Local Rank Estimation of Transformation Models with Functional Coefficients

Youngki Shin
Department of Economics
University of Rochester
Email: yshn@troi.cc.rochester.edu

January 13, 2007
(Job Market Paper)

Abstract

In this paper, I consider a nonparametric functional coefficient model with unknown transformation of the dependent variable. This model provides deeper insight into duration analysis since it reveals heterogeneous effects without assuming a strict hazard functional form. I propose a local rank estimation procedure for the model, which converges to a normal distribution at the optimal nonparametric rate. I also extend the model to the case of an interval-valued regressor and show set identification results under different assumptions. The proposed estimator can be easily combined with other methods in existing literature on various censoring and truncation models. An empirical study on the effect of unemployment insurance (UI) benefits is presented to illustrate the local rank estimation method. The estimation results reveal that the effects of UI benefits are heterogeneous across different poverty levels though it is not very significant. It is also found that UI benefits decrease the length of unemployment spell significantly, which suggests new features of labor market behavior that has not been predicted by current theories.

JEL Classification: C13, C14, C41, J64, J65

Keywords: Rank estimation, transformation model, functional coefficients, set identification, unemployment duration, unemployment insurance.

*I am greatly indebted to my advisors Shakeeb Khan and Werner Ploberger for their encouragement and invaluable guidance. I also thank Árpád Ábrahám, Mark Bils, William Hawkins, Jay Hong, Joseph McMurray, Ronni Pavan, Kyoungwon Seo, Hyemi You and seminar participants at University of Rochester for their helpful comments and suggestions. All errors are mine.
1 Introduction

In economics, some interesting variables such as unemployment spells and the lifetimes of capital goods are defined as the length of time before some event. Although there has been a large and growing literature both in empirical and theoretical duration analysis, most research depends on specific hazard functions that induce severe bias in estimation when the specification is not correct. Transformation models, which are usually defined as below, can be applied to reduce bias (see Ridder (1990)):

\[ T(y_i) = x_i' \beta + \varepsilon_i \] (1)

where \( T(\cdot) \) is an unspecified but monotone transformation function and \( \varepsilon_i \) is an error term whose distribution is unknown. Several semiparametric estimation methods and their extensions to various censored and truncated situations have been suggested. See Han (1987), Cavanagh and Sherman (1998), Abrevaya (1999), Khan and Tamer (2007), and Shin (2006b) among others.

In this paper, I consider transformation models with functional coefficients. Although the unspecified transformation function \( T(\cdot) \) renders much flexibility to a duration model, the linear assumption in coefficients might be still restrictive. We relax this limitation by combining it with a functional coefficient model. Thus, the coefficient \( \beta \) is not assumed to be a constant value anymore but an arbitrary function of some covariate \( w_i \). This model gives much more flexibility without suffering from the curse of dimensionality. The usefulness of the suggested model specification is clearer in the case of a specific empirical application. Let us consider the example of unemployment duration analysis. Suppose that we are interested in the effect of unemployment insurance (UI) benefits. If we construct a model following the standard transformation model in (1), we can control other demographic variables only in a linear way. However, it is possible that UI benefits have different effects on unemployment duration according to age, wealth or education levels. The suggested model enables one to estimate this heterogeneous effect, so the researcher can gain deeper insight. If the effects of UI benefits differ with wealth levels, then policy implication from the model can target to specific poverty levels.

For estimation of the model, I propose local rank estimation procedures. The estimation method of functional coefficient models was pioneered by Cleveland et al. (1991) and Hastie and Tibshirani (1993). Their method, which basically minimizes local least squares, can only apply for models in which the functional form of \( T(\cdot) \) is known. This does not fit well for our analysis of duration models without restrictive assumptions on a hazard function. Instead, I adopt a rank estimation technique, which is an intuitive approach to unspecified monotone transformation models. After unknown transformation, we lose information about the magnitude of \( y_i \), but still have valid information of its rank. The local rank estimators proposed in this paper exploit rank information of \( y_i \) and
maximize the local rank correlation of $y_i$ and $x'_i \beta(w_i)$. Localization can be proceeded by the usual nonparametric estimation methods such as kernel or series estimations. Specifically, I show that the local maximum rank correlation (LMRC) estimator, under standard regularity conditions, satisfies pointwise consistency and converges to a normal distribution at each point $w_i$ at the optimal nonparametric convergence rate.

Extensions of the LMRC estimator are also considered under an interval-valued covariate. In empirical applications, some important variables are observed only as interval values for various reasons. For instance, a survey may ask for wealth levels by brackets instead exact values. Other variables that respondents are reluctant to reveal might be asked as interval values for confidentiality reasons. If we have such an interval-valued regressor, the point identification usually fails. Manski and Tamer (2002) proposed sufficient conditions for a set identification that might be informative in certain cases and suggested a modified maximum score estimator. Magnac and Maurin (2005) proposed a different estimation method under different assumptions. Both papers, however, restrict attention to the binary choice model. Honoré and Lleras-Muney (2005) derive bound identification results of the accelerate failure time (AFT) model specification with competing risks. This paper derives bounds for underlying parameters in transformation models. In fact, the LMRC estimation can be applied to any of the generalized regression models in Han (1987), so parameters in the binary choice model with an interval regressor can also be estimated by the LMRC procedure. I show that combining two methods gives sharper bounds in any case of known transformation function. I also extend the rank estimation of $T(\cdot)$ in Chen (2002) under the set valued $\beta$. It is shown that the transformation function $T(\cdot)$ fails to be identified as a function but is identified as a correspondence.

As an empirical illustration of the local rank estimation, I estimate the heterogeneous effects of UI benefits on unemployment duration. A plausible conjecture is that UI benefits may effect individuals differently according to their wealth levels. There have been many empirical works related to unemployment duration but most of them depend on strict assumptions. For instance, the reduced form estimation in Meyer (1990) assumed proportional and additive hazard functions. Structural estimation based on labor search theory also depends heavily on various functional forms of the model and assumptions on the distribution of exogenous variables (see Eckstein and van den Berg (2003)). Furthermore, to my knowledge, there has been no work that estimates heterogeneous effects of UI benefits across different wealth levels. Thus, the goals of this empirical study are: (i) to estimate UI effects using as weak assumptions as possible and (ii) to check if there are heterogeneous effects of UI benefits.

The rest of this paper is organized as follows: In section 2, I introduce the model and estimation method. Large sample properties are also proposed in the section. In section 3, I consider some
extensions including a model with interval data. The finite sample properties are investigated by means of Monte Carlo simulations in section 4. In section 5, an empirical study of unemployment duration is presented to illustrate the usefulness of the local rank estimation. Section 6 concludes and suggests future research areas. Technical proofs are presented in the appendix unless helpful in understanding arguments in the text.

2 Estimation Procedure and Asymptotic Properties

In this section I propose a new estimation procedure and derive its asymptotic properties. I first introduce transformation models with functional coefficients in detail. Suppose that we have i.i.d. observations \((y_i, x_i', w_i)\) where \(y_i\) and \(w_i\) are scalar random variables and \(x_i\) is a \((k \times 1)\)-dimensional random vector. These variables satisfy the following functional restriction:

\[
T(y_i) = x_i' \beta(w_i) + \varepsilon_i
\]  

(2)

where \(T(\cdot)\) is an unspecified monotone function, \(\beta(\cdot)\) is an unspecified vector-valued function depending on a specific covariate \(w_i\), and \(\varepsilon_i\) is an error term with unknown distribution. Thus, a special variable \(w_i\) affects the coefficients of \(x_i\) through the unknown function \(\beta(\cdot)\). Note that \(w_i\) can be easily generalized to a \(d_w\)-dimensional random vector but we keep the scalar assumption to avoid unnecessarily complex notation.

The right hand side of the model specification in (2) is general enough to include various generalizations of linear models. The identity transformation \(T(y_i) = y_i\) includes generalized linear models, dynamic generalized linear models, and generalized additive models as special cases. See Hastie and Tibshirani (1993) and Christopeit and Hoderlein (2006) for more examples.\(^1\) Although a large volume of literature has been devoted to generalizing linear models, there is to my knowledge no estimation method allowing unknown transformation in dependent variables. In fact, this makes the model suitable for duration analyses with flexible hazard functions, circumventing the curse of dimensionality. It may be also applied to any other regression analysis with less restrictive assumptions on functional structures.

We first focus on estimating \(\beta(w_i)\). Estimation of the transformation function \(T(\cdot)\) is also discussed in the later part of this section. If \(\beta\) is assumed to be a parameter that has a constant value as in standard transformation models, then we can just apply the maximum rank correlation estimation procedure proposed by Han (1987). When \(\beta\) is a function of \(w_i\), that procedure needs to be combined with a nonparametric estimation technique. For this purpose, I assume \(\beta(w_i)\) is

\(^1\)As a related literature, Horowitz (2001) proposed a generalized additive model with an unknown link function. Although it covers various model specifications, it does not contain transformation models with functional coefficients.
an arbitrary but smooth function of \( w_i \). Then we can estimate the function \( \beta(w_i) \) by maximizing the conditional rank correlation defined as:

\[
Q(\beta(w)) = E \left[ 1(y_i > y_j)1\left(x'_i \beta(w_i) > x'_j \beta(w_j)\right) + 1(y_i < y_j)1\left(x'_i \beta(w_i) < x'_j \beta(w_j)\right) | w_i = w_j = w \right]
\]

(3)

where \( 1(\cdot) \) is an indicator function. For a given \( w_0 \), I define the LMRC estimator using a sample analogue of the objective function \( Q(\beta(w_0)) \) as below:

\[
\hat{\beta}(w_0) = \arg \max Q_n(\beta(w_0)) = \arg \max \frac{1}{n(n-1)} \sum_{i \neq j} m_{ij}(\beta(w_0)) K_h(w_i - w_0) K_h(w_j - w_0)
\]

(4)

(5)

where \( m_{ij}(\beta(w_0)) = 1(y_i > y_j)1\left(x'_i \beta(w_0) > x'_j \beta(w_0)\right) \) and \( K_h(w) = h^{-1}K(w/h) \) for a kernel function \( K(\cdot) \) and bandwidth \( h \). Note that the sample objective function is a second order U-process.

