2\}^{1/2}$, where $\hat{g}(X_i)$ is the constrained estimator (it can be HH, DPR or curve) and $\tilde{g}(X_i)$ is the unconstrained estimator. The L_2 distances between the unconstrained Knn

curve and the three constrained curves are $L_{2,HH} = 0.1217$, $L_{2,DPR} = 0.1470$ and $L_{2,Curve} = 0.0412$, respectively. This confirms the fact that our curved based estimate gives the closest fit to the unconstrained curve. The closeness between the constrained and the unconstrained curves also suggests that the monotonic relationship between the job matching probability and the market size (field thickness) is likely to be supported by the 'Job Market Matching' data.

Table 3.8: Proportion of Points Violating Constraints

Violating/Total	Kernel	Knn
Observation	22/128	40/128
Grid	33/97	37/97

We also compute the number of data points that violate the monotone constraints. Table 3.8 reports number of data points that violate the monotone restriction by the unconstrained kernel and the Knn methods. Out of the 128 observations, the kernel method has 22 points violating the monotone restriction, while the Knn method gives 40 violation points. We also used grid points rather than data points in evaluating the unconstrained kernel and Knn estimators. The grid points are evenly distributed grids (x=1, 2, 3, ..., 97). Over the 97 grid evaluations points, the Kernel has 33 points and the Knn has 37 points violating the monotone restriction, respectively. Even the unconstrained Knn estimated curve fits the unconstrained monotone curve quite well, it does not mean that it has a smaller percentage of evaluation points that violate the monotone constrain. It only means that the magnitudes of deviations from the monotone curve are relatively small (compared to those of kernel method estimated curve). Indeed, the relatively smooth Knn estimated curve is quite flat when the data is sparse, there is no guarantee the smooth curve is increasing. It can decrease at many observation or grid evaluation points which lead to a

large number of points violating the monotone constraints. On the other hand, Kernel give very wiggly curve. It can also have many points violating the monotone constraint by large magnitude, but there is no guarantee that it has fewer points violating the monotone constraint. In fact, for the 'Job Market Matching' data, it has less violation points than the that using the Knn method. In summary, the percentage of evaluation points that violate the monotone constraint may not be a good measure of whether the constrained curve is close to the unconstrained curve.

In the next section, we test the null hypothesis that whether the regression function $g(x)$ is a monotonically increasing function (in x) based on the Knn estimation method and use the bootstrap testing procedure discussed in section 3.

3.3.5 The Testing Result

In section 3.3.2, we show that when applying a parametric model to analyze the economists' job matching data, the estimated job matching probability is an increasing function in market size. The next question is whether the monotonic relationship between job matching probability and market size is supported by the robust nonparametric estimation method. Figures 3.8 - 3.10 present the unconstrained and constrained Knn estimated curves using different distance measures. There we see that the constrained curves are close to the unconstrained curves, especially when our new distance measure is employed, suggesting that the monotonic relationship is likely to be true. In this section, we apply the proposed bootstrap testing procedure to formally test the null hypothesis that the job matching probability function is a monotonically increasing function in market size (field thickness) using the 'Job Market Matching' data. Tests with different distance measures are considered. We estimate the constrained regression function $\widehat{g}(x|p)$ using local constant estimation method with a Gaussian kernel as the weight function and least-squares cross-validation method for selecting k. The number of bootstraps is $B - 1 = 999$. Ta-

ble 3.9 reports the testing results. In Table 3.9, D represents the distance measure, which could be $D_{HH,0.5}$ (Hall and Huang's distance function with $\rho = 0.5$), D_{DPR} (Du, Parmeter and Racine's distance function) or D_{curve} (our new distance function). $D(\widehat{p})$ is the test statistic, obtained by using the three different distance measures respectively.

Table 3.9: Bootstrap Tests with Job Matching Data

D	$D_{HH,0.5}$	D_{DPR}	D_{curve}
$D(\widehat{p})$	33.49	43.36	0.13
p-value	0.311	0.065	0.134

As shown in Table 3.9, all three tests with different distance measures have p-values that are greater than 5%. For the test with our new distance measure, the p-value$= 0.134$, which is even greater than 10%. Therefore, we cannot reject the null hypothesis that job matching probability is a monotonically increasing function of market size at the 5% significance level. We confirm the monotone relationship between job matching probability and market size by using the nonparametric Knn method.

4. NONPARAMETRIC IDENTIFICATION AND TESTING OF FIRST-PRICE AUCTIONS WITH ASYMMETRIC BIDDERS

4.1 The Models

In this section, we set up the two alternative models of first-price sealed-bid auctions where bidders display unobserved heterogeneity. We first propose a general baseline model, in which bidders can have arbitrary heterogeneity in preferences and value distributions. Then we restrict the heterogeneity into preference-only or value-distribution-only, which leads to two specific models: asymmetric preferences (AP) model and asymmetric value distributions (AVD) model.

4.1.1 Baseline First-Price Auction Model

A single and indivisible object is sold by auction. All bids are sealed and collected simultaneously. The highest bidder wins the object and pays her own bid to the seller. There is no entry cost and reservation price. The auction is within the independent private values (IPV) paradigm and all of the bidders play Bayesian Nash equilibrium (BNE). Suppose there are n bidders. Each bidder $i \in \mathcal{I} \equiv \{1, 2, \cdots, n\}$ has a private value which is a random draw from an individual-specific value distribution $F_i(\cdot) \in \mathcal{F}$. Values are independent among different bidders. $\mathcal{F}$ is defined as follows.

Definition 1. *Let $\mathcal{F}$ be the set of value distributions $F(\cdot)$ s.t. (a) $F(\cdot)$ is a cumulative distribution function defined on the interval $[\underline{v}, \overline{v}]$. (b) $F(\cdot)$ has at least second order continuous derivatives. (c) The density $f(\cdot)$ is bounded away from zero and infinity on its support $[\underline{v}, \overline{v}]$.*

Each bidder i has a preference described by an individual-specific utility function $u_i(\cdot) \in \mathcal{U}$. The set of utility functions $\mathcal{U}$ is defined as follows.

Definition 2. *Let $\mathcal{U}$ be the set of utility functions $u(\cdot)$ s.t. (a) $u : [0, +\infty) \to [0, +\infty)$. (b) $u(0) = 0$. (c) u has at least second order continuous derivatives and $u'(\cdot) > 0$.*

Given the model setup above, the expected utility maximization problem for bidder i is given by

$$\max_{b \in \mathbb{R}_+} \Pr(b_j < b, j \neq i) \cdot u_i(v_i - b).$$

Denote $G_j(b)$ as bidder j's bid distribution, $j \in \mathcal{I} \setminus i$. Since the model is within the IPV paradigm, bidder i's winning probability can be written as the product of all other bidders' bid distributions.

$$\Pr(b_j < b, j \neq i) = \prod_{j \in \mathcal{I} \setminus i} G_j(b)$$

The utility maximization problem is then given by

$$\max_{b \in \mathbb{R}_+} \prod_{j \in \mathcal{I} \setminus i} G_j(b) \cdot u_i(v_i - b).$$

Denote $\tilde{G}_i(b) \equiv \prod_{j \in \mathcal{I} \setminus i} G_j(b)$. We further rewrite the above equation as

$$\max_{b \in \mathbb{R}_+} \tilde{G}_i(b) \cdot u_i(v_i - b).$$

The first order condition gives

$$v_i = s_i^{-1}(b) = b + \lambda_i^{-1}(\frac{\tilde{G}_i(b)}{\tilde{g}_i(b)}) \tag{4.1}$$

where $\lambda_i \equiv u_i(\cdot)/u_i'(\cdot)$ and $\tilde{g}_i(b) = d\tilde{G}_i(b)/db$. By Definition 2, $\lambda_i : [0, +\infty) \to [0, +\infty)$. λ_i^{-1} is the inverse function of λ_i and thus $\lambda_i^{-1} : [0, +\infty) \to [0, +\infty)$. Equation (4.1) builds up the relationship between values and bids, given other bidders strategies. It characterizes the best response function for bidder i. The equation system that consists of all bidders

best response functions characterizes the BNE, i.e.,

$$\begin{cases} v_1 = s_1^{-1}(b_1) = b_1 + \lambda_1^{-1}(\frac{\tilde{G}_1(b_1)}{\tilde{g}_1(b_1)}) \\ v_2 = s_2^{-1}(b_2) = b_2 + \lambda_2^{-1}(\frac{\tilde{G}_2(b_2)}{\tilde{g}_2(b_2)}) \\ \cdots \\ v_i = s_i^{-1}(b_i) = b_i + \lambda_i^{-1}(\frac{\tilde{G}_i(b_i)}{\tilde{g}_i(b_i)}) \\ \cdots \\ v_n = s_n^{-1}(b_n) = b_n + \lambda_n^{-1}(\frac{\tilde{G}_n(b_n)}{\tilde{g}_n(b_n)}) \end{cases}$$

4.1.2 First-Price Auctions with Unobserved Heterogeneity

The baseline model involves individual-specific utility functions and value distributions, i.e., two-dimensional unobserved heterogeneities of bidders. This is general but leads to unnecessary complexity. We assume that the heterogeneity is discrete and finite, and we call it "type". Suppose there are $K \geq 2$ types[1] and each bidder belongs to type $k \in \mathcal{K} = \{1, 2, \cdots, K\}$ with type probability $p(k)$ with $\sum_{k=1}^{K} p(k) = 1$. The number of types K is unknown to the researchers. Each type k has a type-specific utility function $u_k(\pi)$ and a type-specific value distribution $F_k(v)$. Each bidder knows her own type but does not know any other bidder's type. Each bidder learn the others' types according to a set of type probabilities, $\{p_k\}_{k=1}^{K}$, i.e., a bidder with unknown type belongs to type k with probability p_k. We also assume bidders' types are independent with each other. The common knowledge for all bidders are type probability p_k, type-specific utility function u_k and type-specific value distribution F_k, where $k = 1, \cdots, K$.

Assume bidders within the same type play symmetric strategy, then the bid distributions for the bidders can be categorized according to type. We call them type-specific bid distributions, denoted as $G_k(b)$. We can rewrite Equation (4.1) in terms of types, given

[1] If $K = 1$, the bidders are symmetric.

bidder i belongs to type k

$$v_i = s_i^{-1}(b) = b + \lambda_k^{-1}(\frac{\tilde{G}_i(b)}{\tilde{g}_i(b)}) \tag{4.2}$$

where $\lambda_k = u_k/u_k'$ and λ_k^{-1} is the inverse function of λ_k. We then show that $\tilde{G}_i(b)$ is identical across bidder $i = 1, 2, \cdots, n$. For a specific bidder i, she perceives the bid distribution of bidder $j \neq i$ as an expectation over the unknown types.

$$G_j(b) = \sum_{k \in \mathcal{K}} p_k G_k(b)$$

Since the types among bidders are independent, we have that

$$\tilde{G}_i(b) = \prod_{j \in \mathcal{I} \setminus i} G_j(b) = \left(\sum_{k \in \mathcal{K}} p_k G_k(b) \right)^{n-1}$$

Note that $\tilde{G}_i(b)$ is invariant across i. Denote $\tilde{G}(b) \equiv \left(\sum_{k \in \mathcal{K}} p_k G_k(b)\right)^{n-1}$. We have that

$$\tilde{G}_i(b) = \tilde{G}(b), \quad i = 1, 2, \cdots, n$$

Given the above property, we rewrite Equation (4.2) as

$$v_i = s_i^{-1}(b) = b + \lambda_k^{-1}(\frac{\tilde{G}(b)}{\tilde{g}(b)})$$

Note that the right-hand side of the above equation depends on only type k, instead of individual i. We change the subscript of $s_i^{-1}(b)$ to k, i.e., $s_k^{-1}(b)$. The equation above becomes

$$v_i = s_k^{-1}(b) = b + \lambda_k^{-1}(\frac{\tilde{G}(b)}{\tilde{g}(b)}) \tag{4.3}$$

Equation (4.3) characterizes the best response functions for bidders within type k.

For the purpose of tractability, we restrict our analysis to the cases where bidders' asymmetry is due to one-dimensional unobserved heterogeneity. We introduce two specific first-price auction models with types: asymmetric preference (AP) model and asymmetric value distribution (AVD) model. In AP model, value distributions are identical across all types, i.e., $F_k(v) = F(v)$, while preferences remain type-specific. Denote the strategy for type k as $s_{k,AP}(v)$ and the inverse strategy as $s^{-1}_{k,AP}(b)$. From Equation (4.3), we have the best response function for AP model

$$v_i = s^{-1}_{k,AP}(b) = b + \lambda_k^{-1}(\frac{\tilde{G}(b)}{\tilde{g}(b)}) \tag{4.4}$$

Since $s^{-1}_{k,AP}(b)$ is increasing in b, the type-specific bid distribution for type k is then give by

$$G_k(b) = F(s^{-1}_{k,AP}(b))$$

In AVD model, bidders' preferences are identical, and we further assume that the preference is risk-neutral with utility function $u_k(x) = u(x) = x$. Then,

$$\lambda_k(x) = u_k(x)/u_k'(x) = x$$

and

$$\lambda_k^{-1}(x) = x$$

for all $k \in \mathcal{K}$. Denote the strategy for type k as $s_{k,AVD}(v)$ and the inverse strategy as $s^{-1}_{k,AVD}(b)$. From Equation (4.3), we have the best response function for AVD model

$$v_i = s^{-1}_{k,AVD}(b) = b + \frac{\tilde{G}(b)}{\tilde{g}(b)} \tag{4.5}$$

The value distributions remain type-specific. Since $s_{k,AVD}^{-1}(b)$ is increasing in b, the type-specific bid distribution for type k is then give by

$$G_k(b) = F_k(s_{k,AVD}^{-1}(b))$$

Lebrun (1999) proves the existence and uniqueness of a Bayesian Nash equilibrium (BNE) if all the K distributions have a mass point at $\underline{v}$. At this equilibrium, any two bidders who draw their values from the same distribution have the same bidding strategy. Moreover, if there exists a relation of stochastic dominance between two valuation distributions, the same relation of stochastic dominance extends to the bid distributions, i.e., $F_j(v) \leq F_i(v)$ for all $v \in [\underline{v}, \overline{v}]$ implies $F_j(s_j^{-1}(b)) \leq F_i(s_i^{-1}(b))$ for all $b \in [\underline{v}, \eta]$ with $\eta = s_1(\overline{v}) = ... = s_K(\overline{v})$.

Assumption 1. *There exists a relation of stochastic dominance among the K value distributions $F_1(\cdot)$, ..., $F_K(\cdot)$, i.e., $F_1(\cdot) \leq ... \leq F_K(\cdot)$. Moreover, the K distributions share the common support $[\underline{v}, \overline{v}]$ and have a mass point as $\underline{v}$.*

4.1.3 Model Implications

For notation simplicity, denote the α-th quantile of type k's bid distribution as $b_{k,\alpha} \equiv G_k^{-1}(\alpha)$ and the α-th quantile of type k's value distribution as $v_{k,\alpha} \equiv F_k^{-1}(\alpha)$. Define $T(\alpha) \equiv b_{k,\alpha} - b_{l,\alpha}$, which measures the distance between $G_k(\cdot)$ and $G_l(\cdot)$ horizontally, i.e., the distance between inverse $G_k(\cdot)$ and inverse $G_l(\cdot)$. Denote $H(b) \equiv \tilde{G}(b)/\tilde{g}(b)$.

Assumption 2. *In AP model, there exists a pair of types $k < l \in \mathcal{K}$ s.t. both type k and l have CRRA utility functions, and type k is more risk averse than type l.*

Assumption 3. *In AVD model, there exist a pair of types $k < l \in \mathcal{K}$ and a quantile interval $Q \subset [0, 1]$ s.t. F_k first-order stochastic dominates F_l and $v_{k,\alpha} - v_{l,\alpha}$ is non-increasing on Q.*

Assumption 2 and Assumption 3 are mild. CRRA utility functions are widely used when studying first-price auction models. Assumption 3 places restrictions only within a limited quantile interval.

