Identification and Estimation of Nonlinear Regression Models using Control Functions

Daniel Gutknecht

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

University of Warwick

Department of Economics

May 2012

Acknowledgements

I would like to thank my supervisors Valentina Corradi and Wiji Arulampalam for their continuous encouragement and support throughout the PhD studies. Both provided invaluable comments towards the completion of this thesis and taught me crucial skills, without which I wouldn't have arrived to this point. I am also very thankful to both of them for having been patient with me and the way I delivered work to them (more than once). Finally, I am particularly indebted to Valentina Corradi for her fantastic support and advice during the job market phase.

I would also like to express my gratitude to Mark Stewart for his very useful suggestions and his reference letter as well as to Jennifer Smith for her advice regarding the collection of unemployment data from the British Household Panel Survey. Many thanks also to S. Khan, Y. Shin, and E. Tamer for providing me with their GAUSS routines from the paper *Heteroscedastic Transformation Models with Covariate Dependent Censoring* (JBES 2011, 29(1), p.40-48), which formed the basis of my own GAUSS routines in Chapter two.

Finally, I am grateful to the Department of Economics for providing a stimulating research environment and for the very generous funding support (my special thanks, at this point, go to Abhinay Muthoo). Financial support from the ESRC and the Royal Economic Society is also gratefully acknowledged.

The work of chapter two has been presented at Oxford, the University of Mannheim, the University of Sydney, the University of Vienna, the LACEA-LAMES meeting in Santiago (2011), the EEA-ESEM meeting in Oslo (2011), the University of Warwick, and the Econometric Study Group Meeting in Bristol (2009). Comments received from seminar participants (in particular from Steve Bond and Gerard van den Berg) are gratefully acknowledged.

Declaration

This thesis is submitted to the University of Warwick in accordance with the requirements of the degree of Doctor of Philosophy. I declare that any material contained in this thesis has not been submitted for a degree to any other university. Paper versions of Chapter two and three have recently appeared in the working paper series of the Economics department at the University of Warwick (series no. 961 and 991).

Daniel Gutknecht

24 May 2012

Contents

List of Tables

List of Figures

1 Introduction

According to Blundell and Powell (2003), the development of strategies to identify and estimate certain parameters or even entire functions of regression models under endogeneity has arguably been one of the main contributions of microeconometrics to the statistical literature. The term endogeneity, in this context, refers to a correlation between observable regressor(s) and model unobservable(s), which can arise for multiple reasons such as, among others, omitted variables, measurement error, unobserved heterogeneity, or simultaneous causality. Whereas linear identification and estimation techniques to address endogeneity date back as far as 1928 (Stock and Trebbi, 2003), advances in the field of nonlinear models are much more recent: nonlinear parametric models under endogeneity only came under investigation during the 1970s and 1980s (e.g. Ameniya, 1974; Hansen, 1982), and it was not until the mid 1990s that models of (partially) unknown functional form were considered.[1]

Literature focusing on endogeneity in the latter typically uses two different identification ideas that, despite the common assumption of an available instrumental variable (vector), can be distinguished by the way identification is achieved: nonparametric instrumental variable methods rely on the existence of a suitable moment condition, which is based on the instrument(s) and in turn gives rise to an estimator (e.g. Lewbel, 1998; Ai and Chen, 2003; Darolles, Fan, Florens, and Renault, 2011). Control function methods on the other hand require the existence of a so called control function, that is a function of the model observables and the instrument (vector) satisfying a conditional (mean) independence assumption. Replacing the control function by an estimated correspondent and incorporating this estimate into

[1] The general appeal of nonlinear models with (partially) unknown functional form for economics has been aptly discussed in the literature (e.g. Härdle, 1992): motives range from the ability to capture empirical phenomena such as individual heterogeneity effects to the provision of a direct link to economic theory, which itself often leaves functional relationships between different variables unspecified.

a suitable statistic to control for the endogeneity yields a consistent estimator of the function or parameter of interest. Examples of this literature include e.g. Newey, Powell, and Vella (1999), Blundell and Powell (2003), Imbens and Newey (2009) for an extension to quantiles, and Blundell and Powell (2004) for an extension to binary response models. Since both approaches are complements rather than substitutes as the identification conditions do generally not imply each other, this thesis focuses on the use of control functions as means to identify different semi- and nonparametric regression models. In particular, the thesis contributes to the identification literature of nonlinear models under endogeneity by examining two cases where the correlation between regressor(s) and the model unobservable arises due to measurement error (chapter two) and simultaneous causality between reservation wage and elapsed unemployment duration (chapter three).

Specifically, chapter two studies nonclassical measurement error in the continuous dependent variable of a semiparametric, non-separable transformation model. The latter is a popular choice in practice nesting various nonlinear duration and censored regression models. The main complication arises because the (additive) measurement error is allowed to be correlated with a (continuous) component of the regressors as well as with the true, unobserved dependent variable itself. This problem has not yet been studied in the literature, but it is argued that it is relevant for various empirical setups with mismeasured, continuous survey data like earnings or durations. A framework to identify and consistently estimate (up to scale) the parameter vector of the transformation model is developed. The estimator links a two-step control function approach of Imbens and Newey (2009) with a rank estimator similar to Khan (2001) and is shown to have desirable asymptotic properties. Moreover, it is proven that 'm out of n' bootstrap can be used to obtain a consistent approximation of the asymptotic variance. The estimator's finite sample performance is studied in a Monte Carlo Simulation. To illustrate the empirical usefulness of the procedure, an earnings equation model is estimated using annual data from the Health and Retirement Study (HRS) and its results are compared to the ones of other estimators. Some evidence for a bias in the coefficients of years of education and age is found, emphasizing the importance to adjust for potential measurement error bias in empirical work.

Chapter three develops a test for monotonicity of a (possibly nonlinear) regression function under endogeneity. The novel testing framework is applied to study monotonicity of the reservation wage as a function of elapsed unemployment duration. Hence, the objective of the chapter is twofold: from a theoretical perspective, it proposes a test that formally assesses monotonicity of the regression function in the case of a continuous, endogenous regressor. This is accomplished by combining different nonparametric conditional mean estimators using either control functions or unobservable exogenous variation to address endogeneity with a test statistic based on a functional of a second order U-process. The modified statistic is shown to have a non-standard asymptotic distribution (similar to related tests) from which asymptotic critical values can directly be derived rather than approximated by bootstrap resampling methods. The test is shown to be consistent against general alternatives. From an empirical perspective, the chapter provides a detailed investigation of the effect of elapsed unemployment duration on reservation wages in a nonparametric setup. This effect is difficult to measure due to the simultaneity of both variables. Despite some evidence in the literature for a declining reservation wage function over the course of unemployment, no information about the actual form of this decline has yet been provided. Using a standard job search model, it is shown that monotonicity of the reservation wage function, a restriction imposed by several empirical studies, only holds under certain (rather restrictive) conditions on the variables in the model. The test from above is applied to formally evaluate this shape restriction and it is found that reservation wage functions (conditional on different characteristics) do not decline monotonically.

Finally, all proofs and empirical results are postponed to the corresponding Appendix of each chapter. Numerical computations are carried out in GAUSS (routines are available upon request).

2 Nonclassical Measurement Error in the Dependent Variable of a Nonlinear Model

2.1 Introduction

The paper considers identification and estimation of the parameter vector of the monotone transformation model (Han, 1987) when the continuous dependent variable is subject to nonclassical measurement error, where 'nonclassical' refers to a potential correlation of the measurement error with the true, unobserved dependent variable itself and a (continuous) component of the regressor vector. This setup is of interest from an empirical perspective as survey data is commonly subject to measurement error (Bound, Brown, and Mathiowetz, 2001). In particular for earnings and duration data, evidence suggests that nonclassical measurement error is the rule rather than the exception: Bricker and Engelhardt (2007) for instance study measurement error in matched (annual) earnings data of older workers in the Health and Retirement Study (HRS). Their findings suggest a strong negative ('mean-reverting') relationship between the extent of measurement error, defined as the difference between self-reported survey and administrative records, and 'true' administrative earnings. According to their results for men of the 1991 wave, measurement error falls by approximately \$100 for each additional \$1,000 in 'true' (administrative) annual labour income. In addition, measurement error is found to cause a substantial upwards bias in the effect of education on annual earnings. Cristia and Schwabish (2007) confirm both results using the Survey of Income and Program Participation (SIPP) Panel matched to administrative records.[1] In the

[1] In addition, the study also provides evidence for a correlation of measurement error with other demographic variables such as gender, age, or marital status.

duration context, Jäckle (2008) reveals a similar pattern for benefit recipient history data from the 'Improving Survey Measurement of Income and Employment' project, which employed (among other interviewing techniques) standard questioning methods from the British Household Panel Survey (BHPS) to infer about benefit receipts. Using a non-representative sample and a proportional hazard model, she finds that low educational attainment has a significant negative impact on the exit hazard of benefit income related spells with survey data, but not with administrative records. Moreover, under-reporting of the benefit duration (i.e. reporting a spell length that falls below the actual length) generally increases when a spell spans more than one survey wave.[2] Since durations and earnings typically serve as 'left-hand side' variables in standard censored regression (e.g. Tobit) or duration models, which can be nested within the monotone transformation model, both examples can be accomodated by the framework developed in this paper.[3] The paper addresses identification and estimation of the "parametric" parameter vector of this transformation function (up to scale). Future work will address the recovery of the unknown transformation function, too.

The main contribution of this paper is to provide the researcher with a tool to deal with nonclassical (as defined above) measurement error in continuous survey data such as earnings or durations if the model of interest is the parameter vector of the monotone transformation model (or any other model nested therein).[4] To the best of the author's knowledge, such a tool does not yet exist. The main theoretical complications in the identification and estimation process of the parameter vector arise because of (i) multiple unobservables (the measurement and the equation error) in the model setup, (ii) the "lack" of assumptions on the (conditional) measurement error distribution, and (iii) the potential dependence of the measurement error and continuous component(s) of the regressors. In order to ad-

[2]More precisely, she finds that the share of self-reported durations longer than a year that end at the 'seam' of two survey waves exceed the share of corresponding spells from administrative records by $25 - 35\%$ for certain benefit types.

[3]Notice that in order to apply the framework of this paper to the setup of Bricker and Engelhardt (2007), education needs to be modelled as a continuous variable ('years of schooling').

[4]The Proportional Hazard model can for instance be obtained by restricting the error term distribution (to a type I extreme value distribution), which allows for an interpretation of the transformation function as the integrated baseline hazard, while restricting the transformation function itself (to a logarithmic form) leads to the Accelerated Failure Time model.

dress these points, a three-step identification and estimation procedure is proposed: first, a two-step control function approach (see Blundell and Powell, 2003; Imbens and Newey, 2009) is employed to solve the 'endogeneity problem' arising from the dependence of the measurement error and a continuous component of the regressor vector by estimating the conditional mean of the (mismeasured) dependent variable conditional on all covariates and the estimated control function. Subsequently integrating over the marginal support of the control function eliminates its impact as a conditioning argument and reduces the measurement error to a numerical constant. In a third step, a rank-type argument is then used comparing pairs of observations to eliminate this numerical constant. Since a control function method is employed in the first place, the procedure requires the existence of a suitable instrument vector. Also, notice that in particular the "lack" of assumptions on the (conditional) distribution of the measurement error and the presence of multiple unobservables prohibit the use of other control function estimators such as Rothe (2009). Instead, all three steps outlined above are crucial for identification and consistent estimation of the parameter vector.

Finally, it is argued that for the examples given before, instrumental variables typically suggested by the empirical literature for Mincer-type earnings equations such as parental education, minimum school-leaving age, or (same sex) sibling's educational qualification should also be applicable in this context as they are likely to be correlated with the observed schooling level of the individual, but unlikely to affect the individual's actual response to the survey question.[5] Moreover, as pointed out by Hu and Schennach (2008) and discussed in section 2.2.1, the choice of instrumental variables in the context of measurement error could even comprise repeated measurements if certain conditions are met.

From a technical point of view, the main innovation of the paper is to combine a nonparametric mean estimator with a rank estimation procedure and to derive its asymptotic properties.[6] Since duration models are arguably one of the most

[5] Notice that this argument is valid even when measurement error is actually not related to cognitive ability but to other unobserved determinants.

[6] Concurrently to this work, Jochmans (2010) developed a two-step rank estimator for the monotone transformation model that is weighted by nonparametric control functions. While likely to be superior if the dimension of the covariate vector is large (with the conditional mean estimator

relevant application field of the transformation model in practice, the estimator is extended to allow for random right censoring. The additional estimation step required to accomodate censoring and to obtain the mean function further complicate the asymptotic variance expression, which depends on first and second order derivatives of certain conditional expectations. Thus, in order to construct confidence intervals for the parameter estimates, the use of 'm out of n' bootstrap is suggested to obtain corresponding standard errors and show their first order validity. Finally, to illustrate the methodology empirically, annual earnings data from the HRS is examined, which has been found to be subject to nonclassical measurement error (Bricker and Engelhardt, 2007). A reduced version of an earnings equation is estimated and it is found that the estimator differs substantially from other estimators obtained for comparison purposes. Together with evidence for a mean-reverting non-classical measurement error in annual earnings in the HRS (see Bricker and Engelhardt, 2007), this underlines the need to adjust for measurement error bias when examining the determinants of annual labour income of older workers in the HRS as estimates appear to be strongly affected.

This paper complements the existing literature on nonlcassical measurement error, which has been rather limited regarding measurement error in the response variable of nonlinear models. In the duration context for instance, researchers have limited attention to either fully parametric duration models or classical forms of measurement error (e.g. Skinner and Humphreys, 1999; Augustin, 1999; Abrevaya and Hausman, 2004; Dumangane, 2007), both of which are problematic once the restrictive setup fails to hold. A notable exception is the paper by Abrevaya and Hausman (1999), who consider nonadditive, classical measurement error in the dependent variable. However, relative to the approach proposed here, their setup cannot incorporate a correlation of the measurement error with the true, unobserved dependent variable itself, which often appears to be the more relevant problem in practice. Abstracting from the duration context, Chen, Hong, and Tamer (2005) have considered various semiparametric models under nonclassical measurement error (in the dependent as well as the independent variable(s)) using auxiliary admin-

being potentially less accurate), the procedure has not formally been extended to censoring and is thus of limited interest in the duration case, which is a main focus of this paper.

istrative data to infer about the conditional distribution of the true variables given the mismeasured variables. Matzkin (2007) examines a completely nonparametric framework, but her identification result hinges on the independence of the response error and other model (un-)observables. Hoderlein and Winter (2010) on the other hand use a structural approach to identify marginal effects of linear and nonlinear models under measurement error in either the dependent or the independent variable(s). While their methodology allows them to make detailed statements about the determinants and implications of such a measurement error, the validity of these claims clearly relies on the underlying model assumptions.