The LMRC estimation method is not restricted to transformation models with functional coefficients. Since the MRC estimator is applicable to any generalized regression model defined by

\[
y_i = D \cdot F(x'_i \beta, \varepsilon_i)
\]

(6)

with non-degenerate monotone function \( D(\cdot) \) and \( F(\cdot, \cdot) \), the proposed estimator is also applicable to any local regression of (6) with unknown function \( \beta(\cdot) \). Besides transformation models, it includes Box–Cox transformation models, censored regression models, and binary choice models as special cases. Generally, the LMRC estimator is valid for any model that satisfies the following key identification condition:

\[
P(y_i > y_j|x_i, x_j, w_0) > P(y_i < y_j|x_i, x_j, w_0) \iff x'_i \beta(w_0) > x'_j \beta(w_0).
\]

(7)

Other rank estimation and nonparametric estimation procedures can also be considered. Cavanagh and Sherman (1998) proposed the class of monotone rank estimators (MRE) in which the sample objective function is defined as

\[
\tilde{Q}_n(\beta) = \frac{1}{n(n-1)} \sum_{i \neq j} M(y_i)1(x'_i \beta > x'_j \beta)
\]

(8)
where \( M(y_i) \) is any function satisfying the condition that \( E[M(y_i) | x_i] \) is a non–degenerate increasing function of \( x'_i \beta_0 \). Thus \( M(\cdot) \) can be either a deterministic function or the rank function that would induce the MRC estimator. In transformation models, it is clear that \( E[M(y_i) | x_i] \) is increasing in \( x'_i \beta_0 \) if we set \( M(y) = y \). Considering computational efficiency, we may adopt a MRE objective function instead of using \( m_{ij}(\beta(w_0)) \) in (5). At the same time, we can apply different nonparametric estimation procedures. Note that our objective function is in the form of conditional expectation, and that is why the kernel estimation procedure is combined. Without much difficulty, other nonparametric technique such as local polynomial and series estimation methods can be adopted.

Now we turn our attention to the asymptotic properties of the LMRC estimator. To prove piecewise consistency, we need the following regularity conditions:

A1. The observations \( z_i = (y_i, x'_i, w_i) \) are i.i.d. and the regressors \((x'_i, w_i)\) are independent of the error terms \( \varepsilon_i \).

A2. The domain of \( w_i \), denoted by \( \mathcal{W} \), is a compact subset of \( \mathbb{R} \).

A3. The \( k \times 1 \) dimensional vector valued function \( \beta(\cdot) \) is differentiable.

A4. The objective function \( Q(\beta(w_0)) \) is continuous at \( \beta_0(w_0) \).

A5. The support of \( x_i \) is not contained in a proper linear subspace of \( \mathbb{R}^d \) and the \( d \)th component of \( x_i \) has an everywhere positive Lebesgue density conditional on the remaining components.

A6. The bandwidth \( h_n \) satisfies the following:

\[
\begin{align*}
(a) \quad & \lim_{n \to +\infty} h_n = 0 \\
(b) \quad & \lim_{n \to +\infty} n h_k = \infty \text{ for } k \leq 4. \\
(c) \quad & \lim_{n \to +\infty} n h_5 = \lambda \text{ for } 0 \leq \lambda < \infty.
\end{align*}
\]

A7. The kernel function \( K(x) \) satisfies the following:

\[
\begin{align*}
(a) \quad & \int K(x) \, dx = 1. \\
(b) \quad & \int |K(x)| \, dx < \infty. \\
(c) \quad & \int xK(x) \, dx = 0.
\end{align*}
\]
All assumptions are standard in semiparametric and nonparametric literature. Condition A1 assumes a data generating process. Though it assumes \((x'_i, w_i)\) to be independent of \(\varepsilon_i\), the model can be easily extended to certain conditional heteroscedastic cases by using the generalized regression form in (6). More generally, we can allow arbitrary conditional heteroscedasticity by conducting a two stage estimation method as in Khan (2001). Conditions A2–A3 assures that the parameter space of \(\beta (w_0)\) for a fixed \(w_0\) is compact. Condition A4 is the usual continuity assumption, and Condition A5 is a support condition necessary for the point identification, which is also used in other semiparametric estimation literature (see Han (1987) and Manski (1985)). Condition A6 and A7 are the standard assumptions on the bandwidth and a kernel function.

Before looking at the consistency result, we should keep in mind that the MRC estimator identifies the parameter only up to scale, so a scale normalization is required. In this paper, the \(d\)-th component of \(\beta (w)\) is assumed to be 1. Instead of introducing extra notation, I just keep the notation \((w)\) but it actually stands for \(d(w) = d(w)\) where \(d(w)\) is the first \((d-1)\) elements of \(\beta (w)\). The pointwise consistency result is stated in the next theorem whose proof is in the appendix.

**Theorem 2.1** Suppose that Assumptions A1–A7 holds. Then, for given \(w_0\),

\[
\tilde{\beta} (w_0) - \beta (w_0) = o_p (1).
\]

To prove asymptotic normality, we need the following additional regularity conditions:

A8. The true parameter \(\beta_0 (w_0)\) is an interior point of its parameter set.

A9. Let \(\tau (z, \beta)\) be

\[
\tau (z, \beta) = E \left[ m_{ij} (\beta) K_h (w_i - w_0) K_h (w_j - w_0) | z_i = z \right] + E \left[ m_{ij} (\beta) K_h (w_i - w_0) K_h (w_j - w_0) | z_j = z \right]. \tag{9}
\]

Let \(\mathcal{N}\) be a neighborhood of \(\beta_0\). Then the following holds:

(a) For each \(z \in \mathcal{Z}\), all mixed second partial derivatives of \(\tau (z, \cdot)\) exist on \(\mathcal{N}\).

(b) There is an integrable function \(M (z)\) such that for all \(z \in \mathcal{Z}\) and \(\beta\) in \(\mathcal{N}\)

\[
||\nabla_2 \tau (z, \beta) - \nabla_2 \tau (z, \beta_0)|| \leq M (z) |\beta - \beta_0|.
\]

(c) \(E |\nabla_1 \tau (z, \beta)|^2 < \infty\).
(d) \( E |\nabla_2 \tau (\cdot, \beta) | < \infty \).

(e) The matrix \( E \nabla_2 \tau (\cdot, \theta) \) is negative definite.

Now we are ready to characterize the asymptotic distribution of the LMRC estimator. The following theorem shows that it converges to a normal distribution. With the bandwidth condition in A6, it attains the optimal nonparametric convergence rate. I leave the proof and the explicit expression of the bias term \( B (\theta_0) \) in the appendix.

**Theorem 2.2** Suppose that Assumptions A1–A9 holds. Then, for a given \( w_0 \),

\[
\sqrt{n} h \left( \beta - \beta_0 + 2 \nabla_1 B (\beta_0) \right) \xrightarrow{d} N (0, V^{-1} \Delta V^{-1})
\]

where \( 2V = E [\nabla_2 \tau (z, \beta_0)] \) and \( \Delta = E \left[ (\nabla_1 \tau (z, \beta_0) + 2 \nabla_1 B (\beta_0)) (\nabla_1 \tau (z, \beta_0) + 2 \nabla_1 B (\beta_0))' \right] \).

The variance matrix can be consistently estimated by a numerical derivative form (see Sherman (1993) and Khan and Tamer (2007)) or a kernel method. We may also adopt resampling methods in case of a small sample.

I conclude this section by discussing how to estimate the transformation function \( T (\cdot) \). Existing literature provides several estimation methods. See Chen (2002), Horowitz (1996), Klein (2002), and Ye and Duan (1997) among others. In this paper, I follow the rank estimation method in Chen (2002). Since this estimator does not involve any smoothing parameter, it satisfies the \( \sqrt{n} \)-consistency and shows good finite–sample properties even in boundary areas.

Suppose that we know the true parameter \( \beta_0 \) for each value of \( w \), which can be consistently estimated by the LMRC estimator. For location normalization, we assume that \( T_0 (y_0) = 0 \). Let \( Z_i = x'_i \beta_0 (w_i) \) for each observation \( i \). Then it is clear that the key identification condition in Chen (2002) holds:

\[
E [d_{iy} - d_{jy_0} | x_i, x_j, w_i, w_j] \geq 0 \iff Z_i - Z_j \geq T_0 (y)
\]

where \( d_{iy} = 1 (y_i \geq y) \) and \( d_{jy_0} = 1 (y_j \geq y_0) \) following his notation. Thus, we may apply the rank estimation method and estimate the function \( T (\cdot) \) by maximizing the objective function as follows

\[
\tilde{\tau} (y) = \arg \max_{T} \Gamma_n (y, T, \beta_0)
\]

\[
= \arg \max_{T} \frac{1}{n (n - 1)} \sum_{i \neq j} (d_{iy} - d_{jy_0}) 1 (Z_i - Z_j \geq T).
\]

As noted earlier, we can use the LMRC estimator \( \tilde{\beta} \) that is consistently estimated in the first step. An extension to random censoring and asymptotic properties follows directly from Chen (2002).
3 Extension: Set Identification with Interval Data

In this section I consider an extension of the proposed estimator to interval data. As we have seen in the introduction, it is common for data sets to contain some interval variables. Point identification usually fails in such cases and the LMRC estimation is no exception. The parameters, however, can be identified as a set under mild regularity conditions, which might be small enough to have meaningful interpretation. Recently, Manski and Tamer (2002) and Magnac and Maurin (2005) proposed different set identification results for the binary choice model under different regularity conditions. Following the approach of Manski and Tamer (2002), I show that parameters in the transformation model can be identified up to sets. Since this result holds for all generalized regression models in Han (1987), it could be seen as a generalization that includes those results in the binary choice model as specific cases. I also investigate set identification properties of the rank estimation in Chen (2002) when the coefficient in the first stage has a set value. Finally, I compare those identified sets and suggest a way to get sharper bounds under certain conditions.