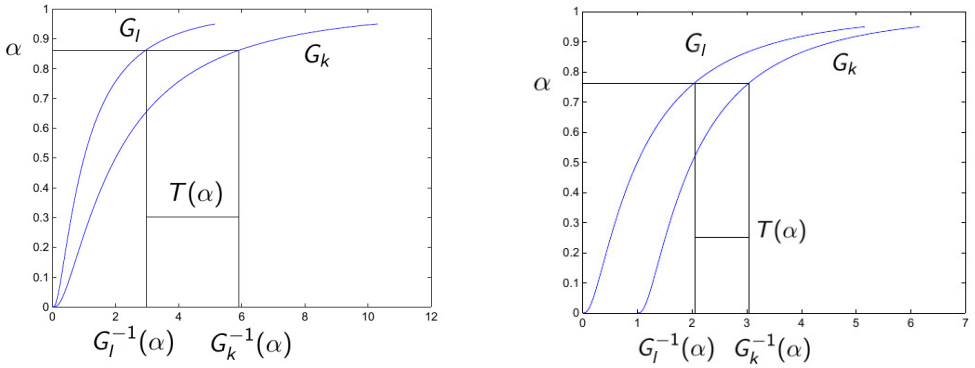
Proposition 1. *(a) Under Assumption 2, if* $0 \leq \frac{\partial H(b_{k,\alpha})}{\partial b_{k,\alpha}} \leq \frac{\partial H(b_{l,\alpha})}{\partial b_{l,\alpha}}$*, then* $T(\alpha)$ *is increasing; (b) under Assumption 3, if* $\frac{\partial H(b_{k,\alpha})}{\partial b_{k,\alpha}} \geq \frac{\partial H(b_{l,\alpha})}{\partial b_{l,\alpha}} \geq 0$*, and* $\alpha \in Q$*, then* $T(\alpha)$ *is decreasing in* α*.*

Proof. See Appendix. □

Since we are able to estimate $H(b)$, the conditions in Proposition 1(a) and (b) are testable. We then visually illustrate Proposition 1 in Figure 4.1.

Panel (a) of Figure 4.1 illustrates the behavior of the function $T(\alpha)$ in part (a) in Proposition 1. For bidders with values at very high quantiles, winning probabilities are high. They are facing a dilemma of sacrificing winning probability and sacrificing winning profit. Bidders with higher levels of risk-aversion tend to choose sacrificing winning profit by bidding high, while those with lower levels of risk-aversion will try to balance the sacrifice between winning probability and winning profit. This leads to a large value of $T(\alpha)$ when α is large. For bidders with values at very low quantiles, winning probability is low and close to zero. They have to bid very close to their own values to obtain an acceptable winning probabilities. Since all bidders' values are i.i.d. draws from an identical value distribution, we should not observe much difference in behavior among different types of bidders at very low quantiles. This explains the small value of $T(\alpha)$ when α is small.

Panel (b) of Figure 4.1 illustrates the behavior of the function $T(\alpha)$ in Proposition 1. For bidders with values at very low quantiles, just like the case in AP model, they will bid very close to their own values. The difference between bid distributions will approximate the difference between value distributions. For bidders with values at very high quantiles,



(a) Illustration of $T(\alpha)$ in AP Model

(b) Illustration of $T(\alpha)$ in AVD Model

Figure 4.1: Illustration of function $T(\alpha)$

they have very high winning probabilities, and thus will tend to bid significantly lower than their own values. To be specific, the strong type will go much further than the weak type does. This is because the strong type is at really high quantiles while the weak type is at sub-high quantiles, if we consider the mixture model. As a result, when α is high, $T(\alpha)$ will be a lot smaller than the distance between value distributions. As long as the distance between value distributions does not increasing in α, $T(\alpha)$ is decreasing in α.

Proposition 1 shows distinct implications on $T(\alpha)$ (increasing and decreasing) from AP and AVD model. In the next, we show that these model implications are testable, and provide corresponding test procedures.

4.2 Identification, Estimation and Testing

The model implications of both AP and AVD model are placed on the function $T(\alpha)$. In order to test the model implications, we first discuss the approach to obtain $T(\alpha)$, i.e., the identification and estimation of the type-specific bid distributions $G_k(b)$.

4.2.1 A Unified Econometric Model

In both AP and AVD model, we observe a set of bids from all bidders without knowing the types. Denote the unconditional bid distribution as $G(b)$, which can be recovered from the data. The unconditional bid distribution is a mixture of the type-specific bid distributions. The type probabilities are the weights in this mixture model.

$$\begin{aligned} G(b) &= \sum_{k\in\mathcal{K}} G_k(b)p_k \\ &= \begin{cases} \sum_{k\in\mathcal{K}} F\big(s^{-1}_{k,AP}(b)\big)p_k, & \text{AP model} \\ \sum_{k\in\mathcal{K}} F_k\big(s^{-1}_{k,AVD}(b)\big)p_k, & \text{AVD model} \end{cases} \end{aligned} \tag{4.6}$$

Equation (4.6) shows that both AP and AVD model lead to an identical form of mixture model. A unified approach can be used for the identification. We want to identify the number of types K, type-specific bid distributions $G_k(b)$ and type probabilities p_k, i.e., to identify all of the elements in the mixture model.

4.2.2 The Number of Types

One key point in Equation (4.6) is that it contains the information of the number of types K. Under our assumption of discrete and finite types, the test of $K = 1$ against $K > 1$ will be equivalent to the test of the existence of heterogeneity, which is our first essential goal. We follow the method proposed by Hu (2008), which was first applied to the measurement-error models. The idea of this method is to obtain three different measures in terms of bids to identify the latent variables, which in our case are bidders' types and bid distributions. As a result, some restrictions must be placed on the data structure.

Assumption 4. *(1) Each bidder participates in at least three homogeneous auctions and her values in the three auctions are independent. (2) The type is invariant across auctions.*

By Assumption 4, we obtain three bids for each bidder, denoted as a bid vector

$$(b_1\ b_2\ b_3)^T.$$

To identify the discrete types, it is convenient to transform two of the measures into discrete form. Without loss of generality, we discretize b_1 and b_3 into d_1 and d_3. The discretization is as follows. Divide the support of b_i into M bins, $i = 1, 3$. Label the bins from left to right as $bin_1, bin_2, \cdots, bin_M$.

$$d_i \quad = \quad \begin{cases} 1, & b_i \in bin_1 \\ 2, & b_i \in bin_2 \\ \cdots & \\ M, & b_i \in bin_M \end{cases}$$

d_i takes values from 1 to M, according to which bin b_i falls into. The choice of M will be discussed later. Another assumption we need for identification is to ensure there is sufficient variation in bidding behavior among different types, i.e., there is no type-specific bid distribution that is a linear combination of the other type-specific bid distributions.

Assumption 5. *There does not exists a non-zero vector of constants $\{c_k\}_{k\in\mathcal{K}}$ such that*

$$\begin{cases} \sum_{k\in\mathcal{K}} c_k F(s^{-1}_{k,AP}(b)) = 0, & \textit{AP model} \\ \sum_{k\in\mathcal{K}} c_k F_k(s^{-1}_{k,AVD}(b)) = 0, & \textit{AVD model} \end{cases} \tag{4.7}$$

Note that Equation (4.7) is equivalent to $\sum_{k\in\mathcal{K}} c_k G_k(b) = 0$. Denote the joint probability mass function (pmf) of d_1 and d_3 as $g(d_1, d_3)$. A matrix form of the joint pmf is expressed as

$$B_{d_1,d_3} \equiv [g(d_1 = i, d_3 = j)]_{i,j}$$

where B_{d_1,d_3} is an $M \times M$ matrix. This matrix is the key to identify the number of types K.

Proposition 2. *Under Assumption 5,* $Rank(B_{d_1,d_3}) = K$ *if* $M \geq K$.

Proof. See proof of Lemma 2 in An (2013) . □

According to Proposition 2, as long as the dimension of B_{d_1,d_3} is greater than K, its rank equals K. If the dimension of B_{d_1,d_3} is less than K, it has full rank (less than K). The procedure to determine K is as follows.

(1) Start from $M = 2$ and test if B_{d_1,d_3} has full rank. If B_{d_1,d_3} has full rank, it implies that $M \leq K$.

(2) Increase M by 1 and test if B_{d_1,d_3} has full rank.

(3) Repeat step (1) and step (2) until B_{d_1,d_3} has *not* full rank.

The final value of M equals $K+1$. The remaining issue is the rank test. We follow the method proposed by Robin and Smith (2000). The idea is to test the number eigenvalues in a matrix that are significantly different from zero. Given a matrix with dimension $M \times M$, our objective is to determine its rank. The null hypothesis H_0 is that $rank(B_{d_1,d_3}) = r^*$ and the alternative hypothesis H_1 is that $rank(B_{d_1,d_3}) > r^*$, where r^* is chosen by researchers and $0 \leq r^* \leq M-1$[2]. This is a sequential test, in which r^* starts from 0 and stops at the true rank. The details of the testing procedure are summarized in Appendix B.1.

4.2.3 The Type-Specific Bid Distribution

Once the number of types is identified, we further identify the bid distributions conditional on types and the corresponding type probabilities. Denote $g(b_1, b_2, b_3)$ as the joint density of the three bids. By law of total probability, we have

$$g(b_1, b_2, b_3) = \sum\nolimits_{k \in \mathcal{K}} g(b_1, b_2, b_3, k) = \sum\nolimits_{k \in \mathcal{K}} g(b_1|b_2, b_3, k) g(b_2|b_3, k) g(k, b_3)$$

[2]If $r^* = M$, the alternative hypothesis can never be true.

By Assumption 4, bids are independent conditional on types. As a result, $g(b_1|b_2, b_3, k) = g(b_1|k)$ and $g(b_2|b_3, k) = g(b_2|k)$. This allows us to rewrite the equation above as

$$g(b_1, b_2, b_3) = \sum\nolimits_{k\in\mathcal{K}} g(b_1|k)g(b_2|k)g(k, b_3)$$

Replacing b_1 and b_3 in the above equation respectively with d_1 and d_3 gives

$$g(d_1, b_2, d_3) = \sum\nolimits_{k\in\mathcal{K}} g(d_1|k)g(b_2|k)g(k, d_3) \tag{4.8}$$

Since type k is discrete, so are d_1 and d_3, it is convenient to express the above equation in matrix form

$$B_{b_2,d_1,d_3} = B_{d_1|k} D_{b_2|k} B_{k,d_3}, \tag{4.9}$$

where the matrices are defined as

$$\begin{aligned}
B_{b_2,d_1,d_3} &\equiv [g(b_2, d_1 = i, d_3 = j)]_{i,j} \\
B_{d_1|k} &\equiv [g(d_1 = i|k = l)]_{i,l} \\
D_{b_2|k} &\equiv \text{diag}\,[g(b_2|k = 1), g(b_2|k = 2), \cdots, g(b_2|k = K)] \\
B_{k,d_3} &\equiv [g(k = l, d_3 = j)]_{l,j}
\end{aligned}$$

Proposition 2 enables us to choose $M = K$, then the matrices B_{d_1,d_3}, $B_{d_1|k}$, and B_{k,d_3} are all full-rank, and thus invertible. Note that $B^{-1}_{d_1,d_3} = B^{-1}_{k,d_3} B^{-1}_{d_1|k}$ holds. If we post-multiply $B^{-1}_{d_1,d_3}$ to both sides of (4.9) and we will have

$$B_{b_2,d_1,d_3} B^{-1}_{d_1,d_3} = B_{d_1|k} D_{b_2|k} B^{-1}_{d_1|k}. \tag{4.10}$$

The right-hand side of the equation above is a form of eigenvalue-eigenvector decomposi-

tion of the left-hand side. The left-hand side consists of the probability density of b_2 and the probability mass of d_1 and d_3, which are observable. Thus, we identify $B_{d_1|k}$ through eigenvalue-eigenvector decomposition. Note that each column of $B_{d_1|k}$ should sum up to 1. We normalize $B_{d_1|k}$ by setting its 1-norm as 1, instead of setting its 2-norm as 1.

$B_{d_1|k}$ is the key to identify type-specific bid distributions and type probabilities. It is a bridge between the unobserved type k and the observed discretized bid d_1. The identification equation for type probabilities is given by

$$p(d_1) = B_{d_1|k} p_k \tag{4.11}$$

where $p(d_1) = (p(d_1 = 1), p(d_1 = 2), \cdots, p(d_1 = M))'$. Since $B_{d_1|k}$ is invertible, Equation (4.11) can be rewritten as $p_k = B_{d_1|k}^{-1} p(d_1)$. The probability $p(d_1)$ can be estimated by sample analog. The equation to identify type-specific bid distributions is given by

$$\begin{cases} G(b_2|k=l) = G(b_2, k=l)/p(k=l) \\ g(b_2|k=l) = g(b_2, k=l)/p(k=l) \end{cases}, l = 1, 2, \cdots, M \tag{4.12}$$

where $G(b_2, k = l)$ can be estimated by empirical CDF estimator and $g(b_2, k = l)$ can be estimated by kernel method.[3]

The identified type-specific bid distributions are anonymous. They can be labeled according to the characteristics of both AP and AVD model. We thus need assumptions to ensure these characteristics are clear.

Assumption 6. *In AP model, different types have different levels of risk-aversion.*

The risk-aversion in Assumption 6 is measured by $R = -u''/u'$.

[3]For more details regarding estimation, see An (2013).

Assumption 7. *In AVD model, there exists first-order stochastic domination ($FOSD$) relationship between any two value distributions.*

Proposition 3. *Under Assumption 6 and 7, in either AP or AVD model, there exists $FOSD$ relationship between any two different type-specific bid distributions.*

Proof. See Lebrun (1999) proof of Corollary 4. □

By Proposition 3, we can thus label the identified types according to the $FOSD$ relationship. Denote type 1 as the strongest type, type 2 as the second strongest type and so on, i.e., $G_1(b)\ FOSD\ G_2(b)$, $G_2(b)\ FOSD\ G_3(b)$, $\cdots$, $G_{K-1}(b)\ FOSD\ G_K(b)$. Given the identified type-specific bid distributions, we further identify the function $T(\alpha) = G_k^{-1}(\alpha) - G_l^{-1}(\alpha)$.

Proposition 4. *The type probability estimator $\hat{p}(k)$ and type-specific bid distribution estimator $\hat{G}_k(b)$ are both $\sqrt{n}$-consistent. The estimator for function $T(\alpha)$ is $\sqrt{n}$-consistent.*

Proof. See Appendix. □

4.2.4 Testing Model Implications

Proposition 1 states that the function $T(\alpha)$ is either increasing under AP model, or decreasing under AVD model. This section introduces the corresponding test to investigate the monotonicity of the function T. Let $\hat{T}(\alpha)$ be the estimator for $T(\alpha)$, and it is a $\sqrt{n}$ consistent estimator. Since testing increasing and testing decreasing are equivalent, we focus on testing increasing, and use “increasing" and “monotone" interchangeably.