The paper is organised as follows: Section 2.2.1 outlines the identification strategy. Section 2.2.2 deals with the corresponding multi-step estimation procedure, its asymptotic distribution is derived in Section 2.2.3 and the validity of the bootstrapped confidence intervals is established in Section 2.2.4. Finally, Section 2.3 explores the finite sample properties in a small scale simulation study and Section 2.4 concludes with an empirical illustration on annual earnings data from the HRS Survey. All tables and proofs are postponed to the appendix.

2.2 Setup

2.2.1 Identification

The monotone transformation model (Han, 1987), which nests several duration and censored regression models, is given by:

$$Y_j^* = m(X_j'\beta_0 + \epsilon_j) \tag{2.1}$$

where Y_j^* is an unobserved, continuous scalar dependent variable, $X_j' = \{X_j^{(c)}, X_j^{(d)}\}'$ is a $(d_x \times 1)$-dimensional covariate vector with $X_j^{(c)}$ $\left(X_j^{(d)}\right)$ containing continuous (discrete) elements, and ϵ_j is a scalar unobservable (independent of X_j). $m(\cdot)$ is a strictly monotonic transformation function giving the model its flexibility and name.[7] Without loss of generality, this function is assumed to be strictly increasing in the following.

Into this setup, an additively separable, nonclassical measurement error η_j is incorporated, which is a scalar random variable:

$$Y_j = Y_j^* + \eta_j \tag{2.2}$$

'Nonclassical' here refers to a potential correlation of the measurement error with the true, underlying dependent variable and (continuous) component(s) of X_j. That is, letting "$\perp$" denote statistical independence of two random variables and "$\not\perp$" their dependence, the following assumptions are made:

- $\epsilon_j \not\perp \eta_j$ and
- $X_{1j} \not\perp \eta_j$, where $X_{1j} \in X_j^{(c)}$.

X_{1j} could possibly also represent a vector of continuous random variables for each of which a reduced form equation such as the one in A1 below holds (e.g. Blundell and Powell, 2004).[8] However, in order to maintain a tractable setup, X_{1j} will be assumed to be a scalar random variable in the following. By contrast, continuity of the endogenous component is crucial to the control function approach and cannot be relaxed (see below).

Regarding the additivity assumption of the measurement error, notice that if Y_j^* is a duration variable taking on positive values only, the expression in (2.2) can be viewed as the log-transformation of $\widetilde{Y}_j = \widetilde{Y}_j^* \cdot \widetilde{\eta}_j$, where both $\widetilde{Y}_j^*$ and $\widetilde{\eta}_j$ have support $[0, \infty)$ and $\widetilde{Y}_j^*, \widetilde{\eta}_j > 0$ except for a set of measure zero. The assumption of additive separability is hence not as restrictive as it might appear at first sight and has in fact been adopted by several authors in the literature (e.g Chesher, Dumangane, and Smith, 2002).

[7] Notice that the lack of restrictions on $m(\cdot)$ (apart from monotonicity) only allows for identification up to a location and size normalization (Sherman, 1993).

[8] Notice that the identification and estimation procedure of this paper is applicable even if, apart from the correlation with the measurement error η_j, $X_{1j} \not\perp \epsilon_j$. That is, as long as the instruments satisfy the independence requirement outlined in assumption A1 below, β_0 can be recovered even if identification and estimation are aggravated by e.g. an omitted variable.

Combining (2.1) and (2.2) yields the observed equation:

$$Y_j = m(X_j'\beta_0 + \epsilon_j) + \eta_j \tag{2.3}$$

The object is to identify β_0 from (2.3). To achieve this, the existence of an instrument vector $Z_j' = (X_{-1j}', Z_{1j}')$ is assumed, where X_{-1j}' refers to all exogenous elements except for X_{1j}:

A1 there exists a $(d_z \times 1)$-dimensional vector $Z_j' = (X_{-1j}', Z_{1j}')$ (with dimension(Z_{1j}) ≥ 1) such that

$$X_{1j} = g(Z_j) + V_j \tag{2.4}$$

with $g(\cdot)$ a real-valued function that is differentiable in its continuous components (with non-zero derivative), $\mathbb{E}[V_j] = 0$, and

$$Z_j \perp \epsilon_j, \eta_j, V_j$$

.

Condition A1 is the "exclusion restriction" typically imposed in the control function literature. It specifies that the correlation between X_{1j} and η_j only runs through a function V_j, the so called control function. As outlined before, continuity of X_{1j} is crucial in this context since, in the discrete case, the distribution of the control function V_j and its relation with η_j will in general depend on Z_j violating independence between Z_j and the model unobservables. Full independence of the instrument vector Z_j on the other hand is required since the model in (2.3) is not additively separable in observables and unobservables. Notice also that the setup in (2.4) allows for parametric or semiparametric restrictions: for instance, the researcher might specify a single-index model of the form $X_{1j} = g(Z_j'\gamma_0) + V_j$ with γ_0 an unknown vector of parameters and $g(\cdot)$ either an unknown or known differentiable function.

Concerning the empirical examples given in the introduction, instruments suggested in the context of (Mincer-type) earnings equations (Glewwe and Patrinos, 1999;

Butcher and Case, 1994; Card, 2001; Ichino and Winter-Ebmer, 1999) are applicable in the measurement error setup, too. However, in line with Hu and Schennach (2008), it is stressed that also a repeated measurement of Y_j^* could be understood as an instrument if it satisfied the independence assumption in A1. That is, if the second observation (together with the possible error contained in that alternative measurement) was independent of the measurement error η_j in the original Y_j conditional on the regressors X_j, the repeated measurement could become a valid instrumental variable (see Chalak and White (2011) for a detailed discussion of identification under various instrument concepts). Finally, since the setup is entirely nonparametric, it is well known that identification condition A1 does not imply nor is it implied by the moment conditions imposed in the nonlinear instrumental variable (NIV) literature.

The second condition required for identification is a "large support condition", which ensures sufficient variation in V_j given X_{1j}.[9]

A2 $\mathcal{W} = \mathcal{X} \times \mathcal{V}$ is a compact, non-empty set, where $\mathcal{X}$ is a subset in the interior of the marginal support of X, while $\mathcal{V}$ denotes the marginal support of V. Assume that the joint density on $\mathcal{W}$ is everywhere continuous and bounded away from zero.

Assumption A2 states that the marginal support of V_j is identical to its conditional support for a compact subset $\mathcal{X}$ of the marginal support of X. As discussed in Imbens and Newey (2009), this might be restrictive in applications where data is scarce or instrumental variables do not vary sufficiently as the above assumption basically requires sufficient strength of the latter. In practice, a verification of A2 can only be carried out approximately on a case by case basis. For instance, various Kolmogorov Smirnov tests on the conditional distributions of the estimated control function for subsets of the data used in the illustration example of section 2.4 indicate that the condition seems to be satisfied for at least a subset of the data. Still, condition A2 remains a drawback in the setup of this paper and future work will be directed

[9]Notice that a further support condition similar to Cavanagh and Sherman (1998) will ensure that identification is not lost by restricting ourselves to the compact set $\mathcal{W}$.

towards identifying sharp bounds similar to Imbens and Newey (2009).

The third assumption sufficient for identification of β_0 is a standard i.i.d. assumption:

A3 $\{X_j, Z_j, \epsilon_j, \eta_j\}_{j=1}^n$ is an i.i.d. sample, where Y_j and the endogenous component X_{1j} are generated according to (2.3) and (2.4), respectively.

In the following, let $\mu(x) := \int \mathbb{E}[Y_j|X_j = x, V_j = v] f_V(v) dv$ with $f_V(\cdot)$ the marginal density of V_j and recall that $m(\cdot)$ is strictly increasing in its argument. Given this setup, we obtain the following lemma, which ensures that the limit of the objective function, introduced in the next section, is uniquely maximized:

Lemma 1. *Under assumptions A1, A2, and A3 and given (2.3) and (2.4) with $m(\cdot)$ strictly increasing in its argument, we have for every $x, \widetilde{x} \in \mathcal{X}$:*

$$\mu(x) > \mu(\widetilde{x}) \quad \textit{if} \quad x'\beta_0 > \widetilde{x}'\beta_0$$

The proof of this lemma can be found in the appendix and proceeds in three steps: firstly, the mean of Y_j conditional on X_j and V_j is computed. Using conditional independence between η_j and X_j given V_j, the 'remainder term' becomes $\mathbb{E}[\eta_j|V_j]$. However, since no assumption about the distribution of $\mathbb{E}[\eta_j|V_j]$ such as $\mathbb{E}[\eta_j|V_j] = 0$ are made, an iterated expectations argument to obtain $\mathbb{E}[\eta_j]$ is subsequently applied by integrating over the support of V_j. That is, it is shown that for every $x \in \mathcal{X}$:

$$\mu(x) = \mathbb{E}[m(x'\beta_0 + \epsilon_j)] + \mathbb{E}[\eta_j] \tag{2.5}$$

where the expectation is taken w.r.t. ϵ_j and η_j, respectively. Notice that $\mathbb{E}[\eta_j]$ is 'reduced' to a numerical constant (which could be non-zero) and that $\mu(x)$, by the properties of $m(\cdot)$, is strictly increasing in $x'\beta_0$ for all $x \in \mathcal{X}$. The latter motivates the use of a rank-type argument (see Cavanagh and Sherman, 1998), which together with the i.i.d. assumption A3 allows for identification of β_0. That is, by A3 we have for every $x \in \mathcal{X}$ and $i, j \in 1, \ldots, n$:

$$\mathbb{E}[m(x'\beta_0 + \epsilon_j)] + \mathbb{E}[\eta_j] = \mathbb{E}[m(x'\beta_0 + \epsilon_i)] + \mathbb{E}[\eta_i]$$

Thus, given x an inequality will only arise for differing β-values. Moreover, it is clear from the above argument that the lack of structure on the transformation function only allows for point identification of β_0 in relative, not in absolute terms (that is, a normalization of β will be required). However, notice that if the researcher is willing to make parametric assumptions about the functional form of $m(\cdot)$, the above identification argument can be strengthened and point identification can be achieved even in absolute terms. Also, it becomes apparent that other estimators using control functions such as Rothe (2009) are not applicable here: the lack of information about $\mathbb{E}[\eta_j|V_j]$ does not make a "normalization" of this conditional expectation to zero innocuous, but further steps to identify the parameter vector of interest are required.

Finally, notice that in a standard linear model with $m(\cdot)$ equal to the identity function, the identification procedure becomes applicable to "nonclassical" measurement error in the independent variable, too. For instance, let:

$$Y_j = X_{1j} + X^*_{2j}\theta_0 + \epsilon_j$$

with $X_{2j} = X^*_{2j} + \eta_j$ and $\eta_j \not\perp X^*_{2j}$ as well as $\eta_j \not\perp \epsilon_j$. In this case, given a suitable instrument vector Z_j satisfying A1 and A2, it holds that: $\int \mathbb{E}\Big[Y_j\Big|X_{1j} = x_1, X_{2j} = x_2, V_j = v]f_V(v)dv = x_1 + x_2\theta_0 + \mathbb{E}\Big[\epsilon_j\Big] + \mathbb{E}\Big[\eta_j\Big]\theta_0$ so that an identical rank argument to above becomes applicable and θ_0 is identified up to scale.

2.2.2 Estimation

The three-step estimation procedure is immediate from the previous identification result:

(i) In a first step, $\widehat{V}_j$ is recovered from a nonparametric first-stage regression of X_{1j} on Z_j.

(ii) Then, $\mu(x,v) := \mathbb{E}[Y_j|X_j = x, V_j = v]$ can be estimated nonparametrically using $Y_j, X_j, \widehat{V}_j$ and the average: $\widehat{\mu}(x) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\mu}(x, \widehat{V}_i)$ for every $x \in \mathcal{X}$ can be computed.

(iii) Finally, a modified version similar to the two-step rank estimator of Khan (2001) can be used to recover β_0 (up to scale).

The last step is similar to a modified rank estimator of Khan's (2001), who uses an estimated conditional quantile function as transformation of the dependent variable. We replace this conditional quantile function and its estimator by the conditional mean $\mu(x)$ and $\widehat{\mu}(x)$, respectively. The replacement (together with the introduction of a control function and censoring) affects the asymptotic variance of our estimator, which will be different from the expression derived in Khan (2001), who does not address endogeneity or random right censoring in his setup.

The estimated control functions $\widehat{V}_j$ stem from the regression equivalent of (2.4):

$$\widehat{V}_j = X_{1j} - \widehat{g}(Z_j)$$

To estimate $g(\cdot)$, the Nadaraya-Watson estimator is proposed (for simplicity, assume that $d_z = 1$) with

$$\widehat{g}(Z_j) = \frac{\sum_{k=1}^{n} X_{1k}\mathbf{k}_h(Z_j - Z_k)}{\sum_{k=1}^{n} \mathbf{k}_h(Z_j - Z_k)}$$

where

$$\mathbf{k}_h(Z_j - Z_k) = \mathbf{k}\left(\frac{Z_j - Z_k}{h}\right)$$

and h is a deterministic sequence satisfying $h \longrightarrow 0$ as $n \longrightarrow \infty$, while $\mathbf{k}(\cdot)$ is a kernel function that satisfies the restrictions in B3 in Appendix A2.1.[10] An optimal bandwidth theory for the estimator is not developed in this paper, but instead standard rules of thumb are employed for the determination of the bandwidth in sections 2.3 and 2.4. Notice that $g(\cdot)$ could also be estimated by series estimators (splines, power series) or local linear smoothers, but the use of the Nadaraya Watson

[10] In practice, if some components of the instrument vector Z_j are discrete, nonparametric estimation of $g(\cdot)$ can proceed by splitting the sample according to the different values of the discrete component and estimating $g(\cdot)$ for each subsample separately.

estimator will facilitate several proofs in the appendix.This argument becomes even more important as the limiting distribution obtained in section 2.2.3 does not depend on the nonparametric first step estimators (a similar result was obtained by Newey (1994) for smooth objective functions with a nonparametric plug-in estimate).