First consider the transformation model that one of the dependent variables, say $v_i$, is observed as a set value. Formally, it can be stated as $P(v_{i0} \leq v_i \leq v_{i1}) = 1$. In addition, the following conditions hold:

S1. The observations $(y_i, x_i, v_{i0}, v_{i1})$ are i.i.d. and independent of the error term $\varepsilon_i$.

S2. $P(\varepsilon|x,v,v_{i0},v_{i1}) = P(\varepsilon|x,v)$.

Condition S1 assumes a data generating process similar to the Condition A1. Note the difference between this condition and the quantile–independence condition of SBR-1 in Manski and Tamer (2002). We cannot say one is more strict than the other in general. The former is stronger in the sense that it requires independence conditions for all quantiles, but weaker at the same time in that it does not require the specified quantile value for location normalization. As we will see later, combining these two different conditions yields sharper set identification results. Condition S2 comes directly from the condition SBR-2 in Manski and Tamer (2002). We also need a scale normalization, so that the coefficient of $v_i$ is assumed to be 1. A key condition for an identified set is stated in the next proposition, whose proof is in the appendix.

**Proposition 3.1** Suppose that Assumptions S1–S2 hold. Let $z_{ij} = (x_i, v_{i0}, v_{i1}) \otimes (x_j, v_{j0}, v_{j1})$ and let $V_1(b)$ be

$$V_1(b) = \{z_{ij} : \Delta x_{ij}\beta + \Delta v_{i0} > 0 \geq \Delta x_{ij}b + \Delta v_{10} \cup \Delta x_{ij}b + \Delta v_{01} > 0 \geq \Delta x_{ij}\beta + \Delta v_{01}\}. \quad (10)$$

Without loss of generality, we assume that the coefficients are now assumed to be a constant. The result can be immediately applied to the functional coefficients case with appropriate localization with respect to $w$. 

9
Then $\beta$ is identified relative to $b$ if and only if $P (V_1 (b)) > 0$.

To estimate the identified set, I consider the following modified maximum rank correlation estimator. This method is very similar to the modified maximum score estimator. Let

$\lambda_1 (z_{ij}) = 1 (P (y_i > y_j|z_{ij}) > P (y_i < y_j|z_{ij}))$ and the objective function $Q_1 (b)$ be defined as

$$Q_1 (b) = E \{1 (y_i > y_j) - 1 (y_i < y_j)\} \times \{\lambda_1 (z_{ij}) \operatorname{sgn} (\Delta x_{ij}' b + \Delta v_{10}) + (1 - \lambda_1 (z_{ij})) \operatorname{sgn} (\Delta x_{ij}' b + \Delta v_{01})\}$$

(11)

Then an identified set $B_1 = \{b \in B : P (V_1 (b)) = 0\}$ is the set of all maximizers of $Q_1 (b)$. This result is summarized in the next lemma whose proof is also in the appendix.

**Lemma 3.1** Suppose that Assumptions S1–S2 hold. Then $Q_1 (b)$ is maximized at $b \in B_1$ and $Q_1 (b) > Q_1 (b^*)$ for all $b^* \in B_1^c$ if $P (P (y_i > y_j|z_{ij}) = P (y_i < y_j|z_{ij})) = 0$.

Note that the function in (11) becomes the usual objective function of the MRC estimation if the covariate $v_i$ is not observed as an interval value but as an exact point, i.e. $\Delta v_{10} = \Delta v_{01} = \Delta v_{ij}$. For estimation, we can apply the analogy principle as in Manski and Tamer (2002), and the consistency result also comes directly from their paper.

Now we turn our attention to set identification of the transformation function, given set-valued parameters $\beta_0$. As noted in Chen (2002), we need a location normalization for estimating the function. Here I use a different location normalization from his for consistency with subsequent proposition. I add the following median restriction on the error term:

S3. $\operatorname{med} (\varepsilon) = 0$

The combined conditions S1 and S3 are now stronger than the assumption SRB-1. We can now estimate the unknown transformation function which was assumed to be known in the binary choice model. In the next proposition, I state a set identification result of the transformation function itself. I leave its proof in the appendix.

**Proposition 3.2** Suppose that Assumptions S1–S3 hold. Let $B_1$ be given. Define the following
notation:

\[ w_m = \min_{b \in B_1} (x_i b + v_{i0}) - \max_{b \in B_1} (x_j b + v_{j1}) \]  
(12)

\[ w_M = \max_{b \in B_1} (x_i b + v_{i1}) - \min_{b \in B_1} (x_j b + v_{j0}) \]  
(13)

\[ T_m = \min_{b \in B_1} (\text{med} \ (x_i b + v_0)) \]  
(14)

\[ T_M = \max_{b \in B_1} (\text{med} \ (x_i b + v_1)) \]  
(15)

Also, for a fixed \( y \), let \( V_2 (T(y)) \) be

\[ V_2 (T(y)) = \{ z_{ij} : (T_0 (y) - T_m < w_m \leq T (y) - T_M) \right. \left. \cup (T (y) - T_m < w_m \leq T_0 (y) - T_M) \right. \} \]  
(16)

Then \( T_0 (y) \) is identified relative to \( T (y) \) if and only if \( P (V_2 (T(y))) > 0 \).

This proposition presents sufficient conditions for a set to be identified. Again, I suggest a modified rank estimation method. Let \( \lambda_2 (z_{ij}) \) be

\[ \lambda_2 (z_{ij}) = 1 (P (y_i > y|z_{ij}) > P (y_j < y|z_{ij})) \]  

I next define the objective function as follows

\[ Q_2 (T(y)) = E \left[ 1 (y_i \geq y) - 1 (y_j \geq y_0) \{ \lambda_2 (z_{ij}) \text{sgn} (w_M - T (y) + T_M) \right. \left. + (1 - \lambda_2 (z_{ij}) \text{sgn} (w_m - T (y) + T_m)) \right] \]  
(17)

Then an identified set \( B_{2y} = \{ T (y) \in R : P (V_2 (T(y))) = 0 \} \) for each \( y \) is the set of all maximizers of \( Q_2 (T(y)) \). This result is shown in the next lemma, whose proof is in the appendix.

Lemma 3.2 Suppose that Assumptions S1–S3 hold. Then \( Q_2 (T(y)) \) is maximized at \( T (y) \in B_{2y} \) and \( Q_2 (T(y)) > Q_2 (T^* (y)) \) for all \( T^* (y) \in B_{2y}^c \).

Collecting those sets for all \( y \) values on the support, we have a correspondence instead of a function. In empirical settings, we do not know the true parameter values of \( \beta_0 \) in advance, so we estimate them in the first stage using the method suggested earlier in this section. Then, we conduct the second stage estimation by plugging in those values into the objective function above. The same analogy principle applies for estimation.
Finally, I consider the case in which we know the lower and upper bounds of the transformation function but do not know its exact value. What we are interested in estimating is the parameter value $\beta$. In this case we need the quantile-independence condition, instead of the condition S1. Without loss of generality, I assume median independence as follows:

$$S_1 \med(\varepsilon_i | x_i, v_i) = 0.$$  

I characterize an identified set using information about the bounds of the transformation function in the next proposition, whose proof is in the appendix.

**Proposition 3.3** Suppose that Assumptions S1', S2 hold. Let $T_0(y) \in [T_m(y), T_M(y)]$ and $d(y) = T_M(y) - T_m(y)$. Let $V_3(b)$ be

$$V_3(b) = \left\{ (x, v_0, v_1) : \begin{array}{l} (x' \beta + v_0 > 0 \geq -d(y) \geq x'b + v_1) \\ \cup (x' \beta + v_1 \leq 0 \leq d(y) < x'b + v_0) \end{array} \right\}$$

Then $\beta$ is identified relative to $b$ if and only if $P(V_3(b)) > 0$.

Small changes in the modified maximum score estimation method give us a way to estimate the set. First, we need some additional notations:

$$P_M = P(T_M(y) > 0 | x, v_0, v_1),$$
$$P_m = P(T_m(y) > 0 | x, v_0, v_1),$$
$$\lambda_{3M} = 1 \left( P_M > \frac{1}{2} \right),$$
$$\lambda_{3m} = \left( P_m > \frac{1}{2} \right).$$

I next define the appropriate objective function:

$$Q_3(b) = E \left[ \begin{array}{c} \left( 1(T_M(y) > 0) - \frac{1}{2} \right) (\lambda_{3M} \sgn(x'b + v_1 + d(y)) + (1 - \lambda_{3M}) \sgn(x'b + v_0 - d(y))) \\ + \left( 1(T_m(y) > 0) - \frac{1}{2} \right) (\lambda_{3m} \sgn(x'b + v_1 + d(y)) + (1 - \lambda_{3m}) \sgn(x'b + v_0 - d(y))) \end{array} \right]$$

In fact, this is a generalization of the objective function used in the binary choice model to any transformation model with known bounds. An identified set $B_3 = \{b \in B : P(V_3(b)) = 0 \}$ is a
collection of all maximizers of (19). This is summarized in the following lemma.

**Lemma 3.3** Suppose that Assumptions S1’ and S2 hold. Then \( Q_3(b) \) is maximized at \( b \in B_3 \) and \( Q_3(b) > Q_3(b^*) \) for all \( b^* \in B_3^c \).

As before, we may apply the analogy principle for estimation. To show its consistency, we can adopt directly the proof of Proposition 3 in Manski and Tamer (2002). For all three estimators above, we can conduct inference procedures by using the subsampling methods in Chernozhukov, Hong, and Tamer (2004) and Shaikh (2006).

Finally, I compare the identified sets \( B_1 \) and \( B_3 \) that are derived under different conditions. Since parameters are identified only up to set, it is natural to ask which method gives tighter bounds when both sets of sufficient conditions are satisfied. It turns out that this depends on how much we know about the bounds of the transformation function. I state the formal result in the following corollary.

**Corollary 3.1** Suppose that Assumptions S1, S1’ and S2 hold. For identified sets \( B_1 \) and \( B_3 \), the following is true:

(i) If \( d(y) \) converges to 0, then there exists a case such that \( B_1 \nsubseteq B_3 \) and \( B_1 \nsubseteq B_3 \).