The monotonicity condition is imposed on grids, i.e.,

$$T(\alpha_i) < T(\alpha_j), \ \ \forall \alpha_i < \alpha_j$$

where $i, j \in \{1, 2, \cdots, m\}$. To accommodate the formation of the monotonicity condition, the function $T(\alpha)$ refers to a vector $(T(\alpha_1) \;\; T(\alpha_2) \cdots T(\alpha_m))^T$, i.e., a vector consists of the function values on the grids. We then set up our test under the context of functional analysis. Denote Λ as the set of increasing functions. The null hypothesis H_0 of our monotonicity test is that $T(\alpha) \in \Lambda$, i.e. $T(\alpha)$ is increasing. The alternative hypothesis H_1 is that $T(\alpha) \notin \Lambda$, i.e., $T(\alpha)$ is not increasing. The main idea of this test is to determine if the estimated function $\hat{T}(\alpha)$ is far enough from the set Λ. If the distance is too large, we tend to believe that the true function $T(\alpha) \notin \Lambda$, and reject the null hypothesis. Otherwise we cannot reject the null hypothesis. To determine the distance between the function $\hat{T}(\alpha)$ and the set Λ, we need to obtain the projection of $\hat{T}(\alpha)$ on Λ. Denote the projection as $\tilde{T}(\alpha)$ and denote Π_Λ as the projection operator on set Λ. Then $\tilde{T}(\alpha) = \Pi_\Lambda \hat{T}(\alpha)$. The distance between function $\hat{T}(\alpha)$ and set Λ is equivalent to the distance between the two functions, $\hat{T}(\alpha)$ and $\tilde{T}(\alpha)$. We use Euclidean norm to define the distance, i.e.,

$$||\hat{T}(\alpha) - \tilde{T}(\alpha)|| = \sqrt{(\hat{T}(\alpha_1) - \tilde{T}(\alpha_1))^2 + (\hat{T}(\alpha_2) - \tilde{T}(\alpha_2))^2 + \cdots + (\hat{T}(\alpha_m) - \tilde{T}(\alpha_m))^2}$$

The projection $\tilde{T}(\alpha)$ is obtained by the following optimization.

$$\begin{aligned} \tilde{T}(\alpha) = \quad & \underset{\tilde{T}(\alpha)}{\text{argmin}} \quad ||\hat{T}(\alpha) - \tilde{T}(\alpha)|| \\ & \text{s.t.} \qquad \tilde{T}(\alpha) \in \Lambda \end{aligned}$$

Recall that $\tilde{T}(\alpha)$ refers to a vector $(\tilde{T}(\alpha_1) \;\; \tilde{T}(\alpha_2) \cdots \tilde{T}(\alpha_m))^T$. The above optimization is in fact choosing $\{\tilde{T}(\alpha_i)\}_{i=1}^m$, i.e., m parameters, to minimize the objective function. Note that we can rewrite this optimization problem by applying a monotone transformation on

the objective function, which is given by

$$\begin{aligned} \tilde{T}(\alpha) = \quad & \underset{\tilde{T}(\alpha)}{\text{argmin}} \quad ||\hat{T}(\alpha) - \tilde{T}(\alpha)||^2 \\ & \text{s.t.} \qquad \tilde{T}(\alpha) \in \Lambda \end{aligned}$$

The new objective function is quadratic in the parameters, and the conditions are linear in the parameters ($\tilde{T}(\alpha_1) < \tilde{T}(\alpha_2) < \cdots < \tilde{T}(\alpha_m)$), which enables us to apply the algorithm of quadratic programming to solve the optimization problem. The quadratic programming algorithm is a highly efficient algorithm. The time complexity of quadratic programming has an order of $O(m \ln m)$. There is barely a computational burden if m takes a value of several hundreds or thousands.

After obtaining the projection $\tilde{T}(\alpha)$, we are able to calculate the distance between $\hat{T}(\alpha)$ and Λ. Denote the distance as $\phi(\hat{T}(\alpha)) \equiv ||\Pi_\Lambda \hat{T}(\alpha) - \hat{T}(\alpha)||$. The next step is to determine if $\phi(\hat{T}(\alpha))$ is large enough so that we can reject the null hypothesis, i.e., to obtain the critical values. Since the projection operator is not Hadamard differentiable, the asymptotic distribution of $\sqrt{n}(\phi(\hat{T}(\alpha)) - \phi(T(\alpha)))$ cannot be consistently estimated by bootstrapping on $\phi(T(\alpha))$. Fang and Santos (2014) provide a consistent bootstrap method based on the delta method. We have that

$$\sqrt{n}(\phi(\hat{T}(\alpha)) - \phi(T(\alpha))) \approx \phi'_{T(\alpha)}(\sqrt{n}(\hat{T}(\alpha) - T(\alpha)))$$

The asymptotic distribution of the derivative $\phi'_{T(\alpha)}(\sqrt{n}(\hat{T}(\alpha) - T(\alpha)))$, however, can be consistently estimated by bootstrapping on $T(\alpha)$. The bootstrap procedure is as follows:

(1) Generate a bootstrap sample by pairwise resampling[4] the bid vector $(b_1 \ b_2 \ b_3)^T$. Denote the bootstrap sample as $(b_1^* \ b_2^* \ b_3^*)^T$.

[4]We allow for other commonly used resampling methods. We can also allow the bootstrap sample size to be different from the observed sample size.

(2) Estimate the function $T(\alpha)$ from the bootstrap sample. Denote the estimates as $\hat{T}^*(\alpha)$. Obtain $\phi'_n(h^*)$ by

$$\phi'_n(h^*) \equiv \frac{\phi(\hat{T}(\alpha) + t_n h^*) - \phi(\hat{T}(\alpha))}{t_n}$$

where $h^* = \sqrt{n}(\hat{T}^*(\alpha) - \hat{T}(\alpha))$. Hong and Li (2014) suggest that t_n goes to 0 at the rate of $n^{-\frac{1}{3}}$.

(3) Repeat step (1) and step (2) for B times, and collect all B values of $\phi'_n(h^*)$. Sort the bootstrap test statistics ascending. Since this is a one-sided test, the upper 5% quantile of the bootstrap test statistics is the critical value for 5% significance level.

4.3 Monte Carlo Studies

In this section, we demonstrate the performance of our method by Monte Carlo simulation. We consider two data generating processes (DGP) from AP and AVD model respectively. For both DGPs, there are two types of bidders, i.e., $K = 2$. We consider three type probabilities, $p_1 = 0.4$, 0.5, 0.6, and $p_2 = 1 - p_1$. The number of qualified bidders (with 3 independent bids) N takes values of 400, 800 and 1200. The bids are collected from homogeneous auctions. Each auction contains 5 bidders. The number of replications is 1000. For the DGP from AP model, type 1 has a CRRA utility function and type 2 has a risk neutral utility function, i.e., $u_1 = \frac{x^{0.5}}{0.5}$ and $u_2 = x$. The common value distribution is log-normal distribution, i.e., $F = e^Z$, where $Z \sim N(0, 1)$. For the DGP from AVD model, every bidder is risk-neutral, i.e., $u(x) = x$. The value distribution of type 2 is log-normal distribution, i.e., $F_2 = e^Z$, where $Z \sim N(0, 1)$. Type 1 has a shifted log-normal value distribution, i.e., $F_1(v) = F_2(v - 0.5)$.

To simulate bids from each auction, we first determine the type of each bidder and let each bidder attend three independent and homogeneous auctions. The private values v for each bidder in each auction are randomly drawn from their own value distributions. We

then calculate each bidder's bids according to the BNE. The equilibrium for both AP and AVD model are characterized by Equation (4.4) and Equation (4.5) respectively. Due to the complexity of the differential equation, we cannot obtain a closed-form solution. Thus, we numerically solve for the equilibrium. The main idea is to do iterations over the system of the best response functions. The initial strategy for bidder i is to bid her/his own value, i.e., $s_i^0(v) = v$. The iteration is as follows.

(1) Generate a large sample of homogeneous auctions with bidders' strategies being $s_i^0(v)$. Estimate the bid distributions for each type.

(2) Let each bidder reacts to the estimated type-specific bid distributions in step (1). Plug the type-specific bid distributions and type probabilities into the system of the best response functions. Update the mapping from value v to bid b for each bidder. Denote this new mapping as a new strategy $s_i^1(v)$.

(3) If $s_i^1(v)$ and $s_i^0(v)$ are not close enough (depends on the numerical convergence condition), override $s_i^1(v)$ with $s_i^0(v)$ and repeat step (1) and step (2). If $s_i^1(v)$ and $s_i^0(v)$ converges, $s_i^1(v)$ is the strategy in equilibrium.

Note that this iteration is a contract mapping process. The convergence in step (3) is thus guaranteed. In the next subsection, we examine the performance of our method bases on the simulated data from both AP and AVD model.

4.3.1 The DGP from AP Model

In this subsection, we discuss the performance of our method for the DGP from AP model. For simplicity, we call this DGP as AP DGP. The first step of our method is to determine the number of types. The type probability we consider here is $p_1 = p_2 = 0.5$.[5] Table 4.1 shows the results of the rank test on the matrix B_{d_1,d_3}. The null hypothesis H_0 is that $rank(B_{d_1,d_3}) = r^*$ and the alternative hypothesis H_1 is that $rank(B_{d_1,d_3}) > r^*$.

[5]The results for $p_1 = 0.4$ and $p_1 = 0.6$ are quite similar to the results for $p_1 = 0.5$.

The significance level is 5%. The number of bins M is 3, and thus B_{d_1,d_3} is a 3×3 matrix. Recall that the true rank of B_{d_1,d_3} equals the number of types, which is 2. The first row of Table 4.1 shows the test results when $r^* = 1$, in which the null hypothesis is false. The results assess the test power. We see that even in the case with the smallest sample size, $N = 400$, the rejection rate is 50.8%. When sample size increases, the rejection rate approaches 100%. The test power is satisfactory. The second row of Table 4.1 shows the test results when $r^* = 2$, in which the null hypothesis is true. The results assess the test size. We see that the rejection rates are very close to the significance level in all three cases with different sample sizes. The test size is normal.

Table 4.1: Rejection Rates of Rank Test with AP DGP

	$N = 400$	$N = 800$	$N = 1200$
$r* = 1$	0.322	0.561	0.818
$r* = 2$	0.034	0.043	0.045

After the rank test, we are able to determine that the number of types K is 2. Given the number of types, we then estimate the type probabilities, i.e., p_1 and p_2, as well as the type-specific bid distributions, i.e., $G_1(b)$ and $G_2(b)$. Table 4.2 shows the estimated type probabilities in all 9 cases, where $p_1 = 0.4$, 0.5 and 0.6, and $N = 400$, 800 and 1200. For each case, we report the mean and the standard errors (in parenthesis) of the estimates in 1000 replications. We see that as the sample size increases, the means are getting closer and closer to the true value, and the standard errors decrease substantially. This pattern suggests that our estimator behaves normally. Figure 4.2 shows the estimated type-specific bid distributions when sample size is 400 and $p_1 = 0.5$. The two solid curves are the true type-specific bid distributions. There are two sequences of dots representing the means of our estimates. They are very close to the true value. The sequence of crosses and sequence

of squares represent the 90% confidence intervals for type 1's bid-distribution estimator and type 2's bid-distribution estimator respectively. Figure 4.3 and 4.4 show the cases when sample size is 800 and 1200 respectively. The confidence intervals shrink quickly when sample size increases. Most of the time, the two confidence intervals do not overlap, indicating that our separation of types is solid.

Table 4.2: Estimated Type Probabilities with AP DGP

p_1	$N = 400$	$N = 800$	$N = 1200$
0.4	0.468	0.440	0.421
	(0.411)	(0.301)	(0.250)
0.5	0.496	0.506	0.491
	(0.358)	(0.285)	(0.236)
0.6	0.551	0.571	0.580
	(0.352)	(0.273)	(0.228)

After the estimation of type-specific bid distributions, we use the estimates to test the model implications for both AP and AVD model. For simplicity, we call the two tests as AP test and AVD test. Recall that the true model is AP model, and thus the null hypothesis for AP test is true and the null hypothesis for AVD test is false. The significance level is 5%. The type probability we consider here is $p_1 = p_2 = 0.5$. The first row of Table 4.3 shows the rejection rates of AP test, whose null hypothesis is true. The rejection rates approximate the test size. For all three sample sizes, the rejection rates are close to the significance level. This indicates a normal test size for AP test. The first row of Table 4.4 shows the rejection rates of AVD test, whose null hypothesis is false. The rejection rates approximate the test power. When sample size is 400, the rejection rate is around 50%. As the sample size increases, the rejection rates increase at a satisfactory speed.

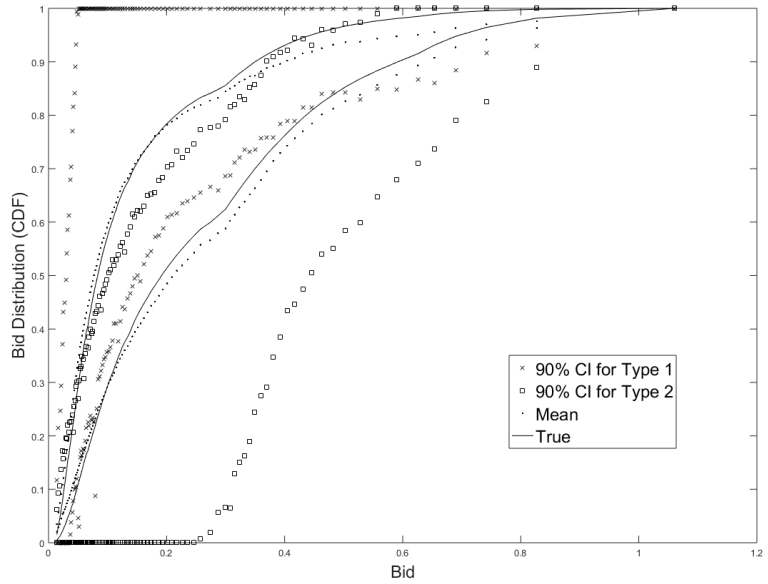


Figure 4.2: Type-Specific Bid Distributions $N = 400$

Table 4.3: AP Test: Size and Power

True Model	$N = 400$	$N = 800$	$N = 1200$
AP	0.066	0.052	0.052
AVD	0.698	0.878	0.948

Table 4.4: AVD Test Size and Power

True Model	$N = 400$	$N = 800$	$N = 1200$
AP	0.542	0.713	0.802
AVD	0.033	0.046	0.053

4.3.2 AVD Model DGP

In this subsection, we turn to discuss the results for AVD DGP. To avoid reporting redundant results, we omit the parts of type separation and directly dive into assessing the

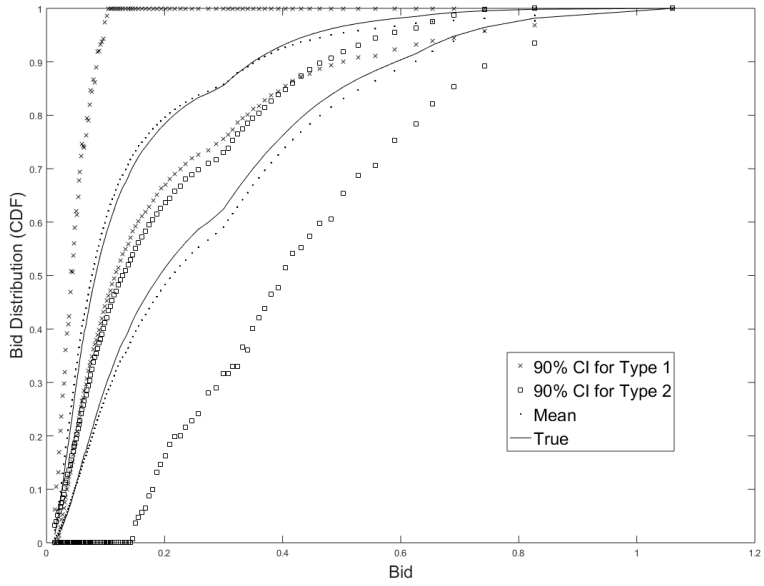


Figure 4.3: Type-Specific Bid Distributions $N = 800$

performance of AP test and AVD test. Since the true model is AVD model, we will be able to learn the test power of AP test and the test size of AVD test. The second row of Table 4.3 shows the rejection rates of AP model test. When sample size is 400, the rejection rate is about 70%. As sample size increases, the rejection rates increase quickly. On the other hand, the second row of Table 4.4 shows the rejection rates of AVD model test. The rejection rates are close to the significance level in all three cases. To sum up, the test power of AP test is satisfactory, and the test size of AVD test is normal.