The conditional mean function $\mu(x)$ can be estimated using again the Nadaraya-Watson kernel estimator. Since the d_x-dimensional covariate vector X_j contains d_c continuous elements and a univariate $\widehat{V}_j$, the following $d = (d_c + 1)$ dimensional product kernel is defined (for simplicity assume that: $h = h_1 = h_2 = ... = h_d$):

$$\mathbf{K}_{h,j}(x, v) = \mathbf{k}\left(\frac{x_1 - X_{1j}}{h}\right) \times \ldots \times \mathbf{k}\left(\frac{x_{d_c} - X_{d_c j}}{h}\right) \times \mathbf{k}\left(\frac{v - \widehat{V}_j}{h}\right)$$

and the following shorthand notation for the first d_x elements is introduced:

$$\mathbf{K}_h(x - X_j) = \mathbf{k}\left(\frac{x_1 - X_{1j}}{h}\right) \times \ldots \times \mathbf{k}\left(\frac{x_{d_c} - X_{d_c j}}{h}\right)$$

To bound the denominator away from zero and to ensure that observations lie within the compact set $\mathcal{W}$, a nonrandom trimming function is introduced:

$$I_{xi} := I[x \in \mathcal{X}, V_i \in \mathcal{V}] \quad \text{and} \quad \widehat{I}_{xi} := I[x \in \mathcal{X}, \widehat{V}_i \in \mathcal{V}]$$

Notice that for simplicity no random trimming is employed, but different trimming techniques might be used in practice.

Finally, in order for the estimator to become applicable in the duration context, the possibility of random (right) censoring is accomodated into the estimation procedure of the conditional mean by using the so called "synthetic data" approach of Koul, Susarla, and van Ryzin (1981).[11] As outlined in section 3.1, duration data is typcially subject to (random) right censoring. Instead of observing the mismeasured duration Y_j for each individual, one typically observes:

$$U_j = \min\{Y_j, C_j\} \quad \text{and} \quad \Delta_j = I\{Y_j \leq C_j\}$$

[11] The setup of this paper cannot straightforwardly be extended to fixed censoring. However, it is conjectured that fixed censoring could possibly be incorporated along the lines of Khan (2001) using a similar kind of weighting function.

where C_j is the censoring time and Δ_j a censoring indicator. We assume $\{C_j, \Delta_j\}$ to be independent of the other model covariates. This assumption, albeit debatable in some settings, is standard in the literature and often justified in practice. In addition, define:

$$U_{jG} = \frac{U_j \Delta_j}{1 - G(U_j-)}$$

and

$$U_{j\widehat{G}} = \frac{U_j \Delta_j}{1 - \widehat{G}(U_j-)}$$

where $G(\cdot-)$ is the left-continuous distribution function of C_j and $\widehat{G}(\cdot-)$ the corresponding Kaplan-Meier estimator (Kaplan and Meier, 1958) with $\widehat{H}(\cdot-)$ the left-continuous empirical distribution function of U_j:

$$\widehat{G}(c) = 1 - \prod_{i:C_i \leq c} \left(1 - \frac{\sum_{j=1}^{n} I[(1-\Delta_j) = 1, C_j \leq C_i]}{1 - \widehat{H}(U_i-)}\right)^{1-\Delta_i}$$

Replacing the partially unobserved Y_j by U_{jG}, Koul, Susarla, and van Ryzin (1981) showed that under condition B1 in appendix A.1:

$$\mathbb{E}[U_{jG}|X_j = x, V_j = v] = \mathbb{E}[Y_j|X_j = x, V_j = v] \tag{2.6}$$

Since U_{jG} is unobserved, we can replace it by $U_{j\widehat{G}}$ and estimate (2.6) as:

$$\widehat{\mu}(x, \widehat{V}_i) = \frac{\sum_{j=1}^{n} \widehat{I}_{xi} U_{j\widehat{G}} \mathbf{K}_{h,j}(x, \widehat{V}_i)}{\sum_{j=1}^{n} \widehat{I}_{xi} \mathbf{K}_{h,j}(x, \widehat{V}_i)} \tag{2.7}$$

while:

$$\widehat{\mu}(x) = \frac{1}{n} \sum_{i=1}^{n} \widehat{\mu}(x, \widehat{V}_i) \tag{2.8}$$

is the average of $\widehat{\mu}(x, \widehat{V}_i)$ over $\widehat{V}_i$. The last stage recovers the parameter vector β_0. As rank estimators only allow an identification of β_0 up to scale, a normalization of an arbitrary component of the parameter vector is required. Following standard procedures, the first component is normalized to one, i.e. $\beta(\theta) \equiv (1, \theta)$.[12] Thus, the

[12] Accordingly, the true parameter vector is $\beta(\theta_0) \equiv (1, \theta_0)$.

third stage rank estimator is given by:

$$\beta(\widehat{\theta}) = \arg\max_{\theta\in\Theta} \frac{1}{n(n-1)} \sum_{k\neq l} I[X_k \in \mathcal{X}] \times \widehat{\mu}(X_k) \times I[X_k'\beta(\theta) \geq X_l'\beta(\theta)] \qquad (2.9)$$

where $\sum_{k\neq l}$ stands for the double sum $\sum_{k=1}^{n}\sum_{l>k}^{n}$ assuming that observations are in ascending order.[13] The form of (2.9) is almost identical to the two-stage rank estimator of Khan (2001) using a conditional mean instead of a conditional quantile function. Notice that for the above estimator to work we require that $\widehat{\mu}(X_k) > 0$ for every X_k in $\mathcal{X}$. Thus, if Y_j also takes on negative values, an upfront transformation of the data needs to be carried out, e.g. $\overline{\overline{Y}}_j = Y_j - \min\{Y_1, \ldots, Y_n\}$, to ensure positivity.

2.2.3 Asymptotic Properties

This subsection considers the asymptotic properties of the estimation procedure. The probability limit of (2.9) evaluated at θ_0 is:

$$\int I[X_k \in \mathcal{X}] \times \mu(X_k) \times I[X_k'\beta(\theta_0) \geq X_l'\beta(\theta_0)] dF_X(X_k, X_l) \qquad (2.10)$$

where $F_X(\cdot,\cdot)$ in this case denotes the distribution function of X_k, X_l. Since the conditions for consistency, $\sqrt{n}$-consistency, and asymptotic normality are standard and rather lengthy (see Cavanagh and Sherman (1998) or Khan (2001) for details), the reader is referred to Appendix A.1 for details, where conditions B1 to B8 used in the theorems below are outlined together with a short discussion of non-standard assumptions. Notice that a higher order kernel function is employed in order to allow for a fairly large dimension of the covariate vector X_j. That is, with an increasing number of covariates used in the estimation of the conditional mean, a kernel function with an increasing number of moments equal to zero is required in order to control for the asymptotic bias.

[13] Summations appearing in the following that involve more than two indices will be defined according to the same logic.

Theorem 2. *Under conditions A1-A3, B1-B5, B7, and B8, we have:*

$$\widehat{\theta} \xrightarrow{p} \theta_0$$

The proof of Theorem 2 parallels the proof of Theorem 3.1 in Khan (2001). The main difference with respect to the latter is to show that replacing $\widehat{\mu}(X_k)$ by its probability limit $\mu(X_k)$ results in an error of smaller order for every $X_k \in \mathcal{X}$. Unlike Khan (2001), however, also the estimated terms $\widehat{V}_j$, $U_{j\widehat{G}}$, and $\widehat{I}_j$ need to be controlled for. One difficulty arises as the $\widehat{V}_j$ also enter the indicator function $\widehat{I}_j$, which in turn prevents a Taylor expansion. An argument from Corradi, Distaso, and Swanson (2011) is borrowed to show that this term can in fact be bounded by an expression approaching zero at rate $\ln(n)^{\frac{1}{2}}/(nh^{d_z})^{\frac{1}{2}} \longrightarrow 0$. Together with the convergence rates of $U_{j\widehat{G}}$ and $\widehat{V}_j$, the overall rate is:

$$\widehat{\mu}(x) - \mu(x) = O_p\left(\left(\frac{\ln(n)}{nh^{d_z}}\right)^{\frac{1}{2}}\right) = o_p(1)$$

for every $x \in \mathcal{X}$.

Given consistency of $\widehat{\theta}$ for θ_0, one can replace the parameter space Θ by a shrinking set around θ_0 to establish $\sqrt{n}$-consistency and asymptotic normality using results of Sherman (1993). To simplify notation in the next theorem, the following expression is defined (see Khan, 2001; Sherman, 1993):

$$\begin{aligned}\psi_1(x,\theta) = &\int \mu(x) \times I[x \in \mathcal{X}]I[x'\beta(\theta) > u'\beta(\theta)] - I[x'\beta_0 > u'\beta_0]dF_x(u) + \\ &\int \mu(u) \times I[u \in \mathcal{X}]I[u'\beta(\theta) > x'\beta(\theta)] - I[u'\beta_0 > x'\beta_0]dF_x(u)\end{aligned} \tag{2.11}$$

Moreover, denote:

$$\psi_2(x,\theta) = \int I[x \in \mathcal{X}]I[x'\beta(\theta) > u'\beta(\theta)]dF_x(u) \tag{2.12}$$

Theorem 3. *Under conditions A1-A3 and B1-B8, it holds that:*

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma)$$

where $\Sigma = J^{-1}\Omega J^{-1}$ *with:*

$$J = \frac{1}{2}\,\mathbb{E}\big[\nabla_{\theta\theta'}\psi_1(X_k,\theta_0)\big]$$

The diagonal elements of the matrix Ω *are given by the sum of the following expressions:*

(i)

$$\Omega_0 = \int \Big(I_m(U_{mG} - \mu(X_m))\nabla_\theta\psi_2(X_m,\theta_0)\Big) \times \Big(I_m(U_{mG} - \mu(X_m))\nabla_\theta\psi_2(X_m,\theta_0)\Big)^{'} dF_{U_G,X,V}(U_{mG},X_m,V_m)$$

(ii) $\Omega_1 = E_1\Phi_1E_1^{'}$ *with:*

$$\Phi_1 = \int V_i^2 dF_V(V_i)$$

and

$$E_1 = \Big(F_V^{(1)}\Big(a\Big) + F_V^{(1)}\Big(b\Big)\Big)\int U_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X}(U_{jG},X_k)$$

where a, b *are real numbers and* $F_V^{(1)}(\cdot)$ *denotes the first-order derivative of the distribution function* $F_V(\cdot)$ *of* V*.*

(iii) $\Omega_2 = E_2\Phi_2E_2^{'}$ *with*

$$\Phi_2 = \Phi_1$$

and

$$E_2 = -\int I_iU_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X,V}(U_{jG},X_k,V_i)$$

(iv) $\Omega_3 = E_3\Phi_3E_3^{'}$ *and*

$$\Phi_3 = \int_0^{\phi_Y} \mathbb{E}\Big[U_{1G}I[s < U_1]\Big]H_{t1}(s)\frac{dG(s)}{(1-G(s-))}$$

and

$$E_3 = \int I_i\nabla_\theta\psi_2(X_k,\theta_0)dF_{X,V}(X_k,X_l,V_i)$$

where ϕ_Y is defined in B1 and $H_{t1}(s) = \mathbb{E}\Big[U_{1G}I[s < U_1]\Big]/\{(1 - F_Y(s-))(1 - G(s-))\}$.

The proof of Theorem (3) follows the proof of Theorem 3.2 in Khan (2001). The conditions of Lemmata A.1 and A.2 therein are explicitly verified, which establish $\sqrt{n}$-consistency and asymptotic normality, respectively. The main differences to Khan (2001) consist in the use of a conditional mean rather than a conditional quantile function and in the estimated first and second stage terms $\widehat{V}_j$, $\widehat{I}_j$, and $U_{j\widehat{G}}$, which complicate the asymptotic analysis in this case further. Both, the estimation of the conditional mean function as well as the estimated $\widehat{V}_j$, $\widehat{I}_j$, $U_{j\widehat{G}}$ yield the extra pieces Ω_0, Ω_1, Ω_2, and Ω_3 in the variance-covariance matrix Σ that differ from the expression derived by Khan (2001). The first step in the proof of the above theorem is to replace $\widehat{\mu}(X_k)$ in (2.9) by $\mu(X_k)$. The term involving $\mu(X_k)$ can be expanded to yield the gradient $J = \frac{1}{2}\mathbb{E}\big[\nabla_{\theta\theta'}\psi_1(X_k, \theta_0)\big]$ plus terms that are of order $o_p(n^{-1})$ once $\sqrt{n}$-consistency of $\|\widehat{\theta} - \theta_0\|$ has been established (notice that Lemmata B.1 and B.2 are verified concurrently and hence expressions shown to be of order $o_p(\|\widehat{\theta} - \theta_0\|/\sqrt{n})$ for instance automatically become $o_p(n^{-1})$ once $\|\widehat{\theta} - \theta_0\| = O_p(1/\sqrt{n})$ has been established via Lemma B.1). The second term containing the estimation error $(\widehat{\mu}(X_k) - \mu(X_k))$ on the other hand can be further expanded to give the different variance pieces plus terms that are again of order $o_p(n^{-1})$ on a set around θ_0 shrinking at rate $\sqrt{n}$.

2.2.4 Bootstrapping Confidence Intervals

The asymptotic variance depends on moments of the derivatives of the unknown functions $\psi_1(\cdot, \cdot)$ and $\psi_2(\cdot, \cdot)$, which can be estimated using either numerical derivatives (e.g. Sherman, 1993; Cavanagh and Sherman, 1998) or kernel-based methods (Abrevaya, 1999). However, since these moments may be difficult to estimate in practice, the use of the 'm out of n' bootstrapping procedure is proposed as an alternative to construct standard errors for our parameter estimates. The 'm out of n' bootstrapping procedure is a widely applicable methodology allowing to approximate the sampling distribution under fairly weak assumptions. Moreover, this bootstrap method is able to replicate the degeneracy of first order terms from the linear U-statistic expansion (Arcones and Gine, 1992) that is used multiple times in the derivation of the asymptotic distribution of our estimator. The nonparametric 'n out of n' bootstrap method fails to replicate this degeneracy and hence an extension of the setup in Subbotin (2008), who recently showed that nonparametric 'n out of n' bootstrap methods consistenly estimate variances and quantiles of standard rank estimators, is not pursued in this paper.