(ii) If \( d(y) \) diverges to +\(\infty\), then \( B_1 \subseteq B_3 \).

The first result implies that combining two sets can induce tighter bounds when we know the functional form of the transformation. This is an intuitive result in the sense that both estimators use different information, and combining them gives us more information about the underlying model. The second result says that \( B_1 \) works better if we do not have any information about the bounds of the transformation function. In that case, \( B_3 \) does not distinguish any of the subsets in the parameter space.

4 Monte Carlo Simulation

In this section I investigate small sample properties of the proposed estimator by conducting a Monte Carlo simulation study. The base design is a transformation model with three covariates. One covariate \( w_i \) affects the model only through the functional coefficient. Specifically, I use the following simulation designs:

\[
T(y_i) = x_{1i} + x_{2i}\beta(w_i) + \epsilon_i
\]
Table 1: Simulations Results of Transformation Models with Functional Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Linear Transformation</th>
<th>Logarithmic Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IMSE</td>
<td>IMAE</td>
</tr>
<tr>
<td>Linear</td>
<td>0.0564</td>
<td>0.1827</td>
</tr>
<tr>
<td>Quadratic</td>
<td>0.0544</td>
<td>0.1705</td>
</tr>
<tr>
<td>Cubic</td>
<td>0.0573</td>
<td>0.1687</td>
</tr>
</tbody>
</table>

where \( x_{1i} \) and \( x_{2i} \) are random variables following the standard normal distribution and the chi-square distribution of order 3, respectively; \( w_i \) is distributed uniformly on the interval \([-1, 1]\); and \( \varepsilon_i \) follows the normal distribution \( N(0, 0.5) \). I consider different functional forms of \( T(\cdot) \) and \( \beta(\cdot) \). For the transformation function, I use the identity function \( T(y) = y \) and the logarithmic function \( T(y) = \log y \). For each transformation function I consider the following three designs of \( \beta(w) \):

Design 1. Linear: \( \beta(w) = w \).
Design 2. Quadratic: \( \beta(w) = w^2 \).
Design 3. Cubic: \( \beta(w) = w^3 \).

Figure 1 shows the simulation results. For each design I conducted 1000 replications with a sample size of 400. The function \( \beta(\cdot) \) was estimated at 101 equispaced points on the interval \([-1, 1]\). The grid search method was used for estimation, and the graph in Figure 1 is the average value of the estimated functions. The solid lines stand for the true function values. I use the Epanechnikov kernel and the optimal bandwidth \( sn^{-1/5} \), where the constant \( s \) was determined by the rule-of-thumb approach in Silverman (1986).

Overall, the LMRC estimator shows good finite sample performance. It remains stable regardless of different transformation function types. Also, it approximates different forms of functional coefficients \( \beta(w) \) very well though it is a bit biased at the boundary—a common problem in kernel estimation. Since the LMRC estimator does not restrict the choice of the specific nonparametric estimation method, we may adopt different methods for different research purposes. I also report the integrated mean squared errors (IMSE) and the integrated mean absolute errors (IMAE) for each design in Table 1. It confirms that the LMRC estimator is stable for different functional shapes.

Consequently, the results from my simulation study indicate that the LMRC estimator introduced in this paper performs well in finite samples with unspecified transformation functions and various types of functional coefficients. Thus, it can be applied to empirical work with very flexible functional restrictions, which I illustrate in the following section.
Figure 1: Simulations Results
5 Empirical Illustration: Analysis of UI Benefits

Unemployment duration has been a topic of extensive research both in theoretic and empirical economics. After the seminal work of Mortensen (1977), various kinds of search theory have been developed to explain labor market phenomena in addition to the length of unemployment such as the coexistence of unemployed workers and unfilled vacancies, wage heterogeneity among homogeneous workers, and interactions between wages and turnover. See Rogerson, Shimer, and Wright (2005) and references therein for details. At the same time, much empirical research has been conducted to test existing theories or to find new stylized facts. In this section, I analyze the heterogeneous effects of UI benefits on unemployment duration across different poverty levels by using the proposed model.

While the number of existing studies is large and growing, most existing models require strict restrictions on the functional forms and/or the distributions of exogenous variables. Meyer (1990) investigated the effects of UI benefits by applying the proportional hazard (PH) and the additive hazard models, and concluded that high benefits decreased the hazard rate, increasing unemployment duration. Han and Hausman (1990) used a mixed proportional hazard (MPH) model with competing risks and rejected the hypothesis of the monotone hazard rate usually assumed or predicted by simple models of labor market behavior. Although (mixed) proportional hazard model approaches have advantages over structural estimation methods (namely, that it can separate policy effects from those caused by individual characteristics), it has been criticized for its assumption that proportional hazard rates across different individuals do not depend on time. Even simple structural model assumptions induce very non–proportional hazard rate expressions.

Recently, Harding (2006) estimated the effect of UI benefits by using the transformation model. Using data from the late 1990s, he showed that those benefits in fact decreased the length of unemployment and that estimations based on more restrictive assumptions such as the accelerated failure time (AFT) model may induce different and possibly wrong conclusions. I keep the advantages of the transformation model approach, such that it allows non-proportional hazard rates and unobserved heterogeneity. In addition, my model allows functional coefficients and a covariate–dependent censoring variable.

5.1 Data

This analysis is based on the data set from Needels, Corson, and Nicholson (2001) submitted to the U.S. Department of Labor and also used by Harding (2006). The data set is based on

---

3Eckstein and van den Berg (2003) surveyed empirical research in labor search models focusing on the structural estimation framework.
Table 2: Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment duration</td>
<td>48.90</td>
<td>45.05</td>
<td>0.14</td>
<td>185.71</td>
</tr>
<tr>
<td>Censor (censored=0)</td>
<td>0.79</td>
<td>0.41</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Age</td>
<td>39.75</td>
<td>11.90</td>
<td>17</td>
<td>83</td>
</tr>
<tr>
<td>Sex (male=1)</td>
<td>0.56</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Race (white=1)</td>
<td>0.66</td>
<td>0.47</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Number of dependents</td>
<td>1.25</td>
<td>1.43</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>High school dropout</td>
<td>0.17</td>
<td>0.38</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Recall</td>
<td>0.44</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>State unemployment rate</td>
<td>4.62</td>
<td>0.97</td>
<td>2.4</td>
<td>6.4</td>
</tr>
<tr>
<td>Log UI benefit level</td>
<td>5.27</td>
<td>0.45</td>
<td>2.71</td>
<td>6.02</td>
</tr>
<tr>
<td>Log pre–UI wage</td>
<td>6.09</td>
<td>0.68</td>
<td>1.79</td>
<td>12.20</td>
</tr>
<tr>
<td>Poverty level</td>
<td>1.46</td>
<td>0.89</td>
<td>0.05</td>
<td>3.98</td>
</tr>
</tbody>
</table>

Data Source: Needels, Corson, and Nicholson (2001), n=2736.

individual–level surveys of the unemployed who received unemployment insurance between 1998 and 2001. Different from some household surveys, the sample is generated from administrative records, so key variables such as unemployment durations and UI benefits do not involve large degrees of measurement error, a point emphasized by Meyer (1990). The data contain more than 600 variables. As a result, the data set includes very substantive information of the UI recipients, such as non–UI benefits during the unemployed period, partner’s working status, participation in unemployment training, etc. Among them, I choose 13 interesting variables and transform them into 12 variables.

Table 2 summarizes the basic descriptive statistics of my sample. The original data contain 3907 observations but I exclude those with a missing variable and outliers leaving a total of 2736 observations to be analyzed. Most variables are self–explanatory, so I will just explain how the poverty level is constructed. They first sum up all other annualized household income and UI benefits during the period of unemployment. Then the poverty level is calculated by dividing those sums by poverty lines from the U.S. Census Bureau’s Statistical Abstract of the United States (2000), which take into account of the number of people per household.

All individual in the data set received UI benefits. Most are white workers with little education. 1065 observations (38.9%) are below or equal to poverty line. In addition, the distribution of poverty level is very right skewed and many observations are close to the poverty line, so this sample represents the poorest economic class. At the time of interview, 570 observations (20.8%) were still unemployed and recorded as censored. It is highly probable that the censoring variable
may depend on other explanatory variables, so I modify our objective function to adopt the partial rank estimation (PRE) procedure in Khan and Tamer (2007), which allows arbitrary covariate-dependent censoring, as will be seen in the next subsection. For UI benefits, I use weekly UI benefit amounts after taking the log transformation. Pre–UI wage is also controlled, along with other demographic variables, as in Meyer (1990).

It is noteworthy that the sample is collected during the economic boom of the late 1990s. The U.S. unemployment rate in 2000 was 4.0%—the lowest level in the past 30 years. Thus, we should keep in mind that the estimated results reflect a strong boom economy and that we need additional data to see the effects in recession period. In the next subsection, we deal with the details of estimation procedures and present the results and interpretation.

5.2 Results

I consider a regression model with the 10 explanatory variables listed in Table 2. For scale normalization, I set the coefficient of age as 1. It is well known that age has positive effect on unemployment duration. Thus, this normalization will give us the right directions of other coefficient values. To use information contained in censored observations, we modify the model slightly, as below:

\[
T(y_i) = \min (x_{A,i} + x'_{B,i} \beta_B (w_i) + \varepsilon_i, c_i) \\
d_i = 1 \{x_{A,i} + x'_{B,i} \beta_B (w_i) + \varepsilon_i \leq c_i\}
\]

where \( y_i \) is the length of unemployment, \( x_{A,i} \) is an age, \( x_{B,i} \) is an \((8 \times 1)\)-dimensional vector of the remaining control variables, and \( w_i \) is the poverty level. The variable \( c_i \) stands for random censoring that may depend on regressors in an arbitrary way, and \( d_i \) is a variable that indicates if \( y_i \) is censored or not. Note that we do not impose any specific error distribution in the model. Thus, it is robust to any heteroskedastic features if they are orthogonal to the regressors. Especially, any endogeneity caused by \( w_i \) does not matter since only the orthogonal part of it remains in the error term after taking functional coefficients.\(^4\) In the model, we assume that only the coefficient of UI benefits is an arbitrary function of \( w \) and the remaining coefficients are usual parameters. As estimating an additional function results in a large efficiency loss, it would be helpful in empirical settings to exploit information that some coefficients are just constant parameters.