4.3.3 Robustness Check: Correlated Values

This subsection aims to address a major concern regarding a fundamental assumption of our method, i.e., what if a bidder's private values in the three auctions are not independent and Assumption 4 fails. To evaluate this situation, we redo the whole simulation in the above with all the same settings except that bidders' private values are correlated

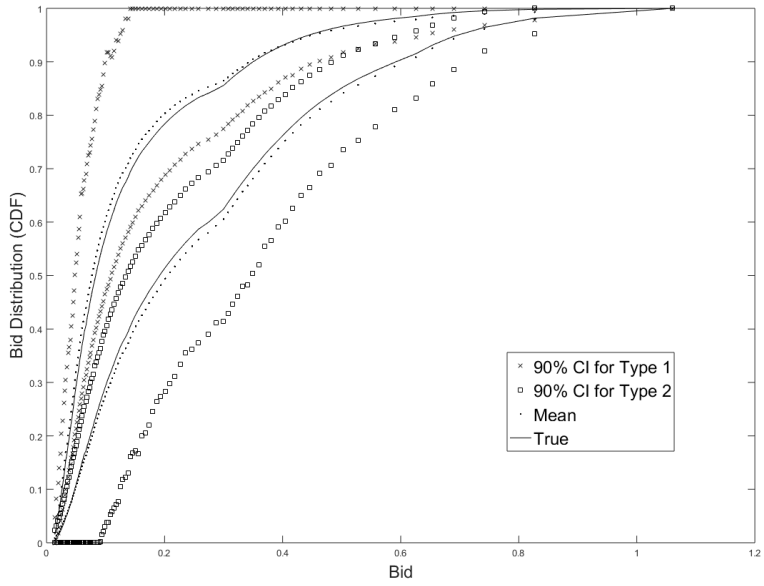


Figure 4.4: Type-Specific Bid Distributions $N = 1200$

across auctions. This correlation is on individual level. To simplify the analysis, we assume the correlation coefficients are identical across all bidders. For each bidder, denote the correlation coefficient between b_i and b_j as ρ_{ij}, $i, j = 1, 2, 3$. We further assume $\rho_{12} = \rho_{23} = \rho_{13} = \rho$. We examine 2 values of ρ, which are 0.1 and 0.2, and compare the results with the results from $\rho = 0$ (uncorrelated values). The type probability we consider here is $p_1 = p_2 = 0.5$.

First, we discuss the determination of the number of types with AP DGP. Table 4.5 shows the rejection rates of the test on $rank(B_{d_1,d_3}) = 2$. Although the number of types is 2, since Assumption 4 fails, $rank(B_{d_1,d_3})$ is no longer 2. We see that as the correlation coefficient ρ increases, it becomes more and more likely to reject that $rank(B_{d_1,d_3}) = 2$. However, as long as the correlation coefficient is not very large, the rejection rates are still acceptable.

Table 4.6 shows the rejection rates of AP test with both AP DGP and AVD DGP,

which approximate the test size and power of AP test respectively. On the other hand, Table 4.7 shows the rejection rates of AVD test with both AP DGP and AVD DGP, which approximate the test power and size of AVD test respectively. We see that both test size and power increase as the correlation coefficient increases, and as the sample size increases. As long as the correlation coefficient is small and the sample size is not too large, the test size distortion is acceptable.

Table 4.5: Rejection Rate of Rank Test under Correlated Data

ρ	$N = 400$	$N = 800$	$N = 1200$
0	0.034	0.043	0.045
0.1	0.048	0.056	0.052
0.2	0.050	0.056	0.061

Table 4.6: AP Test Size and Power under Correlated Data

ρ	$N = 400$	$N = 800$	$N = 1200$
AP Model DGP			
0	0.066	0.052	0.052
0.1	0.038	0.048	0.052
0.2	0.014	0.039	0.067
AVD Model DGP			
ρ	$N = 400$	$N = 800$	$N = 1200$
0	0.698	0.878	0.948
0.1	0.659	0.855	0.923
0.2	0.724	0.887	0.961

4.4 Empirical Studies

In this section, we apply the proposed estimation procedure and model-specification tests to USFS timber auction data. The test results are informative. We can reject AVD

Table 4.7: AVD Test Size and Power under Correlated Data

ρ	$N = 400$	$N = 800$	$N = 1200$
AP Model DGP			
0	0.542	0.713	0.802
0.1	0.821	0.966	0.997
0.2	0.897	0.986	0.998
AVD Model DGP			
ρ	$N = 400$	$N = 800$	$N = 1200$
0	0.033	0.046	0.053
0.1	0.028	0.076	0.072
0.2	0.021	0.090	0.084

model and cannot reject AP model. This is a support of AP model in explaining USFS timber auction data. Note that AP and AVD model are two very specific auction models with asymmetry. Our evidence does not rule out the possibility that the observed bidding behavior can also be explained by other alternative asymmetry of bidders, e.g., mixed asymmetric preferences and value distributions. However, the evidence can extend our understanding of the two basic but popular kinds of bidders' asymmetry.

4.4.1 The Data

The USFS timber auction data has been widely used in auction literature, including the analysis of bidders' asymmetry in either preferences or value distribuitions.[6] As for bidders' preferences, Baldwin (1995) provides empirical evidence which suggests that bidders can have decreasing absolute risk averse preferences. The risk aversion feature in bidders' preferences is also confirmed by Athey and Levin (2001). Campo et al. (2011) estimates the level of bidders' risk aversion by parametrizing their preferences with CRRA utility function. The risk aversion parameter in CRRA uttlity function is estimated to be around 0.30. As for bidders' value distributions, Athey et al. (2011) suggests that

[6]See Baldwin et al. (1997) for detailed description of USFS auctions.

bidders in USFS timber auction can be divided into two groups with high and low value distributions respectively.

The complete data set of timber auction contains both ascending price auctions and first-price auctions. Our analysis focuses on first-price auctions within IPV paradigm, so we rule out ascending price auctions. Among the remaining first-price auctions, we only consider “scaled sale" auctions, which are believed to be with private values. Haile et al. (2003) suggests that there is little empirical evidence supporting common values in “scaled sale" auctions. In addition, various studies assume private values in timber auctions, such as Baldwin et al. (1997), Haile (2001), and Haile and Tamer (2003). In “scaled sale" auctions, the auctioned object is timber to be harvest with unknown volume[7]. Bidders bid on per unit price instead of total price and the winner pays after the timber is harvested. The volume will be measured by a third party when harvesting. Among “scaled sale" auctions, we further eliminate salvage and small-business set-aside sales as they exhibit different features from original auctions. The reservation price is believed to be non-binding in various studies, such as Haile (2001), Baldwin et al. (1997) and Campo et al. (2011). Thus, we confidently apply our models with no settings of reservation price.

Table 4.8: “Scaled Sale" Auctions Summary Statistics

Variable	Mean	Std	Min	Max
Bids	52.45	45.91	0.13	534.58
Appraisal value	31.94	30.23	0.43	450.40
Volume	2808	2958	10	56350
Number of bidders	3.76	1.86	2	12

The time period of our data set is from 1982 to 1993. Auctions in all regions of the US

[7]The timber volume is measured in thousand board feet (mbf)

are considered. Each bidder's identity and bids, each tract's volume and appraisal value are available to us for each auction in the data set. Table 4.8 summarizes the statistics that are of our interest. The sample contains 5901 auctions. Bids and appraisal values are in 1982 real dollars. The auctions contain all kinds of timbers with various numbers of bidders and are clearly not homogeneous. To get rid of this heterogeneity, we first fix the number of bidders as 3 and rule out the other auctions. Among the remaining auctions, we identify 462 bidders who participates in at least 3 auctions. Then we control for appraisal values to capture the heterogeneity caused by different kinds of timbers as suggested by Campo et al. (2011) and Haile (2001). Similarly to the method used in Haile et al. (2003) and Bajari et al. (2007), we run a first-stage linear regression of bids on appraisal values, which is given by

$$b_{ij} = \beta_0 + X_i'\beta_1 + u_{ij}, \;\; u_{ij} \perp X_i$$

where b_{ij} is the j-th bid in i-th auction and X_i are the appraisal value in i-th auction. X_i can contain other auction characteristics if they are reasonable and available. The sum of the intercept and residual, i.e., $\beta_0 + u_{ij}$, is considered as "normalized" bid and will be used in the following estimation and testing procedures.

4.4.2 Estimation and Testing

Our first step is to determine the number of types. We consider four different ways to discretize b_1 and b_3. The number of bins M takes values from 2 to 5. Table 4.9 shows the results of rank test on the matrix B_{d_1,d_3}. The null hypothesis H_0 is that $rank(B_{d_1,d_3}) = r^*$ and the alternative hypothesis H_1 is that $rank(B_{d_1,d_3}) > r^*$. The first row shows the p-values when $r^* = 1$. We see there are straight rejections at 5% significance level across all discretization methods and hence we move to the test on $r^* = 2$. The second row shows the p-values when $r^* = 2$. In the case when $M = 2$, the test is invalid since B_{d_1,d_3} is a 2×2 matrix and has a rank of at most 2. In the cases when $M > 2$, we cannot reject

the null hypothesis at 5% significance level. As a result, we are confident to determine the number of types to be 2 and proceed to type-specific bid distribution estimation.

Table 4.9: Results of Rank Test with Timber Auction Data

P-Value	$M = 2$	$M = 3$	$M = 4$	$M = 5$
$r^* = 1$	0.000	0.000	0.001	0.022
$r^* = 2$	N/A	0.066	0.237	0.173

Figure 4.5 shows the estimated type-specific bid distributions. The three curves on the right-hand side are of high type, labeled as type 1, and the three curves on the left-hand side are of low type, labeled as type 2. The plain solid curves represent the means of the estimates. The curves with squares and cross constitute two bands covers the plain solid curves. These bands are 90% confidence intervals of the estimates for type 1 and type 2 respectively. The confidence intervals are generated by bootstrap. At each quantile, high-type bidders will bid higher than low-type bidders, which displays the property of FOSD. Most of the time, the two confidence intervals have no overlap. It is a sign of the robustness of our type separation. Besides bid distibutions, we are also interested in type probabilities. The estimated type probabilities are $\hat{p}_1 = 0.495$ and $\hat{p}_2 = 0.505$ with standard errors of 0.117. The two estimates share the standard errors because they always sum up to 1. Both type probabilities are significantly different from 0. Given type-specific bid distributions, we can now proceed to model-specification tests.

The AP test gives a p-value of 0.855 while the AVD test gives a p-value of 0.011. We reject AVD model and cannot reject AP model at 5% significance level. These results indicate that AP model might be more favorable in explaining USFS timber auction data. This is a new empirical evidence compared to Athey et al. (2011). Athey et al. (2011)

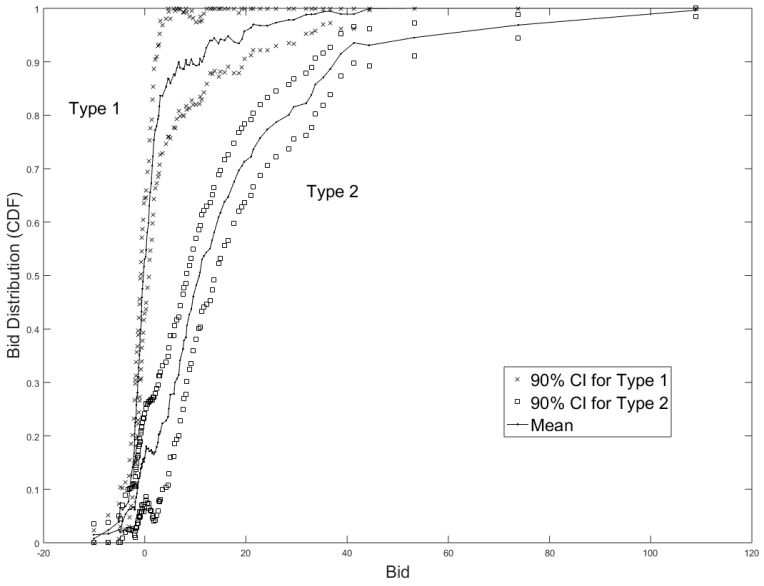


Figure 4.5: Type-Specific Bid Distributions

exploit the characteristics of bidders and realize that bidders are constituted by loggers and mills. By assuming loggers and mills have different value distributions, they deploy AVD model. This assumption could be plausible but it is also possible that loggers and mills share an identical value distribution. If the assumption fails, AP model will become a better choice. Our empirical results do not get into the argument of bidders' characteristics but relies on Assumption 4. It requires that each bidder's values across the three auctions are independent. We cannot eliminate the possibility that their values do have correlations. As shown in the section of Monte Carlo studies, as long as the correlations are small, our method is still robust. The way to handle heavily correlated values remains in question.

5. CONCLUSIONS

In the first essay, we study the estimation of the nonlinear quantile regression, where the regressors include both the continuous and discrete components. Unlike the method proposed in Li et al. (2013) which first estimates the CDF nonparametrically and then inverts the estimated CDF to obtain the quantile estimation. One of the advantage of using our proposed method is that, besides the estimate of a conditional quantile function, we also obtain the derivative function (of the quantile function) estimate and its asymptotic theory, while it is difficulty to obtain derivative function estimation and to derive the related asymptotic theory if one uses the inverse CDF method to estimate conditional quantile function.

We combine the quantile check function and a local linear smoothing technique with the mixed continuous and discrete kernel function to directly estimate the conditional quantile regression function. We establish the asymptotic normal distribution theory for the proposed local linear estimators, which generalizes some existing results which are only applicable to the case of purely continuous regressors. We also study the choice of the tuning parameters in the proposed local linear estimation procedure by proposing a CV approach to directly choose the optimal bandwidths which is different from that in Li et al. (2013), and further derive the asymptotic properties of the CV bandwidth selection approach. A simulation study is provided to examine the finite sample behavior of the proposed method. Through the simulation study, we find that our method has a better small sample performance than the naive local linear quantile estimation without smoothing the discrete regressors and the nonparametric inverse-CDF method.

In the second essay, we consider the problem of estimating a nonparametric regression function with monotonicty restriction using the nonparametric K-nearest-neighbor

method. We show that when data is highly unevenly distributed, the constrained and unconstrained Knn estimators give more reasonable estimation results than those obtained by using the kernel method. We also propose using a new distance function, which directly measures the closeness between the constrained and unconstrained Knn curves. We obtain the constrained Knn estimate by minimizing the new distance function. We use a bootstrap procedure to test the validity of the monotone constraint. Our Monte Carlo simulations demonstrate good finite sample performance of the bootstrap testing procedure. The size of the test is reliable and the power is strong, especially when our new distance measure is employed. We apply our constrained Knn estimation method to analyze the 'Job Market Matching' data and demonstrate the advantages of using Knn method to estimate regression functions with highly unevenly distributed data. The bootstrap testing procedure confirms the monotone relationship between job matching probability and market size.

In the third essay, we have developed a methodology to nonparametrically test the existence of bidders' asymmetry in two models of auctions: (1) Bidders' are risk averse and their risk preference is heterogenous (AP model); (2) Bidders' values are asymmetrically distributed (AVD model). The testing procedure took two steps: in the first step, we nonparametrically recover bid distribution conditional on bidders' unobserved heterogeneity. In the second step, we show that the difference between any two bid distributions for two distinct types is monotonic in its quantile. Specifically, the distance is increasing in AP model and decreasing in AVD model. The application of our method to USFS timber auctions demonstrates that bidders in field auction display heterogenous risk preference.

REFERENCES

An, Y. (2013). Identification of first-price auctions with non-equilibrium beliefs: a measurement error approach. *Working Paper, Texas A&M University*.