The procedure works as follows: $X_1^*, \dots, X_m^*$ and $Z_1^*, \dots, Z_m^*$ are sampled from the original sample of size n (with $m < n$) and $\widehat{V}_1^*, \dots, \widehat{V}_m^*$ are obtained. In total, $1, \dots, B$ of these bootstrap samples of size m are constructed. For each of these samples, the bootstrap equivalent of our estimator is computed:

$$\beta(\theta^*) = \arg\max_{\theta \in \Theta} \frac{1}{m(m-1)} \sum_{k \neq l} I[X_k^* \in \mathcal{X}] \times \widehat{\mu}^*(X_k^*) \times I[X_k^{*\prime}\beta(\theta) \geq X_l^{*\prime}\beta(\theta)] \quad (2.13)$$

where

$$\widehat{\mu}^*(X_k^*) = \frac{1}{m} \sum_{i=1}^m \left\{ \frac{\sum_{j=1}^m \widehat{I}_{ki}^* U_{j\widehat{G}}^* \mathbf{K}_{h^*,j^*}(X_k^*, \widehat{V}_i^*)}{\sum_{j=1}^m \widehat{I}_{ki}^* \mathbf{K}_{h^*,j^*}(X_k^*, \widehat{V}_i^*)} \right\}$$

and the bandwidth sequence h^* is in lieu of h from Section **??** shrinking to zero at a rate depending on m (rather than n). Hence one obtains $\theta_1^*, \dots, \theta_B^*$. The aim is to construct a $1-\alpha$ confidence interval (CI) from the empirical bootstrap distribution. Thus, one needs to recover standard errors from the bootstrap covariance matrix, which is given by:

$$\Sigma^* = \frac{m}{B} \sum_{i=1}^B \left(\theta_i^* - \frac{1}{B} \sum_{i=1}^B \theta_i^* \right) \left(\theta_i^* - \frac{1}{B} \sum_{i=1}^B \theta_i^* \right)'$$

The next theorem establishes that Σ^* is a consistent estimator for Σ:

Theorem 4. *Let $\mathbb{P}_*$ denote the probability distribution induced by the bootstrap sampling. Under assumptions A1-A3 and B1-B8 with h^* and m in place of h and n, respectively, and letting $m, n, \frac{n}{m} \longrightarrow \infty$, it holds for all $\epsilon > 0$:*

$$\mathbb{P}\left(\omega : \mathbb{P}_* \left(\left| \Sigma^* - \Sigma \right| > \epsilon \right) \right) \longrightarrow 0$$

In order to prove the above theorem, it is firstly verified that $\sqrt{m}(\theta^* - \widehat{\theta})$ has the same limiting distribution as $\sqrt{n}(\widehat{\theta} - \theta_0)$ in a similar manner to before. However, since first order validity does not justify the use of the variance of the bootstrap distribution to consistently estimate the asymptotic variance (e.g. Goncalves and White, 2004), it is also shown that uniform integrability holds as well. A sufficient condition for the latter is the existence of a slightly higher moment condition, which in turn ensures consistency of the bootstrap variance estimator.

2.3 Monte Carlo Simulations

To shed some light on the small sample properties of the estimator in 2.9, various Monte Carlo simulations are conducted. The results are displayed in Table 2.1 and 2.2 of Appendix A2.3. The analysis starts by looking at a linear model under non-classical measurement error (as defined in Section 2.2.1) in the dependent variable. This allows to compare the performance of the estimator proposed in this paper relative to other estimators that are consistent (Two Stage Least Squares) or inconsistent (Monotone Rank Estimator, Ordinary Least Squares) in a linear setup.

More precisely, a linear model with two independent variables X_{1j} and X_{2j} is examined:

$$Y_j^* = X_{1j} + X_{2j}\theta_0 + \epsilon_j$$

with the coefficient of X_{1j} normalized to one and θ_0 set equal to .5. The additive measurement error η_j is given by:

$$Y_j = Y_j^* + \eta_j$$

The model unobservables ϵ_j and η_j are generated through a multivariate normal distribution:

$$\begin{pmatrix} \epsilon_j \\ \epsilon_j^M \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 & -.5 \\ -.5 & 1 \end{pmatrix} \right)$$

and the auxiliary equation $\eta_j = \kappa \cdot V_j + \epsilon_j^M$ with $\kappa = .5$. The negative correlation between ϵ_j and ϵ_j^M reflects the mean reverting correlation pattern observed in empirical studies (Bound, Brown, and Mathiowetz, 2001). X_{1j} is simulated from a uniform distribution $U[1,2]$, while X_{2j} is determined by the following reduced form model:

$$X_{2j} = \alpha \cdot Z_j + V_j$$

with $\alpha = 1$. The instrument Z_j and the control function V_j are simulated from two uniform distributions $U[0,1]$ and $U[-1,1]$, respectively. Notice that the chosen range of Z_j and V_j imply that V_j has full support given $0 \leq x_{2j} \leq 1$.

The estimator (labelled RankCF) proposed in section 2.2.3, which is consistent for θ_0, is compared to various other estimation procedures: the Two Stage Least Squares estimator (TSLS), which is also consistent in the linear model setup, is used to evaluate the relative performance of the RankCF in small samples. These results are contrasted with results from the inconsistent Ordinary Least Squares estimator (OLS) and the likewise inconsistent Monotone Rank Estimator (MRE) introduced by Cavanagh and Sherman (1998).[14] The latter has been chosen as it forms the basis for the RankCF and, like the RankCF, also requires an optimization algorithm due to the discontinuous character of the objective function. The chosen algorithm is the Nelder-Mead Simplex method with the normalized TSLS results as starting values. The sample size varies from 50, 100, 200, 400 to 800 observations. For every sample size, 200 replications are conducted. The displayed deviation measures are Mean Bias, Median Bias, Root Mean Square Error (RMSE), and Mean Absolute Deviation (MAD). They are constructed as averages over the number of replications. A second order Epanechnikov kernel is employed using the rule of thumb $std(\cdot) \cdot n^{-\frac{1}{7.5}}$ for the bandwidth selection, where $std(\cdot)$ is the standard deviation of the corresponding argument, while n remarks the sample size.

Starting with the simulation results in Table 2.1 (Design I: Linear Model & No Censoring), one can observe that even at small sample sizes TSLS and RankCF perform well across all bias measures (with a slight advantage for TSLS). Moreover, in line with consistency, their Mean Bias, RMSE, and MAD shrink as the sample size increases (albeit not gradually for the Mean Bias). This is not the case for the

[14] We use the identity function as 'weighting' function of the dependent variable.

MRE and OLS, where the mean bias is still of order .4 even at $n = 800$.

Next a non-linear design is examined. Using again $Y_j = Y_j^* + \eta_j$, the non-linear model is chosen to be:

$$Y_j^* = \ln(X_{1j} + X_{2j}\theta_0 + \epsilon_j)$$

with X_{2j}, and ϵ_j being determined as before, while $X_{1j} \sim U[2,3]$. Notice that in this nonlinear setup with nonclassical measurement error, all estimators except for the RankCF estiamtor are inconsistent either due to the non-linearity or due to the endogeneity of X_{2j}. OLS and TSLS will be dropped from the set of estimators and instead be replaced by the Maximum Rank Correlation Estimator, which was introduced by Han (1987) as the first estimator in the literature using a rank-type argument. The results are displayed in the lower part (Design II) of Table 2.1. Again, one can observe that the theoretical predictions are largely confirmed. Despite a relatively poor performance of all three estimators at $n = 50$, the bias of the MRE and the MRC remain substantial as n increases (albeit a certain decrease in the RMSE and MAD). This is not the case for the RankCF, where the mean bias decreases as the sample size increases (even though the bias has not entirely disappeared at $n = 800$).

Finally, the censoring setup in Table 2.2 is examined comparing our estimation procedure again to its rank competitors, the Monotone Rank Estimator (MRE) and the Maximum Rank Correlation Estimator (MRC). Notice that, in addition to the inconsistency because of the correlation between X_{2j} and η_j, the MRE as well as the MRC have not been formally extended to the case of random right censoring. To evaluate the relative performance of our censoring adjustment, we firstly carry out the simulations for the linear model of Design I without censoring (but $X_{2j} \sim U[2,3]$). To maximize the objective functions, we revert again to the above grid search method. The two censoring cases are also built upon the linear setup of Design II and contain two different average censoring ratios, .25 and .35. The censoring variable C_j is sampled from a uniform distribution $U[1,6]$ (Design IV) and $U[1,8]$ (Design V), respectively. Notice that the support of C_j 'covers' the support of Y_j in both cases so that only the degree of (right) censoring varies.

Turning to the simulation results, one can see that the effect of censoring drives up the biases particularly at small sample sizes. Somewhat surprisingly, the negative impact appears most pronounced for the method proposed in this paper, even though the deterioration slowly vanishes as the sample size increases. Despite this more negative effect of censoring, one can observe that, as expected from a theoretical perspective, the difference in mean and median bias is substantial for all sample sizes excelling the MRE and the MRC in particular for the case of 'light' censoring (Design IV). As the censoring ratio increases, all bias measures become fairly large. Once again, however, one observes a substantial improvement for the RankCF with the size of the sample growing, while the bias measures do only moderately change for the MRE and the MRC.

Overall, the results from this small simulation study indicate a good finite sample performance of the methodology for the chosen setups under different forms of nonclassical measurement error and various degrees of censoring.

2.4 Empirical Illustration

In a recent study, Bricker and Engelhardt (2007) provided empirical evidence for nonclassical measurement error in annual earnings data from the Health and Retirement Study, which is a nationally representative longitudinal survey of the over 50 population in the US.[15] The researchers found a mean-reverting pattern in the data and a significant positive correlation between higher education and measurement error. The mean measurement error (defined as the difference between self-reported HRS and matched administrative annual earnings) was found to be approximately \$1,500 with a standard deviation of \$13,899, which is substantial given that the mean of self-reported and administrative earnings stood at \$33,584 and \$32,071, respectively. The authors also established that for every additional \$1,000 in 'true' earnings, measurement error fell by \$100. Finally, men with a college degree or higher earned 49.2% more than high-school drop-outs based on reported earnings,

[15] See the University of Michigan's webpage http://hrsonline.isr.umich.edu/index.php for a detailed description of the study and the data (access date: 06/2010).

but only 42.1% more based on the matched administrative annual earnings. Unlike in the paper of Bricker and Engelhardt (2007), the 1998 wave is chosen, which also includes the 'War Babies' and the 'Children of the Great Depression' cohorts to broaden the age range in the data and to comply with the assumption of a continuous variable in the covariate vector. The sample is restricted to individuals with positive labour income during that year (i.e. no self-employed) and individuals that were the actual financial respondents of the household.[16] Moreover, to further ensure a certain degree of homogeneity, only white individuals are selelcted for the final dataset. The full support requirement in the assumption setup also meant that persons below the age of 50 and above 70, and those with less than 10 years of schooling were excluded.[17] The final sample size comprised 2,753 observations.

For the earnings equation, (natural) logarithm of annual labour income is taken to be the dependent variable and gender, age (as a proxy for experience), age squared, and years of schooling are considered as model covariates.[18] In a linear setup, using years of schooling as independent variable embeds the assumption of log earnings being a linear function of education, i.e. each additional year of education having the same proportional effect on expected annual earnings. Notice that this constraint does not apply to here though as the model setup allows for a nonlinear, monotonic transformation function. It does only apply to the interpretation of competing estimators imposing a linear model. The possibility of measurement error in the independent variables (which can certainly be put into question) is ruled out and the coefficient of gender is normalized to one as it is a well known result that being a male has a positive effect on earnings. As instruments for the respondent's years of schooling we choose years of schooling of the mother and the father, respectively. These family background covariates are typically correlated with the schooling level

[16]Annual labour income comprises (i) regular wage or salary income, (ii) bonuses, tips, commissions, extra-pay from overtime, (iii) professional practice or trade earnings, and (iv) other income earned from a second job or while in the military reserves.

[17]Various Kolmogorov Smirnov tests were carried out to compare the conditional distributions of the estimated control function residuals for different subsets of the data. The results of these tests indicate that the assumption of a full support is roughly satisfied for this range of the data.

[18]Notice that the use of three regressors plus an (estimated) control function requires the application of a third order kernel. Since simulation results with higher order kernel functions turned out to be less stable, we continue to employ the Epanechnikov kernel from Section 2.3 also in this empirical illustration.

of the individual, but unlikely to be related to the respondent's actual misreporting or his/ her ability.[19]

The estimation results are compared to the ones of the MRE and the MRC as well as a Least Squares (OLS), a Least Absolute Deviations (LAD), and a Two-Stage Least Squares (TSLS) estimator. The latter uses the mother's and the father's education as instrumental variables for the respondent's years of schooling and serves as an additional reference point for the education coefficient. Due to the discontinuous character of the objective function, a Nelder-Mead Simplex method is used to optimize the functions of the three rank estimators. As starting values for the initial simplex the OLS estimates are chosen.[20] To obtain a 90% confidence interval for the parameters, a 'm out of n' bootstrap with subsample size of 1.600 and 200 replications was conducted.

Examining the results in Table 2.3 in Appendix B, one observes that point estimates of age and age squared of our estimator (RankCF) lie amid the range of competing estimates from the MRE, MRC, OLS, LAD, and TSLS. This is in line with the finding of Bricker and Engelhardt (2007), who did not find a correlation between measurement error and age. Naturally, the use of first stage estimates in the final estimator (RankCF and TSLS) does come at the price of larger confidence regions. However, notice that the range of the confidence bands is fairly similar for the TSLS and our estimation procedure, and all point estimates still appear to be significant at a 10% level. Turing to the coefficient of interest, the education coefficient of our estimator differs from its competitiors and falls substantially below their values hinting at an upwards bias in the education coefficient of the other estimators. It can of course not be established whether the size reduction in the estimated education coefficient can be attributed to an elimination of the measurement error or the standard ability bias (by standard arguments, one would expect the abilitiy bias to be positive, which corresponds to the direction of the measurement error bias as found by Bricker and Engelhardt (2007)). However, the relatively unchanged

[19]Despite criticism in the literature about the suitability of parental education as an instrumental variable for childrens' education, it is deemed that these variables still serve the purpose of this small scale illustration.