To deal with both functional and parametric coefficients, I conduct the local PRE procedure twice. First, I estimate the eight coefficients as functions and calculate the seven parametric values by taking weighted averages over \( w \). Next, I estimate the coefficient function of UI benefits again

\(^4\)The proposed estimator is still robust to unobserved heteroskedasticity if it is independent of covariates. However, we cannot apply this method if it is correlated with covariates and unobserved.
Figure 2: Estimation Results: Functional Coefficient of Log Weekly UI Benefit Level

given those parametric values. To achieve robustness, we may repeat the local PRE and PRE procedures given parametric or functional coefficient values until they converge. In all procedures, the simulated annealing method is used for the numerical optimization and confidence regions are constructed by the bootstrapping method.

Figure 2 shows the estimated functional coefficient of the log UI benefit levels. The solid line denotes estimated function values and the dotted lines indicate the 90% confidence region. The horizontal line indicates the poverty level, so the function value means the effect of the UI benefit on unemployment duration at a specific poverty level.

The estimated function confirms heterogeneous effects of UI benefits across different poverty levels though it is not very significant. As its shape is quite nonlinear and nonmonotone, a simple interaction term approach would lead us to wrong conclusions. New features revealed by the proposed model suggest deeper insight into labor market analysis. We will discuss each specific feature and related economic theories.

First, the estimated function has a negative sign for all poverty levels. This result might be counterintuitive for those who think of the classical reservation wage theory. As Harding (2006) noted, however, labor market dynamics are more complicated than can be explained by the classical theory. For instance, suppose an unemployed worker has a lower reservation wage than the market offer, but the offer is valid only at a location far away from his residence. Then, it can be geographic or liquidity constraints that prevent him from getting a job. In such a case, additional resources that lessen the constraint may reduce the unemployment duration. Considering that this data set
is from strong boom period, such an example is plausible. Second, the UI benefit seems to be more effective to those whose poverty levels are around 1.4. This result again suggests the existence of the liquidity constraint in a job search process.

To see how these results can be interpreted theoretically, we consider a simple two-period search model. The main argument in this model is that UI benefits may not always increase the unemployment duration if the search effort (or the probability of getting a job) is a function of both time and goods. Consider an economic agent who is unemployed in the first period and looks for a job in the next period. The probability of getting a job is \( \lambda(i, h) \) where \( i \) stands for goods spent looking for jobs and \( h \) stands for time. If he finds a job, then he will work for \( \bar{h} \) hours and receive the wage \( w \)—both variables are assumed to be fixed for simplicity. His utility maximization problem can be constructed as:

\[
\max_{c, h} U(c, 1 - h) + \beta (\lambda(i, h) U(w, 1 - \bar{h}) + (1 - \lambda(i, h)) U(UI, 1))
\]

s.t. \( c + i = e + UI \)

where \( U \) is a utility function that is increasing, concave and additively separable. The parameter \( \beta \) denotes time preference, \( e \) does an endowment in the first period, and \( UI \) does a UI benefit. We also assume that \( \lambda(i, h) \) is increasing and concave. First, suppose that the probability of getting a job is only a function of time, \( \lambda(i, h) = \lambda(h) \), as in the standard search theory. Then, any increase in the UI benefit will be consumed immediately and only the substitution effect which will decrease \( \lambda(i, h) \):

\[
c^* = e + UI
\]

\[
\frac{dh^*}{dUI} = \frac{\beta \lambda_2 u_1}{\beta \lambda_2 (e - u) + \lambda_2 (c^*, 1 - h^*)} < 0.
\]

Thus, unemployment duration always increases with UI. Now, consider the opposite case that the probability of getting a job is only a function of goods, \( \lambda(i, h) = \lambda(i) \). Then, there is no substitution effect and whether the unemployment will increase or decrease depends on the size of income effect:

\[
h^* = 0
\]

\[
\frac{di^*}{dUI} = \frac{U_{11} + \beta \lambda_1 u_1}{U_{11} + \beta \lambda_1 (e - u)} \leq 0.
\]

Holding other things to be same, \( \lambda_1 \) will be smaller in a boom period, which moves \( di^*/dUI \) in the positive direction. Thus, we can think of the UI benefit as more effective in a boom period. In
Table 3: The Estimation Results with Non-functional coefficient models

<table>
<thead>
<tr>
<th>Variable</th>
<th>AFT</th>
<th>PH</th>
<th>Scaled AFT</th>
<th>Scaled PH</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.0193**</td>
<td>0.0157**</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(0.0019)</td>
<td>(0.0019)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Sex (male=1)</td>
<td>-0.2176**</td>
<td>-0.0534</td>
<td>-11.2746**</td>
<td>-3.4013</td>
<td>-6.8073</td>
</tr>
<tr>
<td>(0.0457)</td>
<td>(0.0476)</td>
<td>(2.3679)</td>
<td>(3.0318)</td>
<td>[-13.8581, 2.2895]</td>
<td></td>
</tr>
<tr>
<td>Race (white=1)</td>
<td>-0.0567**</td>
<td>-0.1446**</td>
<td>-2.9378</td>
<td>-3.2102**</td>
<td>-10.2294**</td>
</tr>
<tr>
<td>(0.0489)</td>
<td>(0.0504)</td>
<td>(2.5337)</td>
<td>(3.2102)</td>
<td>[-22.2792, -1.9242]</td>
<td></td>
</tr>
<tr>
<td>Number of dependents</td>
<td>0.0328**</td>
<td>0.0291*</td>
<td>1.6995**</td>
<td>1.8535*</td>
<td>3.0676*</td>
</tr>
<tr>
<td>(0.0160)</td>
<td>(0.0159)</td>
<td>(0.8290)</td>
<td>(1.0127)</td>
<td>[-0.3338, 5.9882]</td>
<td></td>
</tr>
<tr>
<td>High school dropout</td>
<td>0.4144**</td>
<td>0.2877**</td>
<td>21.4715**</td>
<td>18.3248**</td>
<td>16.0601**</td>
</tr>
<tr>
<td>(0.0640)</td>
<td>(0.0647)</td>
<td>(3.1611)</td>
<td>(4.1210)</td>
<td>[4.2108, 54.2159]</td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td>-0.1996**</td>
<td>-0.2621**</td>
<td>-10.3420</td>
<td>-16.6943**</td>
<td>-32.6510**</td>
</tr>
<tr>
<td>(0.0436)</td>
<td>(0.0442)</td>
<td>(2.2591)</td>
<td>(2.8153)</td>
<td>[-61.5113, 21.7237]</td>
<td></td>
</tr>
<tr>
<td>State unemployment rate</td>
<td>0.1660**</td>
<td>0.0409*</td>
<td>8.6010**</td>
<td>2.6051*</td>
<td>5.5885*</td>
</tr>
<tr>
<td>(0.0205)</td>
<td>(0.0234)</td>
<td>(1.0622)</td>
<td>(1.4904)</td>
<td>[1.3499, 11.2254]</td>
<td></td>
</tr>
<tr>
<td>Log weekly UI benefit level</td>
<td>0.2197**</td>
<td>-0.2735**</td>
<td>11.3834**</td>
<td>-17.4204**</td>
<td>-20.9614**</td>
</tr>
<tr>
<td>(0.0464)</td>
<td>(0.0630)</td>
<td>(2.4041)</td>
<td>(4.0127)</td>
<td>[-36.1997, -9.2592]</td>
<td></td>
</tr>
<tr>
<td>Log pre–UI wage</td>
<td>0.0929**</td>
<td>-0.0264</td>
<td>4.8134**</td>
<td>-1.6815</td>
<td>-4.7092</td>
</tr>
<tr>
<td>(0.0379)</td>
<td>(0.0369)</td>
<td>(1.9637)</td>
<td>(2.3503)</td>
<td>[-12.9182, 5.3928]</td>
<td></td>
</tr>
<tr>
<td>Poverty level</td>
<td>-0.0443*</td>
<td>-0.0308</td>
<td>-2.2953*</td>
<td>-1.9618</td>
<td>3.4526</td>
</tr>
<tr>
<td>(0.0263)</td>
<td>(0.0279)</td>
<td>(1.3627)</td>
<td>(1.7771)</td>
<td>[-1.5040, 8.8539]</td>
<td></td>
</tr>
</tbody>
</table>

addition, if $\lambda(i, h)$ has the form $\lambda(i - K, h)$ for a constant $K$, the liquidity constraint argument can be explained by the model. In the general case of $\lambda(i, h)$, the relative sizes of the substitution and income effects decides the effect of the UI benefit. Thus, the result will heavily depend on the specific functional shapes and coefficient values imposed in each model. As the proposed estimator does not impose many functional restrictions, it may be helpful in finding the appropriate functions and parameter values.

Finally, I estimate the model without functional coefficients and compare the results in different model specifications. The AFT, PH and Transformation models are considered and Table 3 summarizes the estimation results. Since the transformation model requires a scale normalization, I rescale the result of the AFT and PH by using the delta method. Values in parenthesis stand for the standard errors of each estimation, and the superscripts * and ** denote significance at the 90% and 95% confidence levels, respectively. The 95% confidence intervals for the PRE are constructed by the bootstrapping method and denoted in square brackets.

Before looking at the results, recall that the AFT model assuming the Weibull baseline hazard
function is the most restrictive one among the three, followed by the PH model and the transformation model. Thus, the result of the PRE does not assume even a proportional hazard ratio. It is worthwhile to note that the coefficient of the UI benefit in the AFT model is significantly positive while it is significantly negative in the remaining two models. Also, the AFT model alone predicts the positive effect of the pre-UI wage significantly. These results demonstrate the risks of restrictive model specifications. Other coefficient signs are as expected and confirm the standard result of the UI benefit analysis.