An, Y., Y. Hu, and M. Shum (2010). Estimating first-price auctions with an unknown number of bidders: a misclassification approach. *Journal of Econometrics 157*(2), 328–341.

Athey, S. (2001). Single crossing properties and the existence of pure strategy equilibria in games of incomplete information. *Econometrica 69*(4), 861–889.

Athey, S. and J. Levin (2001). Information and competition in US Forest Service timber auctions. *Journal of Political Economy 109*(2), 375–417.

Athey, S., J. Levin, and E. Seira (2011). Comparing open and sealed bid auctions: evidence from timber auctions. *Quarterly Journal of Economics 126*(1), 207–257.

Bajari, P., S. Houghton, and S. Tadelis (2007). Bidding for incomplete contracts: an empirical analysis. *NBER Working Paper #12051*.

Baldwin, L. (1995). Risk aversion in Forest Service timber auctions. *Working Paper, RAND Corporation*.

Baldwin, L., R. Marshall, and J. Richard (1997). Bidder collusion at Forest Service timber sales. *Journal of Political Economy 105*(4), 657–699.

Cai, Z. (2002). Regression quantiles for time series. *Econometric Theory 18*(01), 169–192.

Cai, Z. and Z. Xiao (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics 167*(2), 413–425.

Cai, Z. and X. Xu (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association 103*(484), 1595–1608.

Campo, S. (2012). Risk aversion and asymmetry in procurement auctions: identifica-

tion, estimation and application to construction procurements. *Journal of Econometrics 168*(1), 96–107.

Campo, S., E. Guerre, I. Perrigne, and Q. Vuong (2011). Semiparametric estimation of first-price auctions with risk-averse bidders. *The Review of Economic Studies 78*(1), 112–147.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, 5549–5632.

Delecroix, M. and C. Thomas-Agnan (2000). Spline and kernel regression under shape restrictions. *Smoothing and Regression: Approaches, Computation, and Application*, 109–133.

Du, P., C. F. Parmeter, and J. S. Racine (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statistica Sinica*, 1347–1371.

Fan, J. and I. Gijbels (1996). Local polynomial modelling and its applications: monographs on statistics and applied probability. *CRC Press*.

Fang, Z. and A. Santos (2014). Inference on directionally differentiable functions. *Working Paper, University of California, San Diego*.

Freyberger, J. and J. L. Horowitz (2015). Identification and shape restrictions in nonparametric instrumental variables estimation. *Journal of Econometrics 189*(1), 41–53.

Gan, L. and Q. Li (2016). Efficiency of thin and thick markets. *Journal of Econometrics 192*(1), 40–54.

Green, P. J. and B. W. Silverman (1993). Nonparametric regression and generalized linear models: a roughness penalty approach. *CRC Press*.

Guerre, E., I. Perrigne, and Q. Vuong (2009). Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. *Econometrica 77*(4), 1193–1227.

Haile, P. (2001). Auctions with resale markets: an application to US Forest Service timber sales. *American Economic Review 91*(3), 399–427.

Haile, P., H. Hong, and M. Shum (2003). Nonparametric tests for common values in first-price auctions. *NBER Working Paper #10105*.

Haile, P. and E. Tamer (2003). Inference with an incomplete model of english auctions. *Journal of Political Economy 111*, 1–51.

Hall, P. and L.-S. Huang (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics*, 624–647.

Hall, P., L.-S. Huang, J. A. Gifford, and I. Gijbels (2001). Nonparametric estimation of hazard rate under the constraint of monotonicity. *Journal of Computational and Graphical Statistics 10*(3), 592–614.

Hall, P., J. Racine, and Q. Li (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association 99*(468), 1015–1026.

Hallin, M., Z. Lu, K. Yu, et al. (2009). Local linear spatial quantile regression. *Bernoulli 15*(3), 659–686.

Henderson, D. J., J. A. List, D. L. Millimet, C. F. Parmeter, and M. K. Price (2012). Empirical implementation of nonparametric first-price auction models. *Journal of Econometrics 168*(1), 17–28.

Henderson, D. J. and C. F. Parmeter (2009). Imposing economic constraints in nonparametric regression: survey, implementation, and extension. *Nonparametric Econometric Methods*, 433–469.

Hong, H. and J. Li (2014). The numerical directional delta method. *Working Paper, Stanford University*.

Horowitz, J. L. (2009). Semiparametric and nonparametric methods in econometrics. *Springer*.

Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: a general solution. *Journal of Econometrics 144*,

27–61.

Hu, Y., Y. Kayaba, and M. Shum (2013). Nonparametric learning rules from bandit experiments: the eyes have it! *Games and Economic Behavior 81*, 215–231.

Hu, Y., D. McAdams, and M. Shum (2013). Identification of first-price auctions with non-separable unobserved heterogeneity. *Journal of Econometrics 174*(2), 186–193.

Koenker, R. (2005). Quantile regression. *Cambridge University Press*.

Koenker, R. and G. Bassett Jr (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33–50.

Krasnokutskaya, E. (2011). Identification and estimation of auction models with unobserved heterogeneity. *Review of Economic Studies 78*(1), 293–327.

Lebrun, B. (1999). First price auctions in the asymmetric N bidder case. *International Economic Review 40*(1), 125–142.

Lee, T.-H., Y. Tu, and A. Ullah (2014). Nonparametric and semiparametric regressions subject to monotonicity constraints: estimation and forecasting. *Journal of Econometrics 182*(1), 196–210.

Li, D., L. Simar, and V. Zelenyuk (2016). Generalized nonparametric smoothing with mixed discrete and continuous data. *Computational Statistics & Data Analysis 100*, 424–444.

Li, Q., J. Lin, and J. S. Racine (2013). Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business & Economic Statistics 31*(1), 57–65.

Li, Q. and J. Racine (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica*, 485–512.

Li, Q. and J. S. Racine (2007). Nonparametric econometrics: theory and practice. *Princeton University Press*.

Li, Q., J. S. Racine, and J. M. Wooldridge (2009). Efficient estimation of average treatment

effects with mixed categorical and continuous data. *Journal of Business & Economic Statistics 27*(2), 206–223.

Li, T., I. Perrigne, and Q. Vuong (2000). Conditionally independent private information in ocs wildcat auctions. *Journal of Econometrics 98*, 129–161.

Li, T. and Q. Vuong (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis 65*(2), 139–165.

Li, Z., G. Liu, and Q. Li (2016). Nonparametric knn estimation with monotone constraints. *Econometric Reviews*. forthcoming.

Lin, Z. and D. Li (2007). Asymptotic normality for L1-norm kernel estimator of conditional median under association dependence. *Journal of Multivariate Analysis 98*(6), 1214–1230.

Malikov, E., S. C. Kumbhakar, and Y. Sun (2016). Varying coefficient panel data model in the presence of endogenous selectivity and fixed effects. *Journal of Econometrics 190*(2), 233–251.

Pagan, A. and A. Ullah (1999). Nonparametric econometrics. *Cambridge University Press*.

Racine, J. and Q. Li (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics 119*(1), 99–130.

Robin, J.-M. and R. J. Smith (2000). Tests of rank. *Econometric Theory 16*(02), 151–175.

Sun, K. (2015). Constrained nonparametric estimation of input distance function. *Journal of Productivity Analysis 43*(1), 85–97.

Wand, M. and M. Jones (1995). Kernel smoothing. *Chapman & Hall, London*.

Xu, K.-L. and P. C. Phillips (2011). Tilted nonparametric estimation of volatility functions with empirical applications. *Journal of Business & Economic Statistics 29*(4), 518–528.

Yu, K. and M. Jones (1998). Local linear quantile regression. *Journal of the American Statistical Association 93*(441), 228–237.

Yu, K. and Z. Lu (2004). Local linear additive quantile regression. *Scandinavian Journal of Statistics 31*(3), 333–346.

APPENDIX A

PROOFS IN THE FIRST ESSAY

A.1 Assumptions

In this appendix, we give some regularity conditions which have been used to derive the asymptotic properties of the proposed approaches in Sections 2.1 and 2.2.

ASSUMPTION 1. *The kernel function $\mathbb{K}(\cdot)$ is a continuous and symmetric probability density function with a compact support.*

ASSUMPTION 2. *The sequence of $\{(Y_i, X_i, Z_i)\}$ is composed of independent and identically distributed (i.i.d.) random vectors.*

ASSUMPTION 3. *(i) The conditional density function of e_i for given $X_i = x$ and $Z_i = z$, $f_e(\cdot|x, z)$, exists and is continuous at point zero. Furthermore, $f_e(0|x, z)$ is continuous with respect to x when $z = z_0$ or z_1. Let the conditional cumulative distribution function of e_i for given $X_i = x$ and $Z_i = z$, $\mathbb{F}_e(\cdot|x, z)$, be continuous with respect to x when $z = z_0$ or z_1, and $\mathbb{F}_e(0|x_0, z_0) = \tau$.*

(ii) The probability of $Z_i = z_0$, p_1, satisfies that $0 < p_1 < 1$. The conditional density function of X_i for given $Z_i = z$, $f(x|z)$, is bounded away from infinity and zero for $x \in \mathcal{X}$ and $z = z_0$ or z_1, where $\mathcal{X}$ is the compact support of X_i.

(iii) The conditional quantile regression function $\mathbb{Q}_\tau(\cdot, z)$ and $f(\cdot|z)$ are twice continuously differentiable on $\mathcal{X}$ for $z = z_0$ or z_1.

ASSUMPTION 4. *(i) Let the tuning parameters h and λ satisfy*

$$h \to 0, \;\; nh \to \infty \;\text{ and }\; \lambda = O(h^2).$$

(ii) Let the tuning parameter h satisfy

$$nh^6 \to 0, \;\; nh^{9/2} \to \infty.$$

Assumption 1 imposes some mild conditions on the continuous kernel function in the nonparametric kernel-based smoothing, and several commonly-used kernel functions such as the uniform kernel and the Epanechnikov kernel satisfy these conditions (c.f., Fan and Gijbels, 1996; Li and Racine, 2007). As the kernel function is symmetric, it is easy to check that $\mu_k = 0$ when k is odd. In Assumption 2, we impose the *i.i.d.* condition on the observations, which has been commonly used in the literature on nonparametric estimation with both categorical and continuous data (c.f., Li and Racine, 2004; Racine and Li 2004; Li, Simar and Zelenyuk, 2014). The developed asymptotic theory in this paper can be generalized to the general stationary and weakly dependent processes at the cost of more lengthy proofs. There is no moment condition on e_i to estimate the conditional quantile regression, which indicates that the distribution of e_i is allowed to have heavy tails. Assumption 3 gives some smoothness conditions on the (conditional) density functions and the quantile regression function, which are necessary in order to apply the local linear smoothing method. Assumptions 4 imposes some restrictions on the various tuning parameters. In particular, Assumptions 4(ii) is only needed when we derive the asymptotic expansion of the CV-based loss function.

A.2 Proofs of the Main Results

Before proving the main results in Sections 2.1 and 2.2, we introduce some notations to simplify the presentation. Let

$$u_n(\alpha; x_0, z_0) = \sqrt{nh}\big[\alpha - \mathbb{Q}_\tau(x_0, z_0)\big], \;\; v_n(\beta; x_0, z_0) = \sqrt{nh^3}\big[\beta - \mathbb{Q}_\tau^{(1)}(x_0, z_0)\big].$$

Let

$$\begin{aligned}\Delta_{ni}(\alpha, \beta; x_0, z_0) &= \frac{1}{\sqrt{nh}}\Big[u_n(\alpha; x_0, z_0) + v_n(\beta; x_0, z_0)\big(\frac{X_i - x_0}{h}\big)\Big], \\ b_i(x_0, z_0) &= \mathbb{Q}_\tau(X_i, Z_i) - \mathbb{Q}_\tau(x_0, z_0) - \mathbb{Q}_\tau^{(1)}(x_0, z_0)(X_i - x_0).\end{aligned}$$

With the help of the above notations, it is easy to see that

$$Y_i - \alpha - \beta(X_i - x_0) = e_i - \Delta_{ni}(\alpha, \beta; x_0, z_0) + b_i(x_0, z_0). \tag{B.1}$$

Throughout the proof, we let $a_n \sim b_n$ and $a_n \overset{P}{\sim} b_n$ denote $a_n = b_n(1 + o(1))$ and $a_n = b_n(1 + o_P(1))$, respectively.

We next give the detailed proofs of Theorems 2.1 and 2.2.

PROOF OF THEOREM 2.1

Letting

$$\overline{\mathcal{L}}_n(x_0, z_0) = \sum_{i=1}^{n} \rho_\tau\big[e_i + b_i(x_0)\big]\mathbb{K}_h(X_i - x_0)\Lambda_\lambda(Z_i, z_0)$$

which is unrelated to α and β, and using (B.1), it is straightforward to show that minimiz-

ing $\mathcal{L}_n(\alpha, \beta; x_0, z_0)$ is equivalent to minimizing

$$\begin{aligned}\widetilde{\mathcal{L}}_n(\alpha, \beta; x_0, z_0) &\equiv \mathcal{L}_n(\alpha, \beta; x_0, z_0) - \overline{\mathcal{L}}_n(x_0, z_0), \\ &= \sum_{i=1}^{n} \Big\{ \rho_\tau\big[e_i + b_i(x_0, z_0) - \Delta_{ni}(\alpha, \beta; x_0, z_0)\big] \\ &\quad -\rho_\tau\big[e_i + b_i(x_0, z_0)\big] \Big\} \mathbb{K}_h(X_i - x_0)\Lambda_\lambda(Z_i, z_0). \qquad \text{(B.2)}\end{aligned}$$

Note that the identity result (c.f., Knight, 1998):

$$\rho_\tau(x - y) - \rho_\tau(x) = y\big[I_{\{x \le 0\}} - \tau\big] + \int_0^y \big[I_{\{x \le z\}} - I_{\{x \le 0\}}\big] dz.$$

Letting $x = e_i + b_i(x_0, z_0)$ and $y = \Delta_{ni}(\alpha, \beta; x_0, z_0) \equiv \Delta_{ni}$ in the above identity and denoting $\eta_i(x_0, z_0) = I_{\{e_i \le -b_i(x_0, z_0)\}} - \tau$, $\widetilde{\mathcal{L}}_n(\alpha, \beta; x_0, z_0)$ can be rewritten as

$$\begin{aligned}\widetilde{\mathcal{L}}_n(\alpha, \beta; x_0, z_0) &= \Psi_n(x_0, z_0) \\ &\quad + u_n(\alpha; x_0, z_0)\big[\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \eta_i(x_0, z_0)\mathbb{K}_h(X_i - x_0)\Lambda_\lambda(Z_i, z_0)\big] + \\ &\quad v_n(\beta; x_0, z_0)\big[\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \eta_i(x_0, z_0)\big(\frac{X_i - x_0}{h}\big)\mathbb{K}_h(X_i - x_0)\Lambda_\lambda(Z_i, z_0)\big],\end{aligned}$$

where

$$\Psi_n(x_0, z_0) = \sum_{i=1}^{n} \mathbb{K}_h(X_i - x_0)\Lambda_\lambda(Z_i, z_0) \int_0^{\Delta_{ni}} \Big[I_{\{e_i \le z - b_i(x_0, z_0)\}} - I_{\{e_i \le -b_i(x_0, z_0)\}} dz.$$