[20]Notice that the results were rather insensitive to small variations in the initial simplex, e.g. changing to an average of OLS and LAD or the TSLS estimates.

TSLS estimate of the education coefficient suggests that measurement error might be the reason for the drop in size. This conjecture is supported by the observation that there are no substantial differences between the estimates of OLS and LAD on one hand and MRE and MRC on the other, which suggests that a violation of the linearity restriction is unlikely to be the driving force behind the difference between the TSLS and the RankCF result.

Summarizing this small illustrative example that looks at a log earnings equation with years of education, gender, age, and age squared as covariates using the 1998 wave of the HRS, it is found that point estimates for the education coefficient provided by the estimation procedure proposed in this paper differ quite substantially from those of its competitors. Moreover, since the age coefficient of the estimatior of section 2.2.3 is largely in line with the values obtained from the other estimators (confirming Bricker and Engelhardt (2007), who did not find a substantial correlation of measurement error with other characteristics such as age), the illustrative results hint at the presence of measurement error bias in standard earnings equation regressions based on the HRS. Together with the empirical evidence of Bricker and Engelhardt (2007) for a mean-reverting non-classical measurement error in annual earnings that is correlated with education (Bricker and Engelhardt, 2007) from the 1992 wave, this underlines the need to adjust for measurement error bias when examining the determinants of annual labour income of older workers in the HRS.

2.5 Conclusion

This paper proposes a multi-step procedure to identify and estimate the parameter vector of the monotone transformation model when the continuous dependent variable is subject to nonclassical measurement error. Empirical evidence examining duration and earnings data collected via survey questionnaires often suggests that such a measurement error represents the rule rather than the exception. Taking on a reduced form perspective, a methodology to address measurement error when the researcher does not dispose of any information about the underlying distribution of either the true dependent variable or the measurement error is developed, but he or she only has a suspicion about the correlation pattern of the latter. Combining a modified control function approach, which requires the existence of a suitable instrumental variable vector, with a rank-type argument, it is shown that it is possible to recover the aforementioned parameter vector consistently up to a location and size normalization. We derive the estimator's asymptotic properties and also demonstrate the methodology's good finite sample performance in a small Monte Carlo Study. Finally, an empirical illustration investigating the effect of years of schooling on annual (log) earnings data from the Health and Retirement Study concludes this paper. Substantially different point estimates are found using our estimation procedure (relative to other estimators) suggesting that to account for correct inference is important in this context.

Extensions of the present paper and topics for future research include the nonparametric recovery of the unknown transformation function $m(\cdot)$, which requires a point identification result for the parameter vector hence providing another motivation for the asymptotic result derived in this paper. Being able to identify and nonparametrically estimate the transformation function is of particular interest in survival analysis, where the function is typically labelled 'integrated baseline hazard' and of substantial importance for policy analysis purposes. Alternatively, in contexts where the large support assumption appears to be unjustifiable, the researcher might instead be interested in abandoning the goal of point identification in favour of sharp bounds on the parameter vector. Such an extension was considered by Imbens and Newey (2009) in a similar setup and represents an important future extension of the present work.

The case of measurement error in multiple spell duration models appears to be another important area of future research, too: despite suitable stationarity assumptions on the measurement error (similar to the ones used in Abrevaya (2000) for the idiosyncratic error terms), such an extension is more complex as 'fixed effects' estimators typically exploit 'intra-unit' variation rendering the integration over the support of the control function more difficult. Finally, a last field of interest might be the case of binary dependent variables: duration models in discrete time are often set up using binary dependent variables, where the variable in a particular period takes on the value zero if the spell is on-going and one if the spell fails. Thus, falsely reported or recorded responses turn the nonclassical measurement error into a misclassification rather than a measurement problem, which is non-trivial due to the nonlinear nature of the underlying model (Hausman, Abrevaya, and Scott-Morton, 1998).

A2 Appendix

A2.1 Assumptions

Let $\|\cdot\|$ denote the Euclidean norm and ∇_i the i-th order derivative of a function.

B1 C_j is i.i.d. and independent of Y_j. Moreover, C_j satisfies:

- **(i)** $\mathbb{P}[C_j \leq Y_j | Y_j = y, X_j = x, V_j = v] = \mathbb{P}[C_j \leq Y_j | Y_j = y]$.
- **(ii)** $G(\cdot)$ is continuous.
- **(iii)** $\phi_Y \leq \phi_C$
  with $\phi_Y = \inf\{t : F_Y(t) = 1\}$, $\phi_C = \inf\{t : G(t) = 1\}$, and $F_Y(t) = \mathbb{P}[Y_j \leq t]$, $G(t) = \mathbb{P}[C_j \leq t]$.
- **(iv)** When $\phi_Y < \phi_C$, $\lim\sup_{t\to\phi_Y} \left(\int_t^{\phi_Y}(1 - F_Y(s))dG(s))^{1-\rho}/(1 - F_Y(t))\right) < \infty$, for some $\frac{2}{5} < \rho < \frac{1}{2}$.
- **(v)** When $\phi_Y = \phi_C$, for some $0 \leq \varsigma < 1$, $(1 - G(t))^{\varsigma} = O((1 - F_Y(t-)))$ as $t \to \phi_Y$.
- **(vi)** Let $F_U(t) = \mathbb{P}[U_j \leq t]$ and $H(U_j) = \int_{-\infty}^{U_j} dG(s)/(\{1 - F_U(s)\}\{1 - G(s)\})$. Assume that:

$$\int U_j H^{\frac{1}{2}+\varepsilon}(U_j)[1 - G(U_j-)]^{-1} dF_{U,X,V}(U, X, V) < \infty$$

B2 The elements x in the support of X can be partitioned into subvectors of discrete $x^{(d)}$ and continuous $x^{(c)}$ components. Let $\mathcal{X}^{(d)}$ and $\mathcal{X}^{(c)}$ be the corresponding discrete and continuous parts of $\mathcal{X} \subset \mathcal{W}$. Assume that the conditional density (given $x^{(d)} \in \mathcal{X}^{(d)}$) on $\mathcal{W}$ is everywhere continuous and strictly bounded away from zero. Moreover, assume that $\mathcal{X}$ is not contained in any proper linear subspace of $\mathbb{R}^{d_x}$ and that the subset $\mathcal{X}_{(1)}$ of one component of the d_x-dimensional set $\mathcal{X} = \mathcal{X}^{(d)} \times \mathcal{X}^{(c)}$ contains the interval:

$$\left[\mu(x) - 3\max_{x'_{(-1)}\theta}|x'_{(-1)}\theta| \quad ; \quad \mu(x) + 3\max_{x'_{(-1)}\theta}|x'_{(-1)}\theta|\right]$$

for any $x \in \mathcal{X}$, where $x_{(-1)}$ denotes the remaining $(d_x - 1)$ dimensional component and the maximum is taken over $\mathcal{X}_{(-1)} \times \Theta$ with $\max_{x'_{(-1)}\theta}|x'_{(-1)}\theta| < \infty$.

B3 The multivariate kernel function $\mathbf{K} = k \times \ldots \times k$ with $\mathbf{K} : \mathbb{R}^d \longmapsto \mathbb{R}$ is symmetric, has compact support, and is differentiable (with bounded derivative). In addition, $\mathbf{K}(\cdot)$ satisfies (i) $\int \mathbf{K}(u)du = 1$, (ii) $\int \mathbf{K}(u)u^{\gamma}du = 0$ for $\gamma = 1, \ldots, r-1$, (iii) $\int \mathbf{K}(u)u^r du \neq 0$ and $\int \mathbf{K}(u)u^r du < \infty$, (iv) $\int |\mathbf{K}(u)|du < \infty$, and (v) $\int \mathbf{K}^2(u)du < \infty$.

B4 θ_0 lies in the interior of the parameter space Θ, a compact subset of $\mathbb{R}^{d-1}$.

B5 For any value $x^{(d)} \in \mathcal{X}^{(d)}$, assume that $\mu(\cdot)$ is twice differentiable in $x^{(c)}$. In addition, given $0 < \gamma \leq 1$ and $\delta_0 > 0$, for every $x_1^{(c)}, x_2^{(c)} \in \mathcal{X}^{(c)}$ and $i = 0, 1, 2$:

$$\left\|\nabla_i \mu(x_1^{(c)}, x^{(d)}) - \nabla_i \mu(x_2^{(c)}, x^{(d)})\right\| \leq \delta_0 \left\|x_1^{(c)} - x_2^{(c)}\right\|^{\gamma}$$

where ∇_i denotes the order of derivative w.r.t $x^{(c)}$.

B6 Let $\psi_1(x, \theta)$ and $\psi_2(x, \theta)$ be defined as in (2.11) and (2.12):

- For each x in $\mathcal{X}$, $\psi_1(x, \cdot)$ is twice differentiable with second order Lipschitz derivative.
- $\mathbb{E}[\nabla_{\theta\theta'} \psi_1(\cdot, \theta_0)]$ is negative definite.
- For each $x \in \mathcal{X}$, $\psi_2(x, \cdot)$ is twice continuously differentiable in the second argument.
- Let $\nabla_i f(\cdot, \cdot)$ denote the order of derivative of $f_{X,V}(\cdot, \cdot)$ w.r.t. the first argument: assume that $\mathbb{E}[\|U_G \nabla_\theta \psi_2(X, \theta) \nabla_i f_{X,V}(X, V)\|^{2+\delta}] < \infty$ and $\mathbb{E}[\|U_G \nabla_{\theta\theta'} \psi_2(X, \theta) \nabla_i f_{X,V}(X, V)\|^{2+\delta}] < \infty$ for some $\delta > 0$ and all $\theta \in \Theta$.

B7 Assume that $\mathbb{E}\Big[V^{2+\delta}\Big] < \infty$, $\mathbb{E}\Big[\mu(x, V)^{2+\delta}\Big] < \infty$, $\mathbb{E}\Big[\|U_j / \{1 - G(U_j)\}^2\|^{2+\delta}\Big] < \infty$ and $\mathbb{E}\Big[\|U_G f_{X,V}(X, V)\|^{2+\delta}\Big] < \infty$ for some $\delta > 0$. Moreover, suppose that $F_V(\cdot)$ is continuously differentiable in its argument for every $V \in \mathcal{V}$.

B8 Let $d_z \leq d \leq r + \frac{1}{2} d_z$ (note that r is defined in B3). For $d < 3$, the bandwidth sequence h satisfies:

$$nh^{d+3} \longrightarrow 0 \qquad \text{and} \qquad (n^{\frac{1}{2}} h^{\frac{2r+d_z}{2}}) / \ln(n)^{\frac{1}{2}} \longrightarrow \infty$$

and for $d \geq 3$:

$$nh^{2d} \longrightarrow 0 \qquad \text{and} \qquad (n^{\frac{1}{2}} h^{\frac{2r+d_z}{2}}) / \ln(n)^{\frac{1}{2}} \longrightarrow \infty$$

.

Remark 1: B1 (i) together with the independence assumption of C_j are sufficient for the equality in (2.6). Condition (iii) ensures that we observe the entire distribution and, in combination with (iv) and (v), is relevant for the estimation of $G(U_j-))$ (see Lu and Cheng (2007) for details). The parameter ρ is determined by the "heaviness" of censoring, i.e. the smaller ρ, the fewer uncensored observations actually lie close to the "endpoint" ϕ_C. Finally, (vi) is a square integrability condition used in the proof of Theorem 3.

Remark 2: Assumption B2 extends condition A2 in the text, allowing also for discrete components in the parameter vector. The latter part of the condition ensures that identification is not lost by restricting ourselves to a compact subset of the support. That is, it is assumed that the set $\mathcal{X}_1$ of one regressor is sufficiently large (relative to the others), see Khan (2001) for details.

Remark 3: The requirement $d_z \leq d \leq r + \frac{1}{2} d_z$ of the bandwidth condition B8 allows to neglect the bias of the higher order kernel defined in B3. For a five dimensional instrument vector Z_j ($d_z = 5$)

and a five dimensional covariate vector (four regressors plus the estimated control function $\widehat{V}_j$) as in the empirical illustration of section 2.4 for instance, we thus require the use of a third order kernel function to meet the above restriction and to be able to negelect the bias in the asymptotic distribution.

A2.2 Proofs

Notice that it is implicitly understood that whenever $\mathbf{K}_{h,j}(\cdot,\cdot)$ is evaluated at $\widehat{V}_i$, we sum over $\widehat{V}_j$, while if the kernel function is evaluated at V_i, we sum over V_j. Moreover, we will suppress the dependency of $I[X_k \in \mathcal{X}, V_i \in \mathcal{V}]$ on X_k in the following and write the indicator function as I_i.

Proof of Lemma 1

By A1, we have that $Z_j \perp \eta_j, \epsilon_j, V_j$. Since Z_j is independent of both V_j and η_j, it follows by standard arguments:

$$Z_j \perp \eta_j | V_j$$

Moreover, since $X_j = g(Z_j) + v$ given $V_j = v$ is a function of Z_j only, this implies:

$$X_j \perp \eta_j | V_j$$

By identical arguments and using the fact that ϵ_j is independent of V_j, we can also establish that $X_j \perp \epsilon_j | V_j$. Hence, we obtain:

$$\begin{aligned}
\mathbb{E}[Y_j|X_j = x, V_j = v] &= \mathbb{E}[m(X_j'\beta_0 + \epsilon_j) + \eta_j | X_j = x, V_j = v] \\
&= \mathbb{E}[m(x'\beta_0 + \epsilon_j)|X_j = x, V_j = v] + \mathbb{E}[\eta_j|X_j = x, V_j = v] \\
&= \mathbb{E}[m(x'\beta_0 + \epsilon_j)|V_j = v] + \mathbb{E}[\eta_j|V_j = v]
\end{aligned}$$

where the last equality follows by conditional independence. Using the argument of Blundell and Powell (2003) or Imbens and Newey (2009), by condition A2 we can integrate over the marginal distribution of V and apply iterated expectations to obtain:

$$\begin{aligned}
\int \mathbb{E}[Y_j|X_j = x, V_j = v] f_V(v) dv &= \int \mathbb{E}[m(x'\beta_0 + \epsilon_j)|V_j = v] f_V(v) dv + \int \mathbb{E}[\eta_j|V_j = v] f_V(v) dv \\
&= \mathbb{E}[m(x'\beta_0 + \epsilon_j)] + \mathbb{E}[\eta_j]
\end{aligned}$$

for each $x \in \mathcal{X}$. The result then follows by A3 and the strict monotonicity of $m(\cdot)$. That is, for two observations i, j $(i \neq j)$ with $x, \widetilde{x} \in \mathcal{X}$:

$$\mathbb{E}[m(x'\beta_0 + \epsilon_i)] + \mathbb{E}[\eta_i] > \mathbb{E}[m(\widetilde{x}'\beta_0 + \epsilon_j)] + \mathbb{E}[\eta_j] \quad \text{if} \quad x'\beta_0 > \widetilde{x}'\beta_0$$

Hence, the result follows. $\blacksquare$

Proof of Theorem 2

Using the same steps as in Theorem 3.1 of Khan (2001) and Lemma A1 below, the result follows instantly. ■

Lemma A1. Given B1 to B5, B7, and B8, we have that:

$$\widehat{\mu}(x) - \mu(x) = O_p\left(\left(\frac{\ln(n)}{nh^{d_z}}\right)^{\frac{1}{2}}\right) = o_p(1)$$

for every $x \in \mathcal{X}$.