In summary, relaxing strict restrictions on the hazard function gives very different results in a UI benefit analysis. Also, the functional coefficient model and the proposed estimation method make it possible to see its heterogeneous effects across different poverty levels. Results found in this section may suggest new aspects of labor markets, which can be considered by researchers in related fields, such as optimal social insurance theory and structural estimations of unemployment duration analysis.

6 Concluding Remarks

In this paper I consider nonparametric functional dependency of interesting parameters in unspecified transformation models. To estimate the model, I develop the local maximum rank correlation estimation method and investigate its asymptotic properties. We can identify the flexible model by using local order information. In addition, the proposed estimator has the advantage that it can be easily combined with existing methods that resolve various censoring and truncation problems. We also look at some extensions of set identification under different situations with an interval-valued regressor. These features show that the proposed model can be useful in empirical research, especially in duration analyses.

Using the model, we estimate the heterogeneous effects of UI benefits at different poverty levels. The estimation results indicate different effects of UI benefits across different poverty levels, though it is not very significant. The functional shapes are quite nonlinear, so it may not be easily detected by the conventional interaction terms approach. Furthermore, the function is hump-shaped, which is not predicted by existing economic theories. The estimation results can be explained in the standard search theory framework by generalizing search effort to be a function both of time and of goods. We also see by comparing estimation results in different models that restrictive model specifications might lead a researcher to wrong conclusions.

I conclude this paper by suggesting areas of future research. First, an extension to panel data models can be considered. A specific example would be multiple duration models with time-varying coefficients. The functional parameter $\beta(\cdot)$ is now a function of a time index. A deficiency
of rank estimation methods in duration analysis compared to the partial maximum likelihood approach has been its difficulty to adopt time-varying covariates and/or parameters.\footnote{Recently, Woutersen and Hausman (2005) proposed a rank estimation method with time varying covariates but it requires a more restricted functional form than the transformation model. See Shin (2006a).} An extension of the proposed model to panel data could be a new direction that exploits both the time-varying property and flexible functional assumptions. Second, we may develop a procedure for testing a model specification by adopting the method in Bierens and Ploberger (1997). Third, we can investigate appropriate functional shapes in structural model estimations. Although we have seen heterogeneous effects of UI benefits, we should consider structural models and a general equilibrium approach to fully assess policy implication. The proposed model and estimation method may help us to find proper restrictions. Finally, we may check the heterogeneous UI effects with other data sets. Since the estimation results in this paper are restricted to the strong boom period in the late 1990s, additional research is necessary to see how these change during a recession.
Appendix

Proof of Theorem 2.1: I verify the conditions of Lemma 2.1 in Newey and McFadden (1994). Compactness follows from the Assumption A2 and A3. The condition (7) and the Theorem in Han (1987) assure identification of the model. Continuity follows from the Assumption A4. The part to be added is showing uniform convergence of the sample objective function. Since the first element of \( \beta(w_0) \) is normalized as 1, I denote \((1, \theta) = \beta(w_0)\) and define \( \Gamma_n(\theta) \) and \( \Gamma(\theta) \) as:

\[
\Gamma_n(\theta) = \frac{1}{n(n - 1)} \sum_{i \neq j} (m_{ij}(\beta(w_0)) - m_{ij}(\beta(w_0))) K_h(w_i - w_0) K_h(w_j - w_0) \tag{26}
\]

\[
\Gamma(\theta) = E [m_{ij}(\beta(w_0)) - m_{ij}(\beta_0(w_0))] \phi(w_0)^2. \tag{27}
\]

Note that I suppress the dependency of \( \theta \) on \( w_0 \) for notational simplicity. So, we want to show that

\[
\sup |\Gamma_n(\theta, w_0) - \Gamma(\theta, w_0)| = o_p(1). \tag{28}
\]

By the triangular inequality,

\[
\sup |\Gamma_n(\theta) - \Gamma(\theta)| \leq \sup |\Gamma_n(\theta) - E\Gamma_n(\theta)| + \sup |E\Gamma_n(\theta) - \Gamma(\theta)|. \tag{29}
\]

We first look at the first term in the right hand side. Define \( \tilde{\Gamma}_n = h^2\Gamma_n \), then \( \tilde{\Gamma}_n(\theta) - E\tilde{\Gamma}_n(\theta) \) is the second order U-process with zero-mean. Applying the Corollary 7 in Sherman (1994a), we have

\[
\sup |\Gamma_n(\theta) - E\Gamma_n(\theta)| = \frac{1}{h^2} O_p \left( \frac{1}{\sqrt{n}} \right),
\]

which becomes \( o_p(1) \) by the bandwidth condition.

We next turn our attention to the second term. It is the bias term of the standard nonparametric conditional moment estimation. Applying the Taylor expansion and the change of variable as usual, we get the leading term of \( O(h^2) \), which is \( o_p(1) \) uniformly over parameter space.

Lemma 6.1 Suppose \( \hat{\theta} \) maximizes \( \Gamma_n(\theta) \) and \( \theta_0 \) maximizes \( \Gamma(\theta) \). Let \( \{\delta_n\}, \{\varepsilon_n\} \) and \( \{\eta_n\} \) be sequences of nonnegative real numbers converging to zero as \( n \) goes to infinity. Suppose the following three conditions hold:

(i) \( |\hat{\theta} - \theta_0| = O_p(\delta_n) \).
(ii) There exists a neighborhood $\mathcal{N}$ of $\theta_0$ and a constant $\kappa > 0$ such that

$$\Gamma (\theta) \leq -\kappa |\theta - \theta_0|^2$$

for all $\theta$ in $\mathcal{N}$.

(iii) Uniformly over $O_p(\delta_n)$ neighborhoods of $\theta_0$, $\Gamma_n (\theta) = \Gamma (\theta) + O_p (|\theta - \theta_0| \eta_n) + o_p \left( |\theta - \theta_0|^2 \right) + O_p (\varepsilon_n)$. Then,

$$|\hat{\theta} - \theta_0| = O_p \left( \max \left\{ \varepsilon_n^{1/2}, \eta_n \right\} \right).$$

**Proof.** It follows directly from the proof of the Theorem 1 in Sherman (1993) by replacing $W_n = O_p (\eta_n)$. ■

**Lemma 6.2** Suppose $|\hat{\theta} - \theta_0| = O_p \left( 1/\sqrt{nh} \right)$ for $\theta_0 \in \text{int}(\Theta)$. Suppose that uniformly over $O_p \left( 1/\sqrt{nh} \right)$ neighborhoods of $\theta_0$,

$$\Gamma_n (\theta) = \frac{1}{2} (\theta - \theta_0)' V (\theta - \theta_0) + \frac{1}{\sqrt{nh}} (\theta - \theta_0)' W_n + o_p (1/\sqrt{nh})$$

where $V$ is a negative definite matrix. Suppose that $\lim h^2 \sqrt{nh} = \lambda$ for $0 \leq \lambda < \infty$ and that $W_n$ converges in distribution to $N (\lambda b, \Delta)$ at the rate of $\sqrt{nh}$ for some constant $b$. Then,

$$\sqrt{nh} (\hat{\theta} - \theta_0) \overset{d}{\rightarrow} N (\lambda b, V^{-1} \Delta V^{-1}).$$

**Proof.** It follows directly from the proof of the Theorem 2 in Sherman (1993) by replacing $t_n = \sqrt{nh} (\hat{\theta} - \theta_0)$. ■

**Proof of Theorem 2.2:** I first show that

$$\Gamma_n (\theta) = \frac{1}{2} (\theta - \theta_0)' V (\theta - \theta_0) + \frac{1}{\sqrt{nh}} (\theta - \theta_0)' W_n + o_p \left( |\theta - \theta_0|^2 \right) + O_p (1/\sqrt{nh^2}) \quad (31)$$

uniformly over $o_p (1)$ neighborhoods of $\theta_0$. Then, we have $|\hat{\theta} - \theta_0| = O_p \left( 1/\sqrt{nh} \right)$ by applying the Lemma 6.1 repeatedly. Next, I show that the $O_p (1/\sqrt{nh^2})$ term becomes $o_p (1/\sqrt{nh})$ uniformly over $O_p \left( 1/\sqrt{nh} \right)$ neighborhoods of $\theta_0$, which establish the desired property by the Lemma 6.2.
We can decompose the sample objective function $\Gamma_n(\theta, w_0)$ as below (see Hoeffding (1948) and Serfling (1980)):

\[
\Gamma_n(\theta) = E \Gamma_n(\theta) + U_n^{1} g_1(z, \theta) + U_n^{2} g_2(z_1, z_2, \theta)
\]

where $U_n^{1} g_1(z, \theta)$ and $U_n^{2} g_2(z_1, z_2, \theta)$ are respectively the first and the second order degenerate $U$-processes. The first term can be approximated as a quadratic form around $\theta_0$ as follows:

\[
E \Gamma_n(\theta) = \frac{1}{2} E [\tau(z, \theta) - \tau(z, \theta_0)]
\]

\[
= \frac{1}{2} (\theta - \theta_0)' \left[ \frac{1}{2} E \nabla^2 \tau(z, \theta_0) \right] (\theta - \theta_0) + \frac{1}{2} (\theta - \theta_0)' E \nabla_1 \tau(z, \theta_0) + o_p \left( |\theta - \theta_0|^2 \right)
\]

\[
= \frac{1}{2} (\theta - \theta_0)' V(\theta - \theta_0) + (\theta - \theta_0)' \nabla_1 B(\theta_0) + o_p \left( |\theta - \theta_0|^2 \right).
\]

(32)

where

\[
B(\theta) = h^2 \mu_2 \phi(w_0) \left( \nabla_{ww} E [m_{ij}(\theta)|w_i = w_j = w_0] \phi(w_0) \right.
\]

\[
+ \nabla_w E [m_{ij}(\theta)|w_i = w_j = w_0] \phi'(w_0) + E [m_{ij}(\theta)|w_i = w_j = w_0] \phi^{(2)}(w_0) \Big) + O(h^4)
\]