We first consider $\Psi_n(x_0, z_0)$. By Assumptions 2 and 3(i), we can prove that

$$\begin{aligned}
& \mathbb{E}\big[\Psi_n(x_0,z_0)|\mathcal{F}_n(X,Z)\big] \\
= & \sum_{i=1}^{n} \mathbb{K}_h(X_i - x_0)\Lambda_\lambda(Z_i, z_0) \int_0^{\Delta_{ni}} \Big\{ \mathbb{F}_e\big[z - b_i(x_0,z_0)|X_i, Z_i\big] \\
& - \mathbb{F}_e\big[- b_i(x_0,z_0)|X_i, Z_i\big] \Big\} dz \\
\overset{P}{\sim} & \sum_{i=1}^{n} \mathbb{K}_h(X_i - x_0)\Lambda_\lambda(Z_i, z_0) \int_0^{\Delta_{ni}} z f_e\big[- b_i(x_0,z_0)|X_i, Z_i\big] dz \\
\overset{P}{\sim} & \frac{1}{2}\sum_{i=1}^{n} \mathbb{K}_h(X_i - x_0)\Lambda_\lambda(Z_i, z_0) f_e(0|X_i, Z_i)\Delta_{ni}^2 \\
\equiv & \frac{1}{2}\big[u_n(\alpha; x_0, z_0), v_n(\beta; x_0, z_0)\big] S_n(x_0, z_0) \big[u_n(\alpha; x_0, z_0), v_n(\beta; x_0, z_0)\big]', \quad \text{(B.3)}
\end{aligned}$$

where $\mathcal{F}_n(X, Z)$ is the σ-field generated by (X_i, Z_i), $i = 1, \cdots, n$,

$$S_n(x_0, z_0) = \begin{bmatrix} S_{n1}(x_0,z_0) & S_{n2}(x_0,z_0) \\ S_{n2}(x_0,z_0) & S_{n3}(x_0,z_0) \end{bmatrix}$$

with

$$S_{nk}(x_0, z_0) = \frac{1}{nh}\sum_{i=1}^{n} \big(\frac{X_i - x_0}{h}\big)^{k-1} \mathbb{K}_h(X_i - x_0)\Lambda_\lambda(Z_i, z_0) f_e(0|X_i, Z_i), \quad k = 1, 2, 3.$$

Following the proof of Theorem 3.1 in Li, Simar and Zelenyuk (2014) and using Assumptions 1, 3(i)(ii) and 4(i), we may show that

$$S_{nk}(x_0, z_0) = h^{-1} p_1 \mu_{k-1} f(x_0|z_0) f_e(0|x_0, z_0) + o_P(h^{-1}),$$

which indicates that

$$hS_n(x_0, z_0) \overset{P}{\sim} S(x_0, z_0) \equiv \begin{bmatrix} S_1(x_0, z_0) & S_2(x_0, z_0) \\ S_2(x_0, z_0) & S_3(x_0, z_0) \end{bmatrix}. \tag{B.4}$$

Meanwhile, we can prove that the conditional variance $\Psi_n(x_0, z_0)$ for given $\mathcal{F}_n(X, Z)$ satisfies

$$\sqrt{\mathbb{V}\big[h\Psi_n(x_0, z_0)|\mathcal{F}_n(X, Z)\big]} = o_P(1). \tag{B.5}$$

By (B.3)–(B.5), we can prove

$$h\Psi_n(x_0, z_0) \overset{P}{\sim} \frac{1}{2}\big[u_n(\alpha; x_0, z_0), v_n(\beta; x_0, z_0)\big] S(x_0, z_0) \big[u_n(\alpha; x_0, z_0), v_n(\beta; x_0, z_0)\big]'. \tag{B.6}$$

Let

$$W_n(k) \equiv W_n(k; x_0, z_0) = \frac{h}{\sqrt{nh}} \sum_{i=1}^{n} \eta_i(x_0, z_0) \big(\frac{X_i - x_0}{h}\big)^{k-1} \mathbb{K}_h(X_i - x_0) \Lambda_\lambda(Z_i, z_0)$$

for $k = 1$ and 2, and $W_n(x_0, z_0) = \big[W_n(1), W_n(2)\big]'$. Note that

$$h\widetilde{\mathcal{L}}_n(\alpha, \beta; x_0, z_0) - \big[u_n(\alpha; x_0, z_0), v_n(\beta; x_0, z_0)\big] W_n(x_0, z_0)$$

converges in probability to the right hand side of (B.6), which is a convex function. Then, by Pollard (1991)'s convexity lemma, the minimizer to $\widetilde{\mathcal{L}}_n(\alpha, \beta; x_0, z_0)$ satisfies

$$\big[\widehat{u}_n(x_0, z_0), \widehat{v}_n(x_0, z_0)\big]' \overset{P}{\sim} \big[S(x_0, z_0)\big]^{-1} W_n(x_0, z_0), \tag{B.7}$$

where $\widehat{u}_n(x_0, z_0) = \sqrt{nh}\big[\widehat{\mathbb{Q}}_\tau(x_0, z_0) - \mathbb{Q}_\tau(x_0, z_0)\big]$ and $\widehat{v}_n(x_0, z_0) = \sqrt{nh^3}\big[\widehat{\mathbb{Q}}^{(1)}_\tau(x_0, z_0) -$

$\mathbb{Q}_\tau^{(1)}(x_0, z_0)\big]$. Thus, to establish the asymptotic distribution theory of

$$\big[\widehat{u}_n(x_0, z_0), \widehat{v}_n(x_0, z_0)\big]',$$

we only need to derive the limiting distribution of $W_n(x_0, z_0)$.

Let $W_{n*}(x_0, z_0)$ be defined as $W_n(x_0, z_0)$ with $\eta_i \equiv \eta_i(x_0, z_0)$ replaced by $\eta_{i*} = I_{\{e_i \le 0\}} - \tau$. Then,we have

$$\begin{aligned} W_n(x_0, z_0) - \mathbb{E}\big[W_n(x_0, z_0)\big] &= W_{n*}(x_0, z_0) - \mathbb{E}\big[W_{n*}(x_0, z_0)\big] \\ &\quad + W_n(x_0, z_0) - W_{n*}(x_0, z_0) \\ &\quad - \mathbb{E}\big[W_n(x_0, z_0) - W_{n*}(x_0, z_0)\big]. \end{aligned} \tag{B.8}$$

Note that

$$\begin{aligned} & \mathbb{V}\Big[W_n(x_0, z_0) - W_{n*}(x_0, z_0)\Big] \\ \le\ & \mathbb{E}\Big[\big\|W_n(x_0, z_0) - W_{n*}(x_0, z_0)\big\|^2\Big] \\ =\ & \mathbb{E}\Big[\big\|W_n(x_0, z_0) - W_{n*}(x_0, z_0)\big\|^2 \big| \mathcal{F}_n(X, Z)\Big] \\ \le\ & O\Big(\frac{h}{n}\sum_{i=1}^n \mathbb{E}\Big\{\mathbb{K}_h^2(X_i - x_0)\Lambda_\lambda^2(Z_i, z_0)\mathbb{E}\big[(\eta_i - \eta_{i*})^2 \big| \mathcal{F}_n(X, Z)\big]\Big\}\Big) \\ \le\ & O\Big(\frac{h}{n}\sum_{i=1}^n \mathbb{E}\Big\{\mathbb{K}_h^2(X_i - x_0)\Lambda_\lambda^2(Z_i, z_0)\big[\mathbb{F}_e(-b_i(x_0, z_0)) - \mathbb{F}_e(0)\big]\Big\}\Big) \\ \le\ & o\Big(\frac{1}{nh}\sum_{i=1}^n \mathbb{E}\Big\{\mathbb{K}^2\big(\frac{X_i - x_0}{h}\big)\Lambda_\lambda^2(Z_i, z_0)\Big\}\Big) = o(1), \end{aligned}$$

which implies that $W_{n*}(x_0, z_0) - \mathbb{E}\big[W_{n*}(x_0, z_0)\big]$ is the leading term of $W_n(x_0, z_0) - \mathbb{E}\big[W_n(x_0, z_0)\big]$. We next turn to the proof of

$$W_{n*}(x_0, z_0) - \mathbb{E}\big[W_{n*}(x_0, z_0)\big] \xrightarrow{d} \mathbb{N}\big[\mathbf{0}, \Omega(x_0, z_0)\big], \tag{B.9}$$

where $\Omega(x_0, z_0) = \tau(1-\tau)p_1 f(x_0|z_0)\mathrm{diag}(\nu_0, \nu_2)$ is defined as in Section 2.1. By the so-called Cramér-Wold device (Billingsley, 1968) and the classical central limit theorem for the *i.i.d.* random vectors, we can complete the proof of (B.9). In view of of (B.8) and (B.9), we can show that

$$W_n(x_0, z_0) - \mathbb{E}\big[W_n(x_0, z_0)\big] \xrightarrow{d} \mathbb{N}\big[\mathbf{0}, \Omega(x_0, z_0)\big]. \tag{B.10}$$

Meanwhile, note that when X_i and x_0 are close enough,

$$b_i(x_0, z_0) \approx \frac{1}{2}\mathbb{Q}_\tau^{(2)}(x_0, z_0)(X_i - x_0)^2$$

for $Z_i = z_0$; and

$$b_i(x_0, z_0) \approx \mathbb{Q}(x_0, z_1) - \mathbb{Q}(x_0, z_0)$$

for $Z_i = z_1 \neq z_0$. Then, by some elementary calculations, we may also prove that

$$\frac{1}{\sqrt{nh}}\big[S(x_0, z_0)\big]^{-1}\mathbb{E}\big[W_n(x_0, z_0)\big] \sim \bar{b}(x_0, z_0), \tag{B.11}$$

where $\bar{b}(x_0, z_0)$ is defined as in Section 2.1. By (B.10) and (B.11), we can prove

$$W_n(x_0, z_0) - \sqrt{nh}S(x_0, z_0)\bar{b}(x_0, z_0) \xrightarrow{d} \mathbb{N}\big[\mathbf{0}, \Omega(x_0, z_0)\big]. \tag{B.12}$$

The proof of Theorem 2.1 can be completed by using (B.7) and (B.12). □

PROOF OF THEOREM 2.2

Throughout this proof, we let

$$\widehat{\mathbb{Q}}_{(-i)}(X_i, Z_i) = \widehat{\mathbb{Q}}_{(-i)}(X_i, Z_i; h, \lambda), \zeta_i(X_i, Z_i) = \widehat{\mathbb{Q}}_{(-i)}(X_i, Z_i) - \mathbb{Q}_\tau(X_i, Z_i).$$

Note that

$$\begin{aligned}\mathbb{CV}(h,\lambda) &= \sum_{i=1}^{n}\rho_\tau\big[e_i+\mathbb{Q}_\tau(X_i,Z_i)-\widehat{\mathbb{Q}}_{(-i)}(X_i,Z_i)\big]M(X_i,Z_i)\\ &= \sum_{i=1}^{n}\rho_\tau(e_i)M(X_i,Z_i)+\sum_{i=1}^{n}\big\{\rho_\tau\big[e_i-\zeta_i(X_i,Z_i)\big]-\rho_\tau(e_i)\big\}M(X_i,Z_i)\\ &\equiv \mathbb{CV}_1+\mathbb{CV}_2(h,\lambda). \end{aligned} \tag{B.13}$$

It is easy to see that $\mathbb{CV}_1$ does not rely on the tuning parameters h and λ, which indicates that this term would not play any role in choosing the optimal bandwidths. Thus, to complete the proof of Theorem 2.2, we only need to derive the asymptotic order for $\mathbb{CV}_2(h,\lambda)$.

Using Knight (1998)'s identity result, we may show that

$$\rho_\tau\big[e_i-\zeta_i(X_i,Z_i)\big]-\rho_\tau(e_i)=\zeta_i(X_i,Z_i)\big[I_{\{e_i\le 0\}}-\tau\big]+\int_0^{\zeta_i(X_i,Z_i)}\big[I_{\{e_i\le z\}}-I_{\{e_i\le 0\}}\big]dz. \tag{B.14}$$

Define

$$\begin{aligned}\mathbb{CV}_{21}(h,\lambda) &= \sum_{i=1}^{n}\zeta_i(X_i,Z_i)\big[I_{\{e_i\le 0\}}-\tau\big]M(X_i,Z_i),\\ \mathbb{CV}_{22}(h,\lambda) &= \sum_{i=1}^{n}M(X_i,Z_i)\int_0^{\zeta_i(X_i,Z_i)}\big[I_{\{e_i\le z\}}-I_{\{e_i\le 0\}}\big]dz.\end{aligned}$$

We next derive the asymptotic orders for $\mathbb{CV}_{21}(h,\lambda)$ and $\mathbb{CV}_{22}(h,\lambda)$, respectively.

By the uniform asymptotic approximation of $\zeta_i(X_i,Z_i)$ in Lemma C.1 in Appendix C, we have uniformly for $i=1,\cdots,n$,

$$\zeta_i(X_i,Z_i)\overset{P}{\sim}(1,0)\big[S(X_i,Z_i)\big]^{-1}W_{(-i)}(X_i,Z_i), \tag{B.15}$$

where $W_{(-i)}(x,z) = \big[W_{(-i),1}(x,z), W_{(-i),2}(x,z)\big]'$ with

$$W_{(-i),k}(x,z) = \frac{h}{nh}\sum_{j\neq i}\eta_j(x,z)\big(\frac{X_j - x}{h}\big)^{k-1}\mathbb{K}_h(X_j - x)\Lambda_\lambda(Z_j, z)$$

and $\eta_j(x,z)$ is defined as in the proof of Theorem 2.1. Letting $\eta_{i*} = I_{\{e_i\leq 0\}} - \tau$, by (B.15) and following the proof of (B.12) above, we may show that

$$\begin{aligned}
\mathbb{CV}_{21}(h,\lambda) &= \sum_{i=1}^{n}\zeta_{(-i)}(X_i,Z_i)\eta_{i*}M(X_i,Z_i) \\
&\overset{P}{\sim} \sum_{i=1}^{n}\eta_{i*}\bar{b}_1(X_i,Z_i)M(X_i,Z_i) + \\
&\quad \frac{1}{nh}\sum_{i=1}^{n}\sum_{j\neq i}\eta_{i*}\eta_{j*}\mathbb{K}\big(\frac{X_i - X_j}{h}\big)\Lambda_\lambda(Z_i,Z_j) \\
&\equiv \mathbb{CV}_{211}(h,\lambda) + \mathbb{CV}_{212}(h,\lambda), \qquad \text{(B.16)}
\end{aligned}$$

where $\bar{b}_1(X_i,Z_i)$ is defined as in Section 2.1. Following the proof of Theorem 3.1 in Li, Simar and Zelenyuk (2014), we can prove that

$$\mathbb{CV}_{211}(h,\lambda) = O_P(\sqrt{n}h^2 + \sqrt{n}\lambda) = O_P(\sqrt{n}h^2) \qquad \text{(B.17)}$$

and

$$\mathbb{CV}_{212}(h,\lambda) = O_P(h^{-1/2}). \qquad \text{(B.18)}$$

Using (B.16)–(B.18), we can show that

$$\mathbb{CV}_{21}(h,\lambda) = O_P(\sqrt{n}h^2 + h^{-1/2}). \qquad \text{(B.19)}$$

We next consider $\mathbb{CV}_{22}(h,\lambda)$. It is easy to verify that

$$\mathbb{CV}_{22}(h,\lambda) \overset{P}{\sim} \mathbb{E}\big[\mathbb{CV}_{22}(h,\lambda)\big|\mathcal{F}_n(X,Z)\big] \tag{B.20}$$

and

$$\begin{aligned}\mathbb{E}\big[\mathbb{CV}_{22}(h,\lambda)\big|\mathcal{F}_n(X,Z)\big] &\overset{P}{\sim} \sum_{i=1}^{n} M(X_i,Z_i) f_e(0|X_i,Z_i) \int_0^{\zeta_i(X_i,Z_i)} z dz \\ &= \frac{1}{2}\sum_{i=1}^{n} \zeta_i^2(X_i,Z_i) M(X_i,Z_i) f_e(0|X_i,Z_i). \end{aligned} \tag{B.21}$$

Furthermore, following the proof of Lemma C.2 in Appendix C below, we have

$$\begin{aligned}&\sum_{i=1}^{n} \zeta_i^2(X_i,Z_i) M(X_i,Z_i) f_e(0|X_i,Z_i) \\ \overset{P}{\sim}\ &\sum_{i=1}^{n} b^2(X_i,Z_i;h,\lambda) M(X_i,Z_i) f_e(0|X_i,Z_i) + \\ &\sum_{i=1}^{n} \sigma^2(X_i,Z_i;h) M(X_i,Z_i) f_e(0|X_i,Z_i), \end{aligned} \tag{B.22}$$

where $b(X_i,Z_i;h,\lambda)$ and $\sigma^2(X_i,Z_i;h)$ are defined in Section 2.2. By (B.20)–(B.22), we can show that

$$\mathbb{CV}_{22}(h,\lambda) \overset{P}{\sim} \frac{1}{2}\sum_{i=1}^{n} \big[b^2(X_i,Z_i;h,\lambda) + \sigma^2(X_i,Z_i;h)\big] M(X_i,Z_i) f_e(0|X_i,Z_i). \tag{B.23}$$

Note that $\mathbb{CV}_{21}(h,\lambda)$ is asymptotically negligible (compared with $\mathbb{CV}_{22}(h,\lambda)$) by Assumption 4(ii). We then complete the proof of Theorem 2.2 by (B.13), (B.19) and (B.23). □

A.3 Some Auxiliary Lemmas

In this appendix, we give some auxiliary lemmas which have been used to prove the main results in Appendix A.2.