Proof of Lemma A1

Notice that:

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n}\widehat{\mu}(x,\widehat{V}_i) - \mu(x)\right| \leq & \left|\frac{1}{n}\sum_{i=1}^{n}\widehat{\mu}(x,\widehat{V}_i) - \frac{1}{n}\sum_{i=1}^{n}\widetilde{\mu}(x,V_i)\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{\mu}(x,V_i) - \mu(x)\right| \\ = & L_{1n} + L_{2n}\end{aligned} \tag{A-2.1}$$

where $\widetilde{\mu}(x,\cdot)$ is defined as (2.8) in the text with V_j, I_i, U_{jG} replacing $\widehat{V}_j, \widehat{I}_i, U_{j\widehat{G}}$. We start with L_{1n}, which can be decomposed as:

$$\begin{aligned}L_{1n} = & \left|\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\widehat{s}_{x,V}(x,\widehat{V}_i) - \widetilde{s}_{x,V}(x,V_i)}{\widetilde{f}_{x,V}(x,V_i)} - \frac{\widetilde{f}_{x,V}(x,V_i) - \widehat{f}_{x,\widehat{V}}(x,V_i)}{\widetilde{f}_{x,V}(x,V_i)} \times \widehat{\mu}(x,\widehat{V}_i)\right\}\right| \\ = & L_{11n} + L_{12n}\end{aligned} \tag{A-2.2}$$

where

$$\widehat{s}_{x,\widehat{V}}(x,\widehat{V}_i) = \frac{1}{nh^d}\sum_{j=1}^{n}\widehat{I}_i U_{j\widehat{G}}\mathbf{K}_{h,j}(x,\widehat{V}_i) \tag{A-2.3}$$

and

$$\widehat{f}_{x,\widehat{V}}(x,\widehat{V}_i) = \frac{1}{nh^d}\sum_{j=1}^{n}\widehat{I}_i\mathbf{K}_{h,j}(x,\widehat{V}_i) \tag{A-2.4}$$

with $\widetilde{f}_{x,V}(x,V_i)$ and $\widetilde{s}_{x,V}(x,V_i)$ defined analoguously using U_{jG}, I_i, V_j, respectively (recall that $d = d_x + 1$). We examine L_{11n} first. This term can be further decomposed to tackle the random denominator:

$$L_{11n} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\widehat{s}_{x,V}(x,\widehat{V}_i) - \widetilde{s}_{x,V}(x,V_i)}{f_{x,V}(x,V_i)} + \left[\frac{1}{\widetilde{f}_{x,V}(x,V_i)} - \frac{1}{f_{x,V}(x,V_i)}\right](\widehat{s}_{x,V}(x,\widehat{V}_i) - \widetilde{s}_{x,V}(x,V_i))\right\}$$

By B3 and B8, the second term is of smaller order since $\sup_{x,V\in\mathcal{W}}\left|\widetilde{f}_{x,V}(x,V) - f_{x,V}(x,V)\right| = O_p((\ln(n)/nh^d)^{\frac{1}{2}}) = o_p(1)$ with $f_{x,V}(x,V_i)$ denoting the true density evaluated at x, V_i. As for the first term, $f_{x,V}(x,V)$

is strictly bounded away from zero for all $x \in \mathcal{X}$ and $V \in \mathcal{V}$ by B2. A decomposition of the first term of L_{11n} yields:

$$\begin{aligned}
&\Big|\frac{1}{n^2h^d}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i - I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i)\Big| \\
+&\Big|\frac{1}{n^2h^d}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{f_{x,V}(x,V_i)}I_iU_{jG}\mathbf{K}_h(x-X_j)\big\{k_h(\widehat{V}_i-\widehat{V}_j)-k_h(V_i-V_j)\big\}\Big| \\
+&\Big|\frac{1}{n^2h^d}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{f_{x,V}(x,V_i)}I_i\big(U_{j\widehat{G}}-U_{jG}\big)\mathbf{K}_{h,j}(x,V_i)\Big| \\
+&o_p(1)
\end{aligned} \tag{A-2.5}$$

where $o_p(1)$ captures terms of smaller order containing cross-products. Denote the first, second, and third term as L_{111n}, L_{112n}, and L_{113n}, respectively. We examine each of these terms separately, starting with L_{111n}. Notice that by A3, B2, B8, and standard arguments one can show that:

$$\max_{1\le j\le n}|\widehat{V}_j - V_j| = O_p\left(\left(\frac{\ln(n)}{nh^{d_z}}\right)^{\frac{1}{2}}\right)$$

Noting that $|\widehat{I}_i - I_i| = |I[a \le \widehat{V}_i \le b] - I[a \le V_i \le b]|$, we can use the same argument as in Lemma A3 of Newey, Powell, and Vella (1999) to show that for $\Delta_n = ((\ln(n)/nh^{d_z})^{\frac{1}{2}})$ we have:

$$\begin{aligned}
|\widehat{I}_i - I_i| =& |I[a \le V_i + (\widehat{V}_i - V_i) \le b] - I[a \le V_i \le b]| \\
\le& \big(I[|V_i - a| \le \Delta_n] + I[|V_i - b| \le \Delta_n]\big)
\end{aligned}$$

for $1 \le i \le n$. Turning back to L_{111n}, this term can be expanded as:

$$\begin{aligned}
L_{111n} \le& \Big|\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i - I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i)\Big]\Big| \\
&+\Big|\frac{1}{n^2h^d}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big\{\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i - I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i) \\
&-\mathbb{E}\Big[f_{x,V}(x,V_i))\big(\widehat{I}_i - I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i)\Big]\Big\}\Big| \\
=& M_1 + M_2
\end{aligned}$$

We consider M_1 first. Using the positivity of U_{jG}:

$$\begin{aligned}
M_1 \le& \mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{x,V}(x,V_i)}|\widehat{I}_i - I_i|U_{jG}|\mathbf{K}_{h,j}(x,V_i)|\Big] \\
\le& \mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{x,V}(x,V_i)}\big(I[|V_i - a| \le \Delta_n] + I[|V_i - b| \le \Delta_n]\big)U_{jG}|\mathbf{K}_{h,j}(x,V_i)|\Big]
\end{aligned}$$

We examine only the first term, the second one follows by identical arguments. Setting $a = 0$ without loss of generality and letting $u_1 = ((x - X_j)/h)$, $u_2 = ((V_i - V_j)/h)$, and $f_V(\cdot)$ denote the density of V_i and V_j, after change of variables we obtain:

$$\begin{aligned}
&\int\int\int\int_0^{\Delta_n} U_{jG}|\mathbf{K}_h(u_1)k_h(u_2)|\frac{f_{X,V}(x+u_1h, V_j+u_2h)}{f_{X,V}(x, V_j+u_2h)} f_{X,V}(x,V_j)dV_j du_1 du_2 dF_{U_G}(U_G) \\
=&\int\int_0^{\Delta_n} U_{jG}|\mathbf{K}_h(u_1)k_h(u_2)|f_{X,V}(x,V_j)dV_j dF_{U_G}(U_G)(1+O(h)) \\
=&O(\Delta_n)
\end{aligned}$$

Next we consider M_2. The variance of this term is given by:

$$\mathbb{E}\Big[\Big(\frac{1}{n^2h^d}\sum_{i=1}^n\sum_{j=1}^n\Big\{\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i - I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i) - \mathbb{E}\Big[\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i - I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i)\Big]\Big\}\Big)^2\Big] + O(\Delta_n^2)$$

The first expectation above can be dealt with in the same way as before. This yields:

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\frac{1}{n^2h^d}\sum_{i=1}^n\sum_{j=1}^n\Big\{\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i - I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i) - \mathbb{E}\Big[\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i - I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i)\Big]\Big\}\Big)^2\Big] \\
=&O\Big(\frac{1}{n^2h^d}\Delta_n\Big) + O\Big(\Delta_n^2\Big)
\end{aligned}$$

Using Chebychev's inequality and B8, $M_2 = o_p(\Delta_n)$, so the overall rate becomes:

$$L_{111n} = O_p\left(\left(\frac{\ln(n)}{nh^{d_z}}\right)^{\frac{1}{2}}\right)$$

Next we examine the second term of (A-2.5), L_{112n}. A mean value expansion around $(V_i - V_j)$ yields:

$$L_{112n} = \left|\frac{1}{n^2h^{d+1}}\sum_{i=1}^n\sum_{j=1}^n \frac{1}{f_{x,V}(x,V_i)} I_i U_{jG}\mathbf{K}_h(x-X_j)k_h^{(1)}(\overline{V}_i - \overline{V}_j)\big((\widehat{V}_i - V_i) + (V_j - \widehat{V}_j)\big)\right|$$

where $\overline{V}_i, \overline{V}_j$ denote intermediate values and $k^{(1)}(\cdot)$ is the derivative of the kernel function w.r.t. its argument. We can rewrite the expression as:

$$\left|\frac{1}{n^2h^{d+1}}\sum_{i=1}^n\sum_{j=1}^n \frac{1}{f_{x,V}(x,V_i)} I_i U_{jG}\mathbf{K}_h(x-X_j)k_h^{(1)}(\overline{V}_i - \overline{V}_j)\big(\big(\widehat{g}(Z_i) - g(Z_i)\big) + \big(g(Z_j) - \widehat{g}(Z_j)\big)\big)\right|$$

Since $(\widehat{g}(Z_i) - g(Z_i))$ and $(g(Z_j) - \widehat{g}(Z_j))$ are identical, we only examine the first term involving $(\widehat{g}(Z_i) - g(Z_i))$. Letting $\mathbf{K}^{(1)}_{h,j}(x,\overline{V}_i) = \mathbf{K}_h(x - X_j) \times k_h^{(1)}(\overline{V}_i - \overline{V}_j)$, we can decompose the first term into:

$$
\begin{aligned}
L_{112n} \leq & \left| \frac{1}{nh^{d+1}} \sum_{i=1}^{n} \mathbb{E}\Big[\frac{1}{f_{x,V}(x,V_i)} I_i U_{jG} \mathbf{K}_{h,j}^{(1)}(x,\overline{V}_i) \Big] \times \Big(\widehat{g}(Z_i) - g(Z_i) \Big) \right| + \\
& \left| \frac{1}{n^2 h^{d+1}} \sum_{i=1}^{n} \sum_{j=1}^{n} \Big(\frac{1}{f_{x,V}(x,V_i)} I_i U_{jG} \mathbf{K}_{h,j}^{(1)}(x,\overline{V}_i) - \mathbb{E}\Big[\frac{1}{f_{x,V}(x,V_i)} I_i U_{jG} \mathbf{K}_{h,j}^{(1)}(x,\overline{V}_i) \Big] \Big) \Big(\widehat{g}(Z_i) - g(Z_i) \Big) \right| \\
= & N_{1n} + N_{2n}
\end{aligned}
$$

We start with N_{1n}. The expectation expression can be shown to be $O(1)$ using iterated expectations, change of variables, integration by parts, B1, B2, and B3. Moreover, since $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widehat{g}(Z_i) - g(Z_i)$ converges in distribution (see Proof of Theorem 3), we have that $N_{1n} = O_p(n^{-\frac{1}{2}})$. The second term N_{2n} is of smaller order and can be shown to be $o_p(n^{-\frac{1}{2}})$ using similar arguments. Thus:

$$
L_{112n} = O_p\left(\frac{1}{\sqrt{n}} \right)
$$

It remains to consider L_{113n} of (A-2.5). Using the non-negativity the indicator function together with the decomposition argument of Theorem 2 in Lu and Cheng (2007, p. 1915) for $\left| U_{j\widehat{G}} - U_{jG} \right|$ yields:

$$
\begin{aligned}
L_{113n} \leq & \frac{1}{n^2 h^d} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{f_{x,V}(x,V_i)} I_i \big| U_{j\widehat{G}} - U_{jG} \big| \big| \mathbf{K}_{h,j}(x,V_i) \big| \\
\leq & \sup_{t \leq \phi_F} |\widehat{G}(t) - G(t)| \left[1 + \sup_{t \leq \max_j \{U_j\}} \frac{\big| \{ \widehat{G}(t) - G(t) \} \big|}{\big| 1 - \widehat{G}(t) \big|} \right] \times \\
& \frac{1}{n^2 h^d} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| \frac{U_j}{\{1 - G(U_j)\}^2} \right| \frac{1}{f_{x,V}(x,V_i)} I_i \Big| \mathbf{K}_{h,j}(x,V_i) \Big|
\end{aligned}
$$

By Srinivasan and Zhou (1994, p.199), we have that:

$$
\sup_{t \leq \max_j \{U_j\}} \frac{\big| \{ \widehat{G}(t) - G(t) \} \big|}{\big| 1 - \widehat{G}(t) \big|} = O_p(1)
$$

The term:

$$
\frac{1}{n^2 h^d} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| \frac{U_j}{\{1 - G(U_j)\}^2} \right| \frac{1}{f_{x,V}(x,V_i)} I_i \Big| \mathbf{K}_{h,j}(x,V_i) \Big|
$$

can again be dealt with in the same way as L_{112n} using B1, B3, B7 to show that it is $O_p(1)$. Finally, by conditions B1 and the result of Theorem 3.1 in Chen and Lo (1997):

$$
\sup_{t \leq \phi_F} |\widehat{G}(t) - G(t)| = O_p(n^{-\rho})
$$

for $\frac{2}{5} < \rho < \frac{1}{2}$ (where ρ in turn depends on the "heaviness" of censoring). Putting together these results, the rate of the piece is:

$$L_{113n} = O_p(n^{-\rho})$$

Hence, using B8, the convergence rate of L_{11n} becomes:

$$L_{11n} = O_p\left(\left(\frac{\ln(n)}{nh^{d_z}}\right)^{\frac{1}{2}}\right)$$

The same argument can be used to show that:

$$L_{12n} = O_p\left(\left(\frac{\ln(n)}{nh^{d_z}}\right)^{\frac{1}{2}}\right)$$

and hence the overall rate $O_p((\ln(n)/nh_{d_z})^{\frac{1}{2}})$ of L_{1n} follows.