Note that the bias term $E \nabla_1 \tau(z, \theta_0)$ appears. We next turn out attention to the second term.

\[
U_n^{1} g_1(z, \theta) = \frac{1}{n} \sum \left( \tau(z, \theta) - \tau(z, \theta_0) \right) - 2E \Gamma_n(\theta)
\]

\[
= (\theta - \theta_0)' \left[ \frac{1}{n} \sum \nabla_1 \tau(z, \theta_0) - E \nabla_1 \tau(z, \theta_0) \right]
\]

\[
+ \frac{1}{2} (\theta - \theta_0)' \left[ \frac{1}{n} \sum \nabla_2 \tau(z, \theta_0) - 2V(\theta - \theta_0) + o \left( |\theta - \theta_0|^2 \right) \right].
\]

(33)

Combining (32) and (33), we have uniformly over $o_p(1)$ neighborhoods of $\theta_0$

\[
E \Gamma_n(\theta) + U_n^{1} g_1(z, \theta) = \frac{1}{2} (\theta - \theta_0)' V(\theta - \theta_0) + \frac{(\theta - \theta_0)' W_n + o_p \left( |\theta - \theta_0|^2 \right) + O \left( |\theta - \theta_0| h^4 \right),
\]

where

\[
W_n = \frac{1}{\sqrt{nh}} \sum \left[ \nabla_1 \tau(z, \theta_0) - E \nabla_1 \tau(z, \theta_0) + 2 \nabla_1 B(\theta_0) \right],
\]

which converges in distribution to $N(0, \Delta)$ and the remaining term $O(h^4)$ converges to zero faster than $O_p(1/nh^2)$ by the bandwidth condition. Then, uniformly over $O_p \left( 1/\sqrt{nh^2} \right)$ neighborhoods of $\theta_0$, it converges faster than $o_p(1/nh)$.

Finally, I investigate the third term. Define $\tilde{g}_2(z_1, z_2, \theta) = h^2 g_2(z_1, z_2, \theta)$ which is Euclidean.
with finite envelope. Then, the Theorem 3 in Sherman (1994b) implies that

\[ U_n^2 g_2 (z_1, z_2, \theta) = \frac{1}{h^2} O_p \left( \frac{1}{n} \right) \]

uniformly over \( o_p(1) \) neighborhoods of \( \theta_0 \). Now consider \( |\theta - \theta_0| = O_p \left( \frac{1}{\sqrt{nh^2}} \right) \) that is assured by (31) and the Lemma 6.1. Applying the Theorem 3 again, we have

\[ U_n^2 g_2 (z_1, z_2, \theta) \]

\[ = \frac{1}{h^2} O_p \left( \frac{1}{n(h^2)^{\alpha/2}} \right) \]

\[ = O_p \left( \frac{1}{nh^{\alpha/2 + \alpha/2}} \right) \]

for \( 0 < \alpha < 1 \). Therefore, it becomes \( o_p(1/nh) \) uniformly over \( O_p \left( \frac{1}{\sqrt{nh^2}} \right) \) neighborhoods of \( \theta_0 \) by the bandwidth condition as \( \alpha \) converges to 1. \( \blacksquare \)

**Proof of Proposition 3.1:** The relation between \((x_i, v_{i0}, v_{i1})\) and \((x_j, v_{j0}, v_{j1})\) is partitioned into six cases:

(i) \( \Delta x_{ij}^\beta + \Delta v_{01} > 0 \Rightarrow x_{ij}^\beta + v_i > x_{ij}^\beta + v_j \Rightarrow P(y_i > y_j | x_i, x_j, v_i, v_j) > P(y_i < y_j | x_i, x_j, v_i, v_j) \).

Therefore, \( P(y_i > y_j | z_{ij}) > P(y_i < y_j | z_{ij}) \) since

\[ P(y_i > y_j | z_{ij}) - P(y_i < y_j | z_{ij}) = \int \left( P(y_i > y_j | z_{ij}, v_i, v_j) - P(y_i < y_j | z_{ij}, v_i, v_j) \right) dP(v_i, v_j | z_{ij}) \]

\[ = \int \left( P(y_i | x_i, v_i) - P(y_j | x_j, v_j) \right) dP(v_i, v_j | z_{ij}) > 0. \]

(ii) \( \Delta x_{ij}^\beta + \Delta v_{10} \leq 0 \Rightarrow x_{ij}^\beta + v_i \leq x_{ij}^\beta + v_j \Rightarrow P(y_i > y_j | x_i, v_i) \leq P(y_i < y_j | x_i, v_j) \). Therefore, \( P(y_i > y_j | z_{ij}) \leq P(y_i < y_j | z_{ij}) \).

(iii) For the remaining four cases, \( x_{ij}^\beta + v_i \geq x_{ij}^\beta + v_j \). Thus, \( P(y_i | x_i, v_{i0}, v_{i1}) \geq P(y_j | x_j, v_{j0}, v_{j1}) \).

When we substitute \( b \) for \( \beta \), it gives contradictory relation between \( P(y_i | x_i, v_{i0}, v_{i1}) \) and \( P(y_j | x_j, v_{j0}, v_{j1}) \) if and only if \( \Delta x_{ij}^\beta + \Delta v_{01} > 0 \geq \Delta x_{ij}^b + \Delta v_{10} \) or \( \Delta x_{ij}^b + \Delta v_{01} > 0 \geq \Delta x_{ij}^\beta + \Delta v_{10} \). Therefore, \( \beta \) is identified relative to \( b \) if and only if \( P(V_1(b)) > 0 \). \( \blacksquare \)

**Proof of Lemma 3.1:** We can rewrite the set \( V_1(b) \) as follows:
Let \( b \in B_1 \). We partition the support into the following three cases:

(i) \( \Delta x'_{ij}b + \Delta v_{01} > 0 \). Then, \( \Delta x'_{ij}b + \Delta v_{01} > 0 \) and \( P(y_i > y_j| z_{ij}) > P(y_i < y_j| z_{ij}) \). Thus, the objective function is

\[
Q_1 (b) = E_{z_{ij}} [P(y_i > y_j| z_{ij}) - P(y_i < y_j| z_{ij})].
\]

(ii) \( \Delta x'_{ij}b + \Delta v_{01} \leq 0 \). Then, \( \Delta x'_{ij}b + \Delta v_{01} \leq 0 \) and \( P(y_i > y_j| z_{ij}) \leq P(y_i < y_j| z_{ij}) \). Thus, the objective function is

\[
Q_1 (b) = E_{z_{ij}} [P(y_i > y_j| z_{ij}) - P(y_i < y_j| z_{ij})].
\]

(iii) \( \Delta x'_{ij}b + \Delta v_{01} \leq 0 < \Delta x'_{ij}b + \Delta v_{10} \). Then the objective function is

\[
Q_1 (b) = E_{z_{ij}} [(\{P(y_i > y_j| z_{ij}) - P(y_i < y_j| z_{ij})\} (2\lambda (z_{ij}) - 1)]
\]

\[
= E_{z_{ij}} [P(y_i > y_j| z_{ij}) - P(y_i < y_j| z_{ij})].
\]

Now, consider \( b^* \in B_1^v \). In similar way, we can show that there exist \( z_{ij} \) with positive measure such that the integrand of \( Q_1 \) will be \(-|P(y_i > y_j| z_{ij}) - P(y_i < y_j| z_{ij})|\). Thus, \( Q_1 (b) \geq Q_1 (b^*) \) for all \( b \in B_1 \) and \( b^* \in B_1^v \) and the strict inequality holds if \( P(P(y_i > y_j| z_{ij}) = P(y_i < y_j| z_{ij})) = 0 \).

\[\square\]

**Proof of Proposition 3.2:** We partition the space of \( z_{ij} \) into the following three cases:

(i) \( T_0 (y) - T_m < w_m \Rightarrow T_0 (y) - T_0 (y_0) < (x'_i b + v_i) - (x'_j b + v_j) \Rightarrow P(y_i \geq y|x_i, v_i) > P(y_j \geq y_0|x_j, v_j) \), which implies \( P(y_i \geq y|x_i, v_i, v_{i0}, v_{i1}) > P(y_j \geq y_0|x_j, v_j, v_{j0}, v_{j1}) \) since

\[
P(y_i \geq y|x_i, v_i, v_{i0}, v_{i1}) - P(y_j \geq y_0|x_j, v_j, v_{j0}, v_{j1})
\]

\[
= \int [P(y_i \geq y|x_i, v_i, v_{i0}, v_{i1}) - P(y_j \geq y_0|x_j, v_j, v_{j0}, v_{j1})] dP(v_i, v_j|z_{ij})
\]

\[
= \int [P(y_i \geq y|x_i, v_i) - P(y_j \geq y_0|x_j, v_j)] dP(v_i, v_j|z_{ij}) > 0.
\]
(ii) $w_M \leq T_0(y) - T_M \Rightarrow (x'_i b + v_i) - \left( x'_j b + v_j \right) \leq T_0(y) - T_0(y_0) \Rightarrow P(y_i \geq y|x_i, v_i) \leq P(y_j \geq y_0|x_j, v_{j1})$, which implies $P(y_i \geq y|x_i, v_{i0}, v_{i1}) \leq P(y_j \geq y_0|x_j, v_{j0}, v_{j1})$.

(iii) For the remaining cases, we have $(x'_i b + v_i) - \left( x'_j b + v_j \right) \geq T_0(y)$. Thus, $P(y_i \geq y|x_i, v_{i0}, v_{i1}) \geq P(y_j \geq y_0|x_j, v_{j0}, v_{j1})$.

When we substitute $T(y)$ for $T_0(y)$, it gives contradictory relation between $P(y_i \geq y|x_i, v_{i0}, v_{i1})$ and $P(y_j \geq y_0|x_j, v_{j0}, v_{j1})$ if and only if $T_0(y) < w_m \leq w_M \leq T(y)$ or $T(y) < w_m \leq w_M \leq T_0(y)$. Therefore $T(y_0)$ is identified relative to $T(y)$ if and only if $P(V_2(T(y))) > 0$. 