LEMMA C.1. *Suppose that the conditions of Theorem 2.2 are satisfied. Then we have*

$$\big[\widehat{u}_n(x,z),\widehat{v}_n(x,z)\big]' \overset{P}{\sim} \big[S(x,z)\big]^{-1}W_n(x,z) \tag{C.1}$$

uniformly for $x \in \mathcal{X}$ and $z = z_0$ or z_1, where $\widehat{u}_n(x,z)$, $\widehat{v}_n(x,z)$, $S(x,z)$ and $W_n(x,z)$ are defined in Appendix A.2.

PROOF:

Following the proof of Theorem 2.2 and using the convexity lemma in Pollard (1991), we only need to show that

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}\big(\frac{X_i-x_0}{h}\big)^k \mathbb{K}_h(X_i-x_0)\Lambda_\lambda(Z_i,z_0)f_e(0|X_i,Z_i) \\ = \;& \mu_k f_{XZ}(x,z)f_e(0|x,z)+o_P(1) \end{aligned} \tag{C.2}$$

uniformly for $x \in \mathcal{X}$ and $z = z_0$ or z_1, $k \geq 1$, where $f_{XZ}(x,z_0) = p_1 f(x|z_0)$ and $f_{XZ}(x,z_1) = (1-p_1)f(x|z_1)$. It is easy to prove (C.2) using standard calculation and the uniform consistency result in Mack and Silverman (1982). Thus, the proof of Lemma C.1 has been completed. □

LEMMA C.2. *Suppose that the conditions of Theorem 2.2 are satisfied. Let $\zeta_i(x,z)$ be defined as in the proof of Theorem 2.2. Then we have*

$$\sum_{i=1}^{n}\zeta_i^2(X_i,Z_i) \overset{P}{\sim} \sum_{i=1}^{n}\big[b^2(X_i,Z_i;h,\lambda)+\sigma^2(X_i,Z_i;h)\big], \tag{C.3}$$

where $b^2(x,z;h,\lambda)$ and $\sigma^2(x,z;h)$ are defined in Section 2.2.

PROOF:

By Lemma C.1 and the definition of $\zeta_i(x,z)$ in the proof of Theorem 2.2, we have uniformly for $x \in \mathcal{X}$ and $z = z_0$ or z_1,

$$\zeta_i(x,z) \overset{P}{\sim} (1,0)\big[S(x,z)\big]^{-1} W_{(-i)}(x,z), \tag{C.4}$$

where $W_{(-i)}(x,z) = \big[W_{(-i),1}(x,z), W_{(-i),2}(x,z)\big]'$ with

$$W_{(-i),k}(x,z) = \frac{h}{nh}\sum_{j\neq i}\eta_j(x,z)\big(\frac{X_j - x}{h}\big)^{k-1}\mathbb{K}_h(X_j - x)\Lambda_\lambda(Z_j,z)$$

and $\eta_j(x,z)$ is defined as in the proof of Theorem 2.1.

Recall that $\eta_{i*} = I_{\{e_i \le 0\}} - \tau$. It is easy to show that

$$\begin{aligned} W_{(-i),k}(x,z) &= \frac{1}{nh}\sum_{j\neq i}\eta_j(x,z)\big(\frac{X_j - x}{h}\big)^{k-1}\mathbb{K}\big(\frac{X_j - x}{h}\big)\Lambda_\lambda(Z_j,z) \\ &= \frac{1}{nh}\sum_{j\neq i}\big[\eta_j(x,z) - \eta_{j*}\big]\big(\frac{X_j - x}{h}\big)^{k-1}\mathbb{K}\big(\frac{X_j - x}{h}\big)\Lambda_\lambda(Z_j,z) + \\ &\quad \frac{1}{nh}\sum_{j\neq i}\eta_{j*}\big(\frac{X_j - x}{h}\big)^{k-1}\mathbb{K}\big(\frac{X_j - x}{h}\big)\Lambda_\lambda(Z_j,z) \\ &\equiv B_{ik}(x,z) + V_{ik}(x,z). \end{aligned} \tag{C.5}$$

Let

$$B_i(x,z) = (1,0)\big[S(x,z)\big]^{-1}\big[B_{i1}(x,z), B_{i2}(x,z)\big]'$$

and

$$V_i(x,z) = (1,0)\big[S(x,z)\big]^{-1}\big[V_{i1}(x,z), V_{i2}(x,z)\big]'.$$

Hence, by (C.4) and (C.5), we have

$$\sum_{i=1}^{n} \zeta_i^2(X_i, Z_i) \overset{P}{\sim} \sum_{i=1}^{n} B_i^2(X_i, Z_i) + \sum_{i=1}^{n} V_i^2(X_i, Z_i) + 2\sum_{i=1}^{n} B_i(X_i, Z_i)V_i(X_i, Z_i). \quad \text{(C.6)}$$

We next consider the three terms on the right hand side of (C.6), respectively. It is easy to see that the first term would lead to the asymptotic bias term, i.e.,

$$\sum_{i=1}^{n} B_i^2(X_i, Z_i) \overset{P}{\sim} \sum_{i=1}^{n} b_i^2(X_i, Z_i; h, \lambda). \quad \text{(C.7)}$$

By the uniform consistency of the nonparametric kernel smoothing (c.f., Mack and Silverman, 1982), we can show that

$$\sum_{i=1}^{n} B_i(X_i, Z_i)V_i(X_i, Z_i) = O_P(\sqrt{n}h^2). \quad \text{(C.8)}$$

Hence, to prove (C.3), it is sufficient for us to show that

$$\sum_{i=1}^{n} V_i^2(X_i, Z_i) \overset{P}{\sim} \sum_{i=1}^{n} \sigma_i^2(X_i, Z_i; h, \lambda), \quad \text{(C.9)}$$

which can be proved by showing that

$$\begin{aligned} & \sum_{i=1}^{n} \sum_{j=1,\neq i}^{n} \sum_{k=1,\neq i}^{n} \eta_{j*} \mathbb{K}\big(\frac{X_i - X_j}{h}\big) \mathbb{K}\big(\frac{X_i - X_k}{h}\big) \Lambda_\lambda(Z_j, Z_i) \Lambda_\lambda(Z_k, Z_i) \eta_{k*} \\ \overset{P}{\sim} \; & V_{n*} n^2 h, \end{aligned} \quad \text{(C.10)}$$

where $V_{n*} = \tau(1-\tau)\nu_0 \frac{1}{n} \sum_{i=1}^{n} f_{XZ}(X_i, Z_i)$ with $f_{XZ}(\cdot, \cdot)$ defined in the proof of Lemma

C.1. Observe that

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1,\neq i}^{n}\sum_{k=1,\neq i}^{n}\eta_{j*}\mathbb{K}\big(\frac{X_i-X_j}{h}\big)\mathbb{K}\big(\frac{X_i-X_k}{h}\big)\Lambda_\lambda(Z_j,Z_i)\Lambda_\lambda(Z_k,Z_i)\eta_{k*}\\
=\;&\sum_{i=1}^{n}\sum_{j=1,\neq i}^{n}\sum_{k=1,\neq i,j}^{n}\eta_{j*}\mathbb{K}\big(\frac{X_i-X_j}{h}\big)\mathbb{K}\big(\frac{X_i-X_k}{h}\big)\Lambda_\lambda(Z_j,Z_i)\Lambda_\lambda(Z_k,Z_i)\eta_{k*}+\\
&\sum_{i=1}^{n}\sum_{j=1,\neq i}^{n}\eta_{j*}^2\mathbb{K}^2\big(\frac{X_i-X_j}{h}\big)\Lambda_\lambda^2(Z_j,Z_i).
\end{aligned}$$

By the uniform consistency result of the nonparametric kernel smoothing and the Law of Large Numbers, we can prove

$$\begin{aligned}
&\frac{1}{n^2h}\sum_{i=1}^{n}\sum_{j=1,\neq i}^{n}\eta_{j*}^2\mathbb{K}^2\big(\frac{X_i-X_j}{h}\big)\Lambda_\lambda^2(Z_j,Z_i)\\
=\;&\frac{1}{n}\sum_{j=1}^{n}\eta_{j*}^2\Big[\frac{1}{nh}\sum_{j=1,\neq i}^{n}\mathbb{K}^2\big(\frac{X_i-X_j}{h}\big)\Lambda_\lambda^2(Z_j,Z_i)\Big]\\
=\;&\frac{\nu_0}{n}\sum_{j=1}^{n}\eta_{j*}^2 f_{XZ}(X_j,Z_j)\\
\overset{P}{\sim}\;&\tau(1-\tau)\nu_0\frac{1}{n}\sum_{i=1}^{n}f_{XZ}(X_i,Z_i)\equiv V_{n*}. \qquad \text{(C.11)}
\end{aligned}$$

By some standard calculations, we can show that

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1,\neq i}^{n}\sum_{k=1,\neq i,j}^{n}\eta_{j*}\mathbb{K}\big(\frac{X_i-X_j}{h}\big)\mathbb{K}\big(\frac{X_i-X_k}{h}\big)\Lambda_\lambda(Z_j,Z_i)\Lambda_\lambda(Z_k,Z_i)\eta_{k*}\\
=\;&O_P(n^{3/2}h). \qquad \text{(C.12)}
\end{aligned}$$

The proof of (C.10) can be completed by using (C.11) and (C.12).

APPENDIX B

PROOFS IN THE THIRD ESSAY

B.1 Details of Testing Rank

In this section we briefly describe main idea behind the testing method proposed in Robin and Smith (2000) and how it can be applied to test the rank of B_{d_1,d_3}. The first step is to transform the matrix B_{d_1,d_3} into two quadratic forms: $B_1 = B_{d_1,d_3}B'_{d_1,d_3}$ and $B_2 = B'_{d_1,d_3}B_{d_1,d_3}$. Note that B_1 and B_2 have identical eigenvalues and they have the same number of zero eigenvalues as B_{d_1,d_3}, thus test the rank of B_{d_1,d_3} is equivalent to test the number of non-zero eigenvalues that B_1 has. Under the null hypothesis H_0 (rank(B_{d_1,d_3}) $= r^*$), B_1 has r^* non-zero eigenvalues and $M - r^*$ zero eigenvalues. Recall that B_1 is in quadratic form, its eigenvalues are all non-negative and zero eigenvalues will be the smallest. As a result, the summation of its smallest $M-r^*$ eigenvalues equals to zero under H_0 and we reject H_0 if the summation is too large. This leads to the test statistics $TS = n\sum_{i=1}^{M-r^*} EIG_i$, where $EIG_1, EIG_2, \cdots, EIG_{M-r^*}$ are the smallest estimated eigenvalues of B_1 and n is the number of bidders. Denote the corresponding eigenvectors as $C_1, C_2, \cdots, C_{M-r^*}$ and $C = (C_1\ C_2\ \cdots\ C_{M-r^*})$ is a matrix that collects the eigenvectors. Similarly, we pick up the smallest $M - r^*$ estimated eigenvalues of B_2 and denote the corresponding eigenvectors as $D_1, D_2, \cdots, D_{M-r^*}$ and $D = (D_1\ D_2\ \cdots\ D_{M-r^*})$. Both C and D have a dimension of $M \times (M - r^*)$.

Note that $\hat{B}_{d_1,d_3}$ is a frequency estimator of the joint pmf of d_1 and d_3, then by Central Limit Theorem (CLT)

$$\sqrt{n}(vec(\hat{B}_{d_1,d_3}) - vec(B_{d_1,d_3})) \xrightarrow{d} N_{MM}(0, \Omega) \tag{D.1}$$

where $vec(\cdot)$ is a vectorization operator, which converts a matrix into a vector, and Ω is an $M \times M$ covariance matrix. In practice, Ω may be estimated by bootstrap. The asymptotic distribution of the test statistic is a sum of weighted chi-square distributions, which is described as

$$\sum_{i=1}^{(M-r^*)^2} \lambda_i \cdot \chi^2_{1,i} \tag{D.2}$$

where $\{\lambda_i\}_{i=1}^{(M-r^*)^2}$ are the weights and $\{\chi^2_{1,i}\}_{i=1}^{(M-r^*)^2}$ are i.i.d. chi-square random variables with degree of freedom one. $\{\lambda_i\}_{i=1}^{(M-r^*)^2}$ are the eigenvalues of $(C \otimes D)'\Omega(C \otimes D)$, where $(C \otimes D)'\Omega(C \otimes D)$ is a $(M-r^*)^2 \times (M-r^*)^2$ matrix. The critical value can be generated by simulation.