Next we consider $L_{2n} = \left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{\mu}(x, V_i) - \mu(x)\right|$. We examine the following decomposition:

$$L_{2n} \leq \left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{\mu}(x, V_i) - \mu(x, V_i)\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\mu(x, V_i) - \mu(x)\right|$$

where $\mu(x) = \mathbb{E}\Big[\mu(x, V_i)\Big]$. Since $\widetilde{\mu}(x, V_i)$ is a consistent estimator for $\mu(x, V_i)$ and $\mathbb{E}\Big[(\mu(x, V)^2\Big] < \infty$, we have:

$$\left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{\mu}(x, V_i) - \mu(x, V_i)\right| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

Likewise, since $\mu(x, V_i)$ is continuous (and hence bounded) on $\mathcal{W}$ and B7, we have that:

$$\left|\frac{1}{n}\sum_{i=1}^{n}\mu(x, V_i) - \mu(x)\right| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

■

Proof of Theorem 3

Let $A_{lk}(\theta) = I[X_k \in \mathcal{X}]\big\{I[X_k'\beta(\theta) > X_l'\beta(\theta)] - I[X_k'\beta(\theta_0) > X_l'\beta(\theta_0)]\big\}$. Since the second term involving $\beta(\theta_0)$ does not affect maximization, we note that $\widehat{\theta}$ still maximizes:

$$Q_n(\theta) = \frac{1}{n(n-1)}\sum_{k\neq l}\widehat{\mu}(X_k)A_{lk}(\theta) \tag{A-2.6}$$

and θ_0 its corresponding probability limit:

$$Q(\theta) = \mathbb{E}\Big[\mu(X_k)A_{lk}(\theta)\Big] \tag{A-2.7}$$

Notice in addition that by the normalizations in (A-2.6) and (A-2.7), we have that $Q_n(\theta_0) = Q(\theta_0) = 0$. We expand $Q_n(\theta)$ around the true $\mu(X_k)$ yielding:

$$\begin{aligned}&\frac{1}{n(n-1)}\sum_{k\neq l}\mu(X_k)A_{lk}(\theta)+\frac{1}{n(n-1)}\sum_{k\neq l}\Big(\widehat{\mu}(X_k)-\mu(X_k)\Big)A_{lk}(\theta)\\=&S_{1n}+S_{2n}\end{aligned}$$

In the following, we proceed by examining S_{1n} and S_{2n} in turn, starting with S_{1n}. Identical arguments to the proof of Lemma A.3 in Khan (2001) can be used to show that S_{1n} yields the gradient term plus terms that are of order $o_p(n^{-1})$ once $\sqrt{n}$-consistency of $\|\theta-\theta_0\|$ has been established. That is:

$$S_{1n}=(\theta-\theta_0)^{'}J(\theta-\theta_0)+o_p\Big(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\Big)+o_p\Big(\|\theta-\theta_0\|^2\Big)+o_p\Big(\frac{1}{n}\Big)$$

with

$$J=\frac{1}{2}\,\mathbb{E}\Big[\nabla_{\theta\theta'}\psi_1(X_k,\theta_0)\Big]$$

and $\psi_1(x,\theta)$ defined in (2.11) of section 2.2.3.

S_{2n} on the other hand can be further expanded to give:

$$\begin{aligned}&\frac{1}{n(n-1)}\sum_{k\neq l}\Big(\widetilde{\mu}(X_k)-\mu(X_k)\Big)A_{lk}(\theta)+\frac{1}{n(n-1)}\sum_{k\neq l}\Big(\widehat{\mu}(X_k)-\widetilde{\mu}(X_k)\Big)A_{lk}(\theta)\\=&S_{21n}+S_{22n}\end{aligned}$$

where $\widetilde{\mu}(x)$ is defined analoguously to $\widehat{\mu}(x)$ using the true U_{jG}, I_i, V_j. S_{21n} and S_{22n} determine the components of the variance. They can be tackled through Lemma B3 and Lemmata B4 to B6, respectively: using the result of Lemma B3 below, $S_{21n} = (\theta-\theta_0)^{'}\frac{1}{\sqrt{n}}W_{0n} + o_p\big(\|\theta-\theta_0\|/\sqrt{n}\big)$, where W_{0n} is a sum of zero mean vector random variables that converges in distribution to a random vector defined in Lemma B3. It remains to examine S_{22n}, which can be expanded as in the proof of Theorem 2:

$$\begin{aligned}S_{22n}=&\frac{1}{n(n-1)}\sum_{k\neq l}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\widehat{s}_{X,V}(X_k,\widehat{V}_i)-\widetilde{s}_{X,V}(X_k,V_i)}{\widetilde{f}_{X,V}(X_k,V_i)}+\frac{\widetilde{f}_{X,V}(X_k,V_i)-\widehat{f}_{X,V}(X_k,\widehat{V}_i)}{\widetilde{f}_{X,V}(X_k,V_i)}\widehat{\mu}(X_k,\widehat{V}_i)\right\}A_{lk}(\theta)\\=&S_{22n}^{(1)}+S_{22n}^{(2)}\end{aligned}$$

where $\widehat{s}_{X,V}(\cdot,\cdot)$ and $\widehat{f}_{X,V}(\cdot,\cdot)$ are defined in (A-2.3) and (A-2.4), respectively, and $\widetilde{s}_{X,V}(\cdot,\cdot)$ and

$\widetilde{f}_{X,V}(\cdot,\cdot)$ follow accordingly. We start with $S_{22n}^{(1)}$, which can be further decomposed into:

$$\begin{aligned}
S_{22n}^{(1)} =& \frac{1}{n(n-1)}\sum_{k\neq l}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{1}{nh^d}\sum_{j=1}^{n}\left(\widehat{I}_i - I_i\right)U_{jG}\mathbf{K}_{h,j}(X_k,V_i)}{\frac{1}{nh^d}\sum_{j=1}^{n}I_i\mathbf{K}_{h,j}(X_k,V_i)}\right\}\times A_{lk}(\theta)\\
&+\frac{1}{n(n-1)}\sum_{k\neq l}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{1}{nh^d}\sum_{j=1}^{n}I_iU_{jG}\left(\mathbf{K}_{h,j}(X_k,\widehat{V}_i)-\mathbf{K}_{h,j}(X_k,V_i)\right)}{\frac{1}{nh^d}\sum_{j=1}^{n}I_i\mathbf{K}_{h,j}(X_k,V_i)}\right\}\times A_{lk}(\theta)\\
&+\frac{1}{n(n-1)}\sum_{k\neq l}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{1}{nh^d}\sum_{j=1}^{n}I_i\left(U_{j\widehat{G}}-U_{jG}\right)\mathbf{K}_{h,j}(X_k,V_i)}{\frac{1}{nh^d}\sum_{j=1}^{n}I_i\mathbf{K}_{h,j}(X_k,V_i)}\right\}\times A_{lk}(\theta)\\
&+o_p(1)\\
=&S_{221n}+S_{222n}+S_{223n}+o_p(1)
\end{aligned} \tag{A-2.8}$$

where the $o_p(1)$ term contains cross-products of smaller order. We examine each of the three terms seperately starting with S_{221n}, which by Lemma B4 is equal to:

$$(\theta-\theta_0)'\frac{1}{\sqrt{n}}W_{1n}+o_p\left(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\right)+o_p\left(\frac{1}{n}\right)$$

where W_{1n} is again defined in Lemma B4 below. Likewise, for S_{222n} and S_{223n}, we can apply Lemma B5 and B6 to obtain:

$$S_{222n}=(\theta-\theta_0)'\frac{1}{\sqrt{n}}W_{2n}+o_p\left(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\right)+o_p\left(\frac{1}{n}\right)$$

and

$$S_{223n}=(\theta-\theta_0)'\frac{1}{\sqrt{n}}W_{3n}+o_p\left(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\right)+o_p\left(\frac{1}{n}\right)$$

with W_{2n} and W_{3n} being again sums of zero mean vector random variables that converge to a normal distribution defined in Lemma B5 and B6, respectively. Next we consider $S_{22n}^{(2)}$. A similar decomposition as for $S_{22n}^{(1)}$ and arguments as in Lemmata B4 and B5 can be used to show that the limiting distribution of this term is the same as that of S_{221n} and S_{222n}.

Taking these decompositions of S_{1n} and S_{2n} together and using B5, Lemma B1 and B2 below become directly applicable establishing $\sqrt{n}$-consistency and asymptotic normality. Notice that for (i) of Lemma B1, b_n can be set to be $o(1)$ by the consistency result of Theorem 2. (ii) of the same lemma is satisfied by B2 and B5 in combination with a second order Taylor expansion of $Q(\theta)$ in (A-2.7) around θ_0: $Q(\theta)=\frac{1}{2}(\theta-\theta_0)'\nabla_{\theta\theta'}Q(\overline{\theta})(\theta-\theta_0)\leq-\kappa\|\theta-\theta_0\|^2$ for some constant κ and $\overline{\theta}\in\Theta$. ■

The following two lemmata are from Theorem 3.2 of Khan (2001) (we adapt notation of the original paper to our setup).

Lemma B1. (Lemma A.1 of Khan (2001)) Let $\widehat{\theta}$ maximize $Q_n(\theta)$ in (A-2.6) and θ_0 maximizes $Q(\theta)$ in (A-2.7). Let $b_n, l_n \longrightarrow 0$ as $n \longrightarrow \infty$. If:

(i) $\widehat{\theta} - \theta_0 = O_p(b_n)$

(ii) there exists a neighbourhood $\mathcal{N}$ of 0 and a positive constant κ for which:

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

for all θ in $\mathcal{N}$,

(iii) uniformly over $O_p(b_n)$ neighbourhoods of 0,

$$Q_n(\theta) = Q(\theta) + O_p(\|\theta - \theta_0\|/\sqrt{n}) + o_p(\|\theta - \theta_0\|^2) + O_p(l_n) \tag{A-2.9}$$

then:

$$\|\widehat{\theta} - \theta_0\| = O_p\Big(\max\{l_n^{\frac{1}{2}}, 1/\sqrt{n}\}\Big)$$

Lemma B2. (Lemma A.2 of Khan (2001)) Suppose $\widehat{\theta}$ is $\sqrt{n}$-consistent for θ_0, an interior point of Θ. Suppose also that uniformly over $O_p(1/\sqrt{n})$ neighbourhoods of 0:

$$Q_n(\theta) = (\theta - \theta_0)' J (\theta - \theta_0) + \frac{1}{\sqrt{n}} (\theta - \theta_0)' W_n + o_p(1/n) \tag{A-2.10}$$

where J is a negative definite matrix, and W_n converges in distribution to a $N(0, \Sigma)$ random vector. Then

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, J^{-1} \Sigma J^{-1})$$

.

Lemma B3. Under assumptions A1-A3, B1-B6, and B8, the term $\frac{1}{n(n-1)} \sum_{k \neq l} \Big(\widetilde{\mu}(X_k) - \mu(X_k)\Big) A_{lk}(\theta)$ is equal to:

$$(\theta - \theta_0)' \frac{1}{\sqrt{n}} W_{0n} + o_p\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\right) = (\theta - \theta_0)' \frac{1}{n} \sum_{m=1}^{n} I_m (U_{mG} - \mu(X_m)) \nabla_\theta \psi_2(X_k = X_m, \theta_0) + o_p\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\right)$$

where

$$\frac{1}{\sqrt{n}} \sum_{m=1}^{n} I_m (U_{mG} - \mu(X_m)) \nabla_\theta \psi_2(X_k = X_m, \theta_0) \xrightarrow{d} N(0, \Omega_0)$$

with

$$\Omega_0 = \int \Big(I_m(U_{mG} - \mu(X_m))\nabla_\theta \psi_2(X_m, \theta_0)\Big) \times \Big(I_m(U_{mG} - \mu(X_m))\nabla_\theta \psi_2(X_m, \theta_0)\Big)' dF_{U_G,X,V}(U_{mG}, X_m, V_m)$$

where $\psi_2(\cdot,\cdot)$ is defined in (2.12) of section 2.2.3.