**Proof of Lemma 3.2:** We can rewrite the set $V_2(b)$ as follows:

$$V_2(T(y)) = \{ z_{ij} : (P(y_i \geq y|x_i, v_{i0}, v_{i1}) > P(y_j \geq y_0|x_j, v_{j0}, v_{j1}) \cap w_M \leq T(y) - T_M) \cup (T(y) - T_m < w_m \cap P(y_i \geq y|x_i, v_{i0}, v_{i1}) \leq P(y_j \geq y_0|x_j, v_{j0}, v_{j1})) \}$$

Let $T(y) \in B_{2y}$. We partition the support into the following five cases:

(i) $T(y) - T_m < w_m$. Then, $T(y) - T_M < w_M$ and $P(y_i \geq y|x_i, v_{i0}, v_{i1}) > P(y_j \geq y_0|x_j, v_{j0}, v_{j1})$.

Thus, the objective function will be

$$Q_2(T(y)) = E_{z_{ij}} [P(y_i \geq y|x_i, v_{i0}, v_{i1}) - P(y_j \geq y_0|x_j, v_{j0}, v_{j1})].$$

(ii) $w_M \leq T(y) - T_M$. Then, $w_m \leq T(y) - T_m$ $P(y_i \geq y|x_i, v_{i0}, v_{i1}) \leq P(y_j \geq y_0|x_j, v_{j0}, v_{j1})$.

Thus the objective function will be

$$Q_2(T(y)) = E_{z_{ij}} [P(y_i \geq y|x_i, v_{i0}, v_{i1}) - P(y_j \geq y_0|x_j, v_{j0}, v_{j1})].$$

(iii) $T(y) - T_M < w_m \leq T(y) - T_m$. Then, $w_M > T(y) - T_M$, so the objective function will be

$$Q_2(T(y)) = E_{z_{ij}} [(P(y_i \geq y|x_i, v_{i0}, v_{i1}) - P(y_j \geq y_0|x_j, v_{j0}, v_{j1})) (2\lambda_2(z_{ij}) - 1)]$$

$$= E_{z_{ij}} [P(y_i \geq y|x_i, v_{i0}, v_{i1}) - P(y_j \geq y_0|x_j, v_{j0}, v_{j1})].$$

(iv) $T(y) - T_M < w_M \leq T(y) - T_m$. Then, $w_m \leq T(y) - T_m$, so the objective function will be

$$Q_2(T(y)) = E_{z_{ij}} [(P(y_i \geq y|x_i, v_{i0}, v_{i1}) - P(y_j \geq y_0|x_j, v_{j0}, v_{j1})) (2\lambda_2(z_{ij}) - 1)]$$

$$= E_{z_{ij}} [P(y_i \geq y|x_i, v_{i0}, v_{i1}) - P(y_j \geq y_0|x_j, v_{j0}, v_{j1})].$$
Thus the objective function will be
\[
Q_2(T(y)) = E_{z_{ij}} [(P(y_i \geq y|x_i, v_{i0}, v_{i1}) - P(y_j \geq y_0|x_j, v_{j0}, v_{j1})) (2\lambda_2(z_{ij}) - 1)] \\
= E_{z_{ij}} |P(y_i \geq y|x_i, v_{i0}, v_{i1}) - P(y_j \geq y_0|x_j, v_{j0}, v_{j1})|.
\]

Now consider \(T^*(y) \in B_{2y}^c\). In similar way, we can show that there exist \(z_{ij}\) with positive measure such that the integrand of \(Q_2\) will be \(-|P(y_i \geq y|x_i, v_{i0}, v_{i1}) - P(y_j \geq y_0|x_j, v_{j0}, v_{j1})|\). Thus, \(Q_2(T(y)) \geq Q_1(T^*(y))\) for all \(T(y) \in B_{2y}\) and \(T^*(y) \in B_{2y}^c\) and the strict inequality holds if \(P(P(y_i \geq y|x_i, v_{i0}, v_{i1}) = P(y_j \geq y_0|x_j, v_{j0}, v_{j1})) = 0\).

**Proof of Proposition 3.3:** We partition the space of \((x, v_0, v_1)\) into the following three cases:

(i) \(x' \beta + v_0 > 0 \implies x' \beta + v > 0 \implies P(T_0(y) > 0|x, v) > \frac{1}{2} \implies P(T_M(y) > 0|x, v) > \frac{1}{2}\), which implies \(P(T_M(y) > 0|x, v_0, v_1) > \frac{1}{2}\) since

\[
P(T_M(y) > 0|x, v_0, v_1) = \int P(T_M(y) > 0|x, v, v_0, v_1) dP(v|x, v_0, v_1)
= \int P(T_M(y) > 0|x, v) dP(v|x, v_0, v_1) > \frac{1}{2}.
\]

(ii) \(x' \beta + v_1 < 0 \implies x' \beta + v < 0 \implies P(T_0(y) > 0|x, v) < \frac{1}{2} \implies P(T_m(y) > 0|x, v) < \frac{1}{2}\), which implies \(P(T_m(y) > 0|x, v_0, v_1) < \frac{1}{2}\).

(iii) \(x' \beta + v_0 \leq 0 \leq x' \beta + v_1 \implies x' \beta + v \geq 0\). Thus, \(0 \leq P(T_m(y) > 0|x, v_0, v_1) \leq P(T_M(y) > 0|x, v_0, v_1) \leq 1\).

Now we consider an imposter \(b\) such that \(x'b + v_1 \leq -d(y)\). Then, \(x'b + v \leq -d(y)\) and it follows that

\[
P(T_M(y) > 0|x, v) = P(T_0(y) > -(T_M(y) - T_0(y))|x, v)
\leq P(T_0(y) > -d(y)|x, v)
= P(\varepsilon > -(x'b + v) - d(y)|x, v) \leq \frac{1}{2}.
\]

Thus, \(P(T_M(y) > 0) \geq \frac{1}{2}\) and it gives contradictory result if \(x' \beta + v_0 > 0 \geq -d(y) \geq x' \beta + v_1\). Similarly, we can show that it contradicts if \(x' \beta + v_1 < 0 \leq d(y) \leq x' \beta + v_0\). Therefore, \(\beta\) is identified relative to \(b\) if and only if \(P(V_3(b)) > 0\).
Proof of Lemma 3.3: We can rewrite the set $V_3(b)$ as follows:

$$V_3(b) = \left\{ (x, v_0, v_1) : \left( P_M > \frac{1}{2} \cap x'b + v_1 \leq -d(y) \right) \right\} \cup \left( P_m \leq \frac{1}{2} \cap x'b + v_0 > d(y) \right) \right\}$$

(35)

Let $b \in B_3$. We partition the support into the following three cases:

(i) $x'b + v_1 \leq -d(y)$. Then, $x'b + v_0 \leq d(y)$ and $P_M \leq \frac{1}{2}$. Thus, $Q_3(b) = \left| P_M - \frac{1}{2} \right| + \left| P_m - \frac{1}{2} \right|$. 

(ii) $x'b + v_0 \geq d(y)$. Then, $x'b + v_1 \geq -d(y)$ and $P_m \geq \frac{1}{2}$. Thus, $Q_3(b) = \left| P_M - \frac{1}{2} \right| + \left| P_m - \frac{1}{2} \right|$. 

(iii) For the remaining cases,

$$Q_3(b) = \left\{ \left( P_M - \frac{1}{2} \right) (2\lambda_3M - 1) + \left( P_m - \frac{1}{2} \right) (2\lambda_3m - 1) \right\} (1 - \lambda_3M) + \lambda_3m$$

$$= \left\{ \left| P_M - \frac{1}{2} \right| + \left| P_m - \frac{1}{2} \right| \text{ if } P_M < \frac{1}{2} \text{ or } P_m > \frac{1}{2} \right\}$$

otherwise.

Now, consider $b^* \in B_3^c$. For those $b^*$, we have the following two cases with positive measure.

(i) $x'b^* + v_1 \leq -d(y)$. Then, $x'b^* + v_0 \leq d(y)$ and $P_M > \frac{1}{2}$. Thus, the integrand of $Q_3(b^*)$ contains $-\left| P_M - \frac{1}{2} \right| - \left( P_m - \frac{1}{2} \right)$ terms which is less than $\left| P_M - \frac{1}{2} \right| + \left| P_m - \frac{1}{2} \right|$. 

(ii) $x'b^* + v_0 \geq d(y)$. Then, $x'b^* + v_1 \geq -d(y)$ and $P_m < \frac{1}{2}$. Thus, the integrand of $Q_3(b^*)$ contains $\left( P_M - \frac{1}{2} \right) - \left| P_m - \frac{1}{2} \right|$ terms which is less than $\left| P_M - \frac{1}{2} \right| + \left| P_m - \frac{1}{2} \right|$. 

Therefore, $Q_3(b) \geq Q_3(b^*)$ for all $b \in B_3$ and $b^* \in B_3^c$. The strict inequality holds if $P \left( P_M = \frac{1}{2} \text{ or } P_m = \frac{1}{2} \right) = 0$. 

Proof of Corollary 3.1: (i) This can be shown by an example. Consider the following data generating process. A covariate $x$ is a binary variable with equal probability and another covariate $v$ follows the uniform distribution on $[0, 2]$. Suppose that $v$ is observed as an interval value so that

$$\{(x, v_0, v_1) : (0, 0, 1), (0, 1, 2), (1, 0, 1), (1, 1, 2)\}.$$

The error term $\varepsilon$ can follow any distribution with $\text{med} \ (\varepsilon) = 0$. The identity transformation function $T(y) = y$ is known for a moment. It can be calculated from the conditions (34) and (35) in the
proof of the proposition 3.1 and 3.3 respectively that $B_1 = (-4, 2)$ and $B_3 = (-2, +\infty)$. Therefore $B_1$ and $B_3$ do not include each other.

(ii) It follows from the fact that $V_3(b)$ becomes the null set from the condition (35) if $d(y)$ diverges to infinity.
References