B.2 Proof of Proposition 1

Part(a)

Proof. From the property of strategy functions, we know that the mapping from value to bid is increasing. For each type of bidders, the α-th quantile of value must be mapped to the α-th quantile of bid, i.e., $s_k(v_{k,\alpha}) = b_{k,\alpha}$ and $s_l(v_{l,\alpha}) = b_{l,\alpha}$. Then we have $b_{k,\alpha} - b_{l,\alpha} = s_k(v_{k,\alpha}) - s_l(v_{l,\alpha})$. The statement of Proposition 1 is equivalent to that $s_k(v_{k,\alpha}) - s_l(v_{l,\alpha})$ is increasing in α on Q. Since both $s_k(v_{k,\alpha})$ and $s_l(v_{l,\alpha})$ are differentiable with respect to α, we can write the statement of Proposition 1 in terms of derivatives as

$$\frac{\partial s_k(v_{k,\alpha})}{\partial \alpha} - \frac{\partial s_l(v_{l,\alpha})}{\partial \alpha} \geq 0, \quad \forall \alpha \in Q$$

By the chain rule, we have

$$\frac{\partial s_k(v_{k,\alpha})}{\partial \alpha} - \frac{\partial s_l(v_{k,\alpha})}{\partial \alpha} = \frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} \cdot \frac{\partial v_{k,\alpha}}{\partial \alpha} - \frac{\partial s_l(v_{l,\alpha})}{\partial v_{l,\alpha}} \cdot \frac{\partial v_{l,\alpha}}{\partial \alpha}$$

Since type k and l share the same value distribution, we have $v_{k,\alpha} = v_{l,\alpha}$.

$$\frac{\partial s_k(v_{k,\alpha})}{\partial \alpha} - \frac{\partial s_l(v_{k,\alpha})}{\partial \alpha} = (\frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} - \frac{\partial s_l(v_{k,\alpha})}{\partial v_{k,\alpha}}) \cdot \frac{\partial v_{k,\alpha}}{\partial \alpha}$$

Since $v_{k,\alpha}$ is the α-th quantile of type k's value distribution, $\frac{\partial v_{k,\alpha}}{\partial \alpha} > 0$. We only need to show that

$$\frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} - \frac{\partial s_l(v_{k,\alpha})}{\partial v_{k,\alpha}} \geq 0, \quad \forall \alpha \in Q \tag{D.3}$$

Since we do not have the closed-form solutions for s_k and s_l, but have the closed-form solutions of s_k^{-1} and s_l^{-1} instead. We consider rewrite the above inequality in terms of inverse strategy functions. By the chain rule, we have $\frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} \cdot \frac{\partial s_k^{-1}(b_{k,\alpha})}{\partial b_{k,\alpha}} = 1$. Since the mapping from value to bid is increasing, we have both $\frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} > 0$ and $\frac{\partial s_k^{-1}(b_{k,\alpha})}{\partial b_{k,\alpha}} > 0$. Similarly, for type l, we have $\frac{\partial s_l(v_{l,\alpha})}{\partial v_{l,\alpha}} \cdot \frac{\partial s_l^{-1}(b_{l,\alpha})}{\partial b_{l,\alpha}} = 1$, $\frac{\partial s_l(v_{l,\alpha})}{\partial v_{l,\alpha}} > 0$ and $\frac{\partial s_l^{-1}(b_{l,\alpha})}{\partial b_{l,\alpha}} > 0$. Equation (D.3) is equivalent to the following.

$$\frac{\partial s_k^{-1}(b_{k,\alpha})}{\partial b_{k,\alpha}} - \frac{\partial s_l^{-1}(b_{l,\alpha})}{\partial b_{l,\alpha}} \leq 0, \quad \forall \alpha \in Q \tag{D.4}$$

Recall that we have closed-form solutions for both s_k^{-1} and s_l^{-1} in Section 2, as showed in the following.

$$s_k^{-1}(b_{k,\alpha}) = b_{k,\alpha} + \lambda_k^{-1}(H(b_{k,\alpha}))$$

$$s_l^{-1}(b_{l,\alpha}) = b_{l,\alpha} + \lambda_l^{-1}(H(b_{l,\alpha}))$$

Furthermore, we have

$$\frac{\partial s_k^{-1}(b_{k,\alpha})}{\partial b_{k,\alpha}} = 1 + \frac{\partial \lambda_k^{-1}(H(b_{k,\alpha}))}{\partial H(b_{k,\alpha})} \cdot \frac{\partial H(b_{k,\alpha})}{\partial b_{k,\alpha}} \tag{D.5}$$

$$\frac{\partial s_l^{-1}(b_{l,\alpha})}{\partial b_{l,\alpha}} = 1 + \frac{\partial \lambda_l^{-1}(H(b_{l,\alpha}))}{\partial H(b_{l,\alpha})} \cdot \frac{\partial H(b_{l,\alpha})}{\partial b_{l,\alpha}} \tag{D.6}$$

Plug Equation (D.5) and Equation (D.6) into Equation (D.4), then we have

$$\frac{\partial \lambda_k^{-1}(H(b_{k,\alpha}))}{\partial H(b_{k,\alpha})} \cdot \frac{\partial H(b_{k,\alpha})}{\partial b_{k,\alpha}} \leq \frac{\partial \lambda_l^{-1}(H(b_{l,\alpha}))}{\partial H(b_{l,\alpha})} \cdot \frac{\partial H(b_{l,\alpha})}{\partial b_{l,\alpha}}, \quad \forall \alpha \in Q \tag{D.7}$$

By Assumption 2(b), we have $0 \leq \frac{\partial H_k(b_{k,\alpha})}{\partial b_{k,\alpha}} \leq \frac{\partial H_l(b_{l,\alpha})}{\partial b_{l,\alpha}}$. By Assumption 2(a), $u_k(x) = x^{\beta_k}/\beta_k$ and $\lambda_k(x) = u_k(x)/u_k'(x) = x/\beta_k$. As a result, λ_k^{-1} reduces to a linear function, as well as λ_l^{-1}, i.e., $\frac{\partial \lambda_k^{-1}(H_k(b_{k,\alpha}))}{\partial H_k(b_{k,\alpha})} = \beta_k$ and $\frac{\partial \lambda_l^{-1}(H_l(b_{l,\alpha}))}{\partial H_l(b_{l,\alpha})} = \beta_l$. By Assumption 6, we have $0 < \beta_k < \beta_l$. Then $\frac{\partial \lambda_k^{-1}(H_k(b_{k,\alpha}))}{\partial H_k(b_{k,\alpha})} \cdot \frac{\partial H_k(b_{k,\alpha})}{\partial b_{k,\alpha}} \leq \frac{\partial \lambda_l^{-1}(H_l(b_{l,\alpha}))}{\partial H_l(b_{l,\alpha})} \cdot \frac{\partial H_l(b_{l,\alpha})}{\partial b_{l,\alpha}}$. The proof is completed.

□

Part(b)

Proof. Since both s_k and s_l are increasing, the quantiles will thus transfer from values to bids, i.e., $s_k(v_{k,\alpha}) = b_{k,\alpha}$ and $s_l(v_{l,\alpha}) = b_{l,\alpha}$. Then we have $b_{k,\alpha} - b_{l,\alpha} = s_k(v_{k,\alpha}) - s_l(v_{l,\alpha})$. The statement that $T(\alpha)$ is decreasing is equivalent to that $s_k(v_{k,\alpha}) - s_l(v_{l,\alpha})$ is decreasing in α, i.e., $\frac{\partial s_k(v_{k,\alpha})}{\partial \alpha} < \frac{\partial s_l(v_{l,\alpha})}{\partial \alpha}$. For simplicity in writing the proof, the quantile notation α we are using here means $\forall \alpha \in Q$. The quantiles outside of Q are excluded. By Assumption 3(a), we have

$$v_{l,\alpha} = v_{k,\alpha} - c(v_{k,\alpha})$$

The statement that $T(\alpha)$ is decreasing is further equivalent to that

$$\frac{\partial s_k(v_{k,\alpha})}{\partial \alpha} < \frac{\partial s_l(v_{k,\alpha} - c(v_{k,\alpha}))}{\partial \alpha}.$$

$$
\begin{aligned}
& \frac{\partial s_k(v_{k,\alpha})}{\partial \alpha} - \frac{\partial s_l(v_{k,\alpha} - c(v_{k,\alpha}))}{\partial \alpha} \\
= & \frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} \cdot \frac{\partial v_{k,\alpha}}{\partial \alpha} - \frac{\partial s_l(v_{k,\alpha} - c(v_{k,\alpha}))}{\partial v_{k,\alpha}} \cdot \frac{\partial v_{k,\alpha}}{\partial \alpha} \\
= & (\frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} - \frac{\partial s_l(v_{k,\alpha} - c(v_{k,\alpha}))}{\partial v_{k,\alpha}}) \cdot \frac{\partial v_{k,\alpha}}{\partial \alpha}
\end{aligned}
$$

Since $v_{k,\alpha}$ is the α-th quantile of type k's value distribution, $\frac{\partial v_{k,\alpha}}{\partial \alpha} > 0$. Only to show that $\frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} - \frac{\partial s_l(v_{k,\alpha} - c(v_{k,\alpha}))}{\partial v_{k,\alpha}} < 0$. In the following we will decompose $\frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} - \frac{\partial s_l(v_{k,\alpha} - c(v_{k,\alpha}))}{\partial v_{k,\alpha}}$.

$$
\begin{aligned}
& \frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} - \frac{\partial s_l(v_{k,\alpha} - c(v_{k,\alpha}))}{\partial v_{k,\alpha}} \\
= & \frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} - \frac{\partial s_l(v_{k,\alpha} - c(v_{k,\alpha}))}{\partial v_{l,\alpha}} \cdot \frac{\partial v_{l,\alpha}}{\partial v_{k,\alpha}} \\
= & \frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} - \frac{\partial s_l(v_{l,\alpha})}{\partial v_{l,\alpha}} \cdot (1 - \frac{\partial c(v_{k,\alpha})}{\partial v_{k,\alpha}}) \\
= & \frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} - \frac{\partial s_l(v_{l,\alpha})}{\partial v_{l,\alpha}} + \frac{\partial s_l(v_{l,\alpha})}{\partial v_{l,\alpha}} \cdot \frac{\partial c(v_{k,\alpha})}{\partial v_{k,\alpha}} \\
\equiv & A + B
\end{aligned}
$$

where $A \equiv \frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} - \frac{\partial s_l(v_{l,\alpha})}{\partial v_{l,\alpha}}$ and $B \equiv \frac{\partial s_l(v_{l,\alpha})}{\partial v_{l,\alpha}} \cdot \frac{\partial c(v_{k,\alpha})}{\partial v_{k,\alpha}}$. In the following we will show that $B < 0$. By Assumption 3(a), $\frac{\partial c(v_{k,\alpha})}{\partial v_{k,\alpha}} < 0$. Since s_l is type l's strategy function, it must be increasing in value, i.e., $\frac{\partial s_l(v_{l,\alpha})}{\partial v_{l,\alpha}} > 0$. The combination of these two inequalities gives

$$B = \frac{\partial s_l(v_{l,\alpha})}{\partial v_{l,\alpha}} \cdot \frac{\partial c(v_{k,\alpha})}{\partial v_{k,\alpha}} < 0$$

The last step is to show that $A < 0$. From $s_k(v_{k,\alpha}) = b_{k,\alpha}$ and $s_k^{-1}(b_{k,\alpha}) = v_{k,\alpha}$, by the law of derivatives of inverse function, we have $\frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} \cdot \frac{\partial s_k^{-1}(b_{k,\alpha})}{\partial b_{k,\alpha}} = 1$. Since $\frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} > 0$, we have $\frac{\partial s_k^{-1}(b_{k,\alpha})}{\partial b_{k,\alpha}} > 0$. This will also apply to type l. Then $A < 0$ is equivalent to that $\frac{\partial s_k(v_{k,\alpha})}{\partial v_{k,\alpha}} < \frac{\partial s_l(v_{l,\alpha})}{\partial v_{l,\alpha}}$ and further equivalent to that $\frac{\partial s_k^{-1}(b_{k,\alpha})}{\partial b_{k,\alpha}} > \frac{\partial s_l^{-1}(b_{l,\alpha})}{\partial b_{l,\alpha}}$. By Equation (4.5),

$\frac{\partial s_k^{-1}(b_{k,\alpha})}{\partial b_{k,\alpha}}$ and $\frac{\partial s_l^{-1}(b_{l,\alpha})}{\partial b_{l,\alpha}}$ can be explicitly expressed as

$$\frac{\partial s_k^{-1}(b_{k,\alpha})}{\partial b_{k,\alpha}} = 1 + \frac{\partial H_k(b_{k,\alpha})}{\partial b_{k,\alpha}}$$

and

$$\frac{\partial s_l^{-1}(b_{l,\alpha})}{\partial b_{l,\alpha}} = 1 + \frac{\partial H_l(b_{l,\alpha})}{\partial b_{l,\alpha}}$$

By Assumption 3(b), we have $\frac{\partial H_k(b_{k,\alpha})}{\partial b_{k,\alpha}} > \frac{\partial H_l(b_{l,\alpha})}{\partial b_{l,\alpha}}$, i.e., $\frac{\partial s_k^{-1}(b_{k,\alpha})}{\partial b_{k,\alpha}} > \frac{\partial s_l^{-1}(b_{l,\alpha})}{\partial b_{l,\alpha}}$ and $A < 0$. The proof is completed.

□

B.3 Proof of Proposition 4

The goal of this section is to derive the asymptotic properties of the estimator. It is easy to see that both $\hat{B}_{b_2,d_1,d_3}$ and $\hat{B}_{d_1,d_3}$ are $\sqrt{n}$ consistent by empirical CDF/pmf estimator. By Hu (2008), the eigen decomposition function $\phi_E(\cdot)$ is an analytical function. As a result,

$$\hat{B}_{d_1|k} = \phi_E(\hat{B}_{b_2,d_1,d_3}\hat{B}_{d_1,d_3}) \xrightarrow{p} \phi_E(B_{b_2,d_1,d_3}B_{d_1,d_3}) = B_{d_1|k}$$

The inverse of $B_{d_1|k}$ is also consistent.

$$\hat{B}_{d_1|k}^{-1} \xrightarrow{p} B_{d_1|k}^{-1}$$

Since $\hat{\vec{p}}(d_1)$ is $\sqrt{n}$ consistent by empirical pmf estimator, we have that

$$\hat{\vec{p}}(k) = \hat{B}_{d_1|k}^{-1}\hat{\vec{p}}(d_1) \xrightarrow{p} B_{d_1|k}^{-1}\vec{p}(d_1) = \vec{p}(k)$$

Recall that $\hat{G}(b_2|k)$ is obtained by

$$\hat{G}(b_2|k) = \frac{\vec{e}_k \hat{B}_{d_1|k}^{-1} \hat{G}(b_2, d_1)}{\vec{e}_k \hat{\vec{p}}(k)}$$

$\vec{e}_k$ is a deterministic vector. Therefore, $\vec{e}_k \hat{B}_{d_1|k}^{-1} \xrightarrow{p} \vec{e}_k B_{d_1|k}^{-1}$ and $\vec{e}_k \hat{\vec{p}}(k) \xrightarrow{p} \vec{e}_k \vec{p}(k)$. The only component left unconsidered is $\hat{G}(b_2, d_1)$. The joint CDF/pmf is estimated by empirical CDF/pmf estimator with

$$\sqrt{n}(\hat{G}(b_2, d_1) - G(b_2, d_1)) \xrightarrow{d} N(0, V_{b_2,b_2})$$

where V_{b_2,b_2} is the variance for empirical CDF estimator. By Slutsky theorem, we have that

$$\begin{aligned} & \sqrt{n}(\hat{G}(b_2|k) - G(b_2|k)) \\ = & \frac{\vec{e}_k \hat{B}_{d_1|k}^{-1} \sqrt{n}(\hat{G}(b_2, d_1) - G(b_2, d_1))}{\vec{e}_k \hat{\vec{p}}(k)} \\ \xrightarrow{d} & N(0, \frac{\vec{e}_k B_{d_1|k}^{-1} V_{b_2,b_2} (\vec{e}_k B_{d_1|k}^{-1})^T}{(\vec{e}_k \vec{p}(k))^2}) \end{aligned}$$

$$nVar(\hat{G}(b_2|k)) \to \frac{\vec{e}_k B_{d_1|k}^{-1} V_{b_2,b_2} (\vec{e}_k B_{d_1|k}^{-1})^T}{(\vec{e}_k \vec{p}(k))^2} \equiv V_{b_2,b_2,kk}$$

$$nCov(\hat{G}(b_2|k), \hat{G}(b_2'|l)) \to \frac{\vec{e}_k B_{d_1|k}^{-1} V_{b_2,b_2'} (\vec{e}_l B_{d_1|l}^{-1})^T}{\vec{e}_k \vec{p}(k) \vec{e}_l \vec{p}(k)} \equiv V_{b_2,b_2',kl}$$

According to the asymptotic properties of the conditional CDF estimator, we can further derive the asymptotic properties of the conditional inverse CDF estimator by Bahadur's representation. For notation simplicity, denote $b_{k,\alpha} \equiv G_k^{-1}(\alpha)$. By Bahadur's

representation, we have that

$$\hat{G}_k^{-1}(\alpha) = b_{k,\alpha} + \frac{\alpha - \hat{G}_k(b_{k,\alpha})}{g_k(b_{k,\alpha})} + o_p(\frac{1}{\sqrt{n}})$$

The conditional inverse CDF estimator is a $\sqrt{n}$ consistent estimator. Furthermore, $\hat{T}(\alpha) \equiv \hat{G}_k^{-1}(\alpha) - \hat{G}_l^{-1}(\alpha)$ is a root n consistent estimator for $T(\alpha)$.