Proof of Lemma B3

Notice that $\frac{1}{n(n-1)} \sum\limits_{k\neq l} \Big(\widetilde{\mu}(X_k) - \mu(X_k)\Big) A_{lk}(\theta)$ can be rewritten as:

$$\frac{1}{n(n-1)} \sum_{k\neq l} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\frac{1}{nh^d} \sum\limits_{j=1}^{n} I_i\big(U_{jG} - \mu(X_k)\big)\mathbf{K}_{h,j}(X_k, V_i)}{\frac{1}{nh^d} \sum\limits_{j=1}^{n} I_i \mathbf{K}_{h,j}(X_k, V_i)} \right\} A_{lk}(\theta)$$

$$= \frac{1}{n(n-1)} \sum_{k\neq l} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\frac{1}{nh^d} \sum\limits_{j=1}^{n} I_i\big(U_{jG} - \mu(X_k)\big)\mathbf{K}_{h,j}(X_k, V_i)}{f_{X,V}(X_k, V_i)} \right\} A_{lk}(\theta) + o_p(1)$$

where $\widehat{f}_{X,V}(X_k, V_i) = \frac{1}{nh^d} \sum\limits_{j=1}^{n} I_i \mathbf{K}_{h,j}(X_k, V_i)$ and the $o_p(1)$ term follows as in the proof of Theorem 2 by B3 and the bandwidth condition B8. As for the first term, $f_{X,V}(X, V)$ is strictly bounded away from zero for every $X, V \in \mathcal{W}$ by B2 and can be restated as:

$$\frac{1}{n(n-1)(n-2)(n-3)} \sum_{k\neq l\neq i\neq j} \frac{1}{h^d} \frac{1}{f_{X,V}(X_k, V_i)} I_i\big(U_{jG} - \mu(X_k)\big)\mathbf{K}_{h,j}(X_k, V_i) A_{lk}(\theta) \qquad \text{(A-2.11)}$$

where omitting terms with $k = l = i = j$ results in an error of magnitude $o_p(\|\theta - \theta_0\|/nh^d)$. The expression in (A-2.11) is a fourth order U-statistic for each $\theta \in \Theta$. Letting $\xi_k = \{I_k, U_{kG}, X_k, V_k\}$ (ξ_l, ξ_i, ξ_j are defined accordingly), (A-2.11) is:

$$\frac{1}{n(n-1)(n-2)(n-3)} \sum_{k\neq l\neq i\neq j} q_n(\xi_i, \xi_j, \xi_k, \xi_l; \theta) \qquad \text{(A-2.12)}$$

where $q_n(\cdot,\cdot,\cdot,\cdot;\theta) = \frac{1}{h^d} I_i\big(U_{jG} - \mu(X_k)\big)\mathbf{K}_{h,j}(X_k, V_i) A_{lk}(\theta)$ is the 'kernel' function of the U-statistic. Using iterated expectations repeatedly, change of variables, together with B1, B3, B5, and B6 one can show that $q_n(\cdot,\cdot,\cdot,\cdot)$ is degenerate in ξ_k, ξ_l, and ξ_i for each $\theta \in \Theta$ since the expectation of $(U_{jG} - \mu(X_k))$ conditional on X_k is zero. By iterated expectations, this in turn implies that $\mathbb{E}\Big[q_n(\xi_k, \xi_l, \xi_i, \xi_j, \theta)\Big] = 0$. By contrast, after change of variables, iterated expectations, and dominated convergence, the term $\mathbb{E}\Big[q_n(\xi_k,\xi_l,\xi_i,\xi_j,\theta)\Big|\xi_j\Big]$ yields:

$$\begin{aligned}\lim_{n\longrightarrow\infty}\mathbb{E}\Big[q_n(\xi_k,\xi_l,\xi_i,\xi_j,\theta)\Big|\xi_j\Big] =&\mathbb{E}\Big[q(\xi_k,\xi_l,\xi_i,\xi_j,\theta)\Big|\xi_j\Big]\\ =&\Big(U_{jG}-\mu(X_j)\Big)I_j\,\mathbb{E}\Big[A_{lk}(\theta)\Big|X_k=X_j\Big]\\ =&(\theta-\theta_0)'\Big(U_{jG}-\mu(X_j)\Big)I_j\nabla_\theta\psi_2(X_j,\theta_0)+O(\|\theta-\theta_0\|^2)\end{aligned}$$

where the last equality follows by a second order Taylor expansion of $\mathbb{E}\Big[A_{lk}(\theta)\Big|X_k=X_j\Big]$ around θ_0. Applying the Hoeffding decomposition to the degenerate fourth order U-process in (A-2.12) (Serfling, 1980) and noting that, by Lemma A.6 in Khan (2001) and the arguments used therein, all terms except the leading term are of order $o_p\Big(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\Big)$, yields:

$$\frac{1}{n(n-1)(n-2)(n-3)}\sum_{k\neq l\neq i\neq j}q_n(\xi_i,\xi_j,\xi_k,\xi_l;\theta)=\frac{1}{n}\sum_{m=1}^{n}\mathbb{E}\Big[q(\xi_k,\xi_l,\xi_i,\xi_j,\theta)\Big|\xi_j=\xi_m\Big]+o_p\Bigg(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\Bigg)$$

and hence:

$$(\theta-\theta_0)'\frac{1}{n}\sum_{m=1}^{n}\big(U_{mG}-\mu(X_m)\big)I_m\nabla_\theta\psi_2(X_k=X_m,\theta_0)+o_p\Bigg(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\Bigg)$$

The expression:

$$\frac{1}{n}\sum_{m=1}^{n}\big(U_{mG}-\mu(X_m)\big)I_m\nabla_\theta\psi_2(X_k=X_m,\theta_0)\tag{A-2.13}$$

is a sum of zero mean random variables. Applying Lindberg Levy's Central Limit Theorem (CLT) yields the result of the lemma. ■

Lemma B4. Under assumptions A1-A3, B1-B8, the term S_{221n} defined in the proof of Theorem 3 is equal to:

$$\begin{aligned}&(\theta-\theta_0)'\frac{1}{\sqrt{n}}W_{1n}+o_p\Bigg(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\Bigg)+o_p\Bigg(\frac{1}{n}\Bigg)\\ =&(\theta-\theta_0)'\frac{1}{n}\sum_{m=1}^{n}\Big(\widehat{g}(Z_m)-g(Z_m)\Big)\Big(F_V^{(1)}\Big(a\Big)+F_V^{(1)}\Big(b\Big)\Big)\int U_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X,V}(U_G,X_k,V_i)\Big)\\ &+o_p\Bigg(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\Bigg)+o_p\Bigg(\frac{1}{n}\Bigg)\end{aligned}$$

where

$$\frac{1}{\sqrt{n}}\sum_{m=1}^{n}\Big(\widehat{g}(Z_m)-g(Z_m)\Big)\Big(F_V^{(1)}\Big(\overline{V}_a\Big)+F_V^{(1)}\Big(\overline{V}_b\Big)\Big)\int U_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X}(U_G,X_k)\Big)\xrightarrow{d}N\Big(0,\Omega_1\Big)$$

with $\Omega_1=E_1\Phi_1E_1'$:

$$\Phi_1=\int V_i^2dF_V(V_i)$$

and

$$E_1 = \Big(F_V^{(1)}\Big(a\Big) + F_V^{(1)}\Big(b\Big)\Big) \int U_{jG} \nabla_\theta \psi_2(X_k, \theta_0) dF_{U_G,X}(U_{jG}, X_k)$$

where $F_{U_G,X}(\cdot,\cdot)$ denotes the joint distribution function of U_{jG} and X_k.

Proof of Lemma B4

As before, we start by replacing $\widehat{f}_{X,V}(X_k, V_i)$ with the true density $f_{X,V}(X_k, V_i)$ using B2, B3, and B8. Moreover, notice that $I\{a \le \widehat{V}_j \le b\} - I\{a \le V_i \le b\} = I\{\widehat{V}_i \le b\} + I\{\widehat{V}_i \ge a\} - I\{V_i \le b\} - I\{V_i \ge a\} = \big(I\{\widehat{V}_i \le b\} - I\{V_i \le b\}\big) + \big(I\{\widehat{V}_i \ge a\} - I\{V_i \ge a\}\big)$. We focus on $\big(I\{\widehat{V}_i \le b\} - I\{V_i \le b\}\big)$, the other term will follow by an identical argument. Let $F_V(b)$ denote the distribution function of V_i evaluated at b and $\mathbf{B}_{ijkl}(\theta) = f^{-1}_{X,V}(X_k, V_i) U_{jG} \mathbf{K}_{h,j}(X_k, V_i) A_{lk}(\theta)$. Then we can decompose S_{221n} as follows:

$$\begin{aligned}
&\frac{1}{n(n-1)(n-2)(n-3)} \sum_{k \neq l \neq i \neq j} \big(I\{V_i \le b + (\widehat{V}_i - V_i)\} - I\{V_i \le b\}\big) \frac{1}{h^d} \mathbf{B}_{ijkl}(\theta) \\
=&\frac{1}{n(n-1)(n-2)(n-3)} \sum_{k \neq l \neq i \neq j} \Big\{ I\{V_i \le b + (\widehat{V}_i - V_i)\} - F_V\Big(b + (\widehat{V}_i - V_i)\Big) \\
&- I\{V_i \le b\} + F_V\Big(b\Big)\Big\} \times \frac{1}{h^d} \mathbf{B}_{ijkl}(\theta) \\
&+ \frac{1}{n(n-1)(n-2)(n-3)} \sum_{k \neq l \neq i \neq j} \Big(F_V\Big(b + (\widehat{V}_i - V_i)\Big) + F_V\Big(b\Big)\Big) \frac{1}{h^d} \mathbf{B}_{ijkl}(\theta) \\
=&T_{1n}(\theta) + T_{2n}(\theta)
\end{aligned}$$

We start with $T_{1n}(\theta)$. We examine the term involving $F_V\Big(b\Big) - I\{V_i \le b\}$, the term with $I\{V_i \le b + (\widehat{V}_i - V_i)\} - F_V\Big(b + (\widehat{V}_i - V_i)\Big)$ follows by the same argument. Adding and subtracting $\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijkl}(\theta)\Big]$ yields:

$$\begin{aligned}
T_{1n}(\theta) =& \frac{1}{n(n-1)(n-2)(n-3)} \sum_{k \neq l \neq i \neq j} \Big(F_V\Big(b\Big) - I\{V_i \le b\}\Big) \mathbb{E}\Big[\frac{1}{h^d} \mathbf{B}_{ijlk}(\theta)\Big] \\
&+ \frac{1}{n(n-1)(n-2)(n-3)} \sum_{k \neq l \neq i \neq j} \Big(F_V\Big(b\Big) - I\{V_i \le b\}\Big) \times \\
&\Big(\frac{1}{h^d} \mathbf{B}_{ijlk}(\theta) - \mathbb{E}\Big[\frac{1}{h^d} \mathbf{B}_{ijlk}(\theta)\Big]\Big) \\
=&T_{11n}(\theta) + T_{12n}(\theta)
\end{aligned}$$

We start with the first piece, which can be simplified since no term depends on k,l, or j:

$$\frac{1}{n} \sum_{i=1}^{n} \Big(F_V\Big(b\Big) - I\{V_i \le b\}\Big) \mathbb{E}\Big[\frac{1}{h^d} \mathbf{B}_{ijlk}(\theta)\Big]$$

Since $\mathbb{E}\Big[F_V\Big(b\Big) - I\{V_i \leq b\}\Big] = 0$, notice that by change of variables, iterated expectations, and a second order Taylor expansion of $\mathbb{E}\Big[A_{lk}(\theta)\Big|X_k = X_j\Big]$ around θ_0, the variance of T_{11n} is $\mathbb{E}\Big[\Big(T_{11n}(\theta)\Big)^2\Big] = O\Big(\frac{\|\theta-\theta_0\|^2}{n}\Big)$. Thus, using Chebychev's inequality, we have that $T_{11n}(\theta) = o_p(n^{-1})$ uniformly over $o_p(1)$ neighbourhoods of θ_0.

Next, we consider $T_{12n}(\theta)$. To derive an upper bound for the convergence rate of $T_{12n}(\theta)$ via Rosenthal's inequality, we first examine:

$$\begin{aligned}&\mathbb{E}\Bigg[\Big(I\{V_i \leq b\} - F_V\Big(b\Big)\Big)^2\Bigg\{\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{X,V}(X_k,V_i)}U_{jG}\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)\Big]^2\\ &+2\frac{1}{h^d}\frac{1}{f_{X,V}(X_k,V_i)}U_{jG}\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)\times\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{X,V}(X_k,V_i)}U_{jG}\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)\Big]\\ &+\frac{1}{h^{2d}}\frac{1}{f_{X,V}(X_k,V_i)^2}U_{jG}^2\mathbf{K}_{h,j}^2(X_k,V_i)A_{lk}^2(\theta)\Bigg\}\Bigg]\\ =&T_{121}(\theta)+T_{122}(\theta)+T_{123}(\theta)\end{aligned}$$

We start with $T_{121}(\theta)$. Using change of variables, iterated expectations, a second order Taylor expansion of $\mathbb{E}\Big[A_{lk}(\theta)\Big|X_k = X_j\Big]$ around θ_0, B3, B7, and the boundedness of the indicator function, we have that $T_{121}(\theta) = O(\|\theta-\theta_0\|^2) + o(\|\theta-\theta_0\|^2)$. By the same line of argument, the same rates can be obtained for $T_{122}(\theta)$. Using again $u_1 = (X_k - X_j)/h$, $u_2 = (V_i - V_j)/h$, boundedness of the indicator function, B2, B3, B6, B7, and the equality $A_{lk}^2(\theta) = |A_{lk}(\theta)|$, $T_{123}(\theta)$ on the other hand is given by:

$$\begin{aligned}T_{123}(\theta)=&\int\Big(I\{V_i\leq b\}-F_V\Big(b\Big)\Big)^2\frac{1}{h^d}\frac{1}{f_{X,V}(X_j+hu,V_j+hu_2)^2}U_{jG}^2\mathbf{K}_j^2(X_j+hu_1,V_j+hu_2)\\ &\times\Big|(\theta-\theta_0)'\nabla_\theta\psi_2(X_j+hu_1,\theta_0)\Big|f_{X,V}(X_j,V_j)f_{X,V}(X_j+hu_1,V_j+hu_2)\\ &\times dx_j du_1 dv_j du_2 dU_G + o\Big(\frac{\|\theta-\theta_0\|}{h^d}\Big)\\ =&O\Big(\frac{\|\theta-\theta_0\|}{h^d}\Big)(1+h)+o\Big(\frac{\|\theta-\theta_0\|}{h^d}\Big)\end{aligned}$$

Moreover, using identical arguments:

$$\mathbb{E}\Bigg[\Bigg|\Big(F_V\Big(b\Big)-I\{V_i\leq b\}\Big)\Big(\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)-\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big]\Big)\Bigg|^\kappa\Bigg]=O(\|\theta-\theta_0\|h^{-d(\kappa-1)})+o(\|\theta-\theta_0\|h^{-d(\kappa-1)})$$

for $\kappa \geq 1$. Applying Rosenthal's inequality yields:

$$\begin{aligned}\mathbb{E}\Big[\Big(T_{12n}(\theta)\Big)^{2\kappa}\Big]\leq&n^{-8\kappa}\Xi_\kappa\Big(\|\theta-\theta_0\|n^{4\kappa}h^{-d\kappa}+\|\theta-\theta_0\|n^4h^{-2\kappa d+d}\Big)\\ =&O\Big(\|\theta-\theta_0\|n^{-4\kappa}h^{-d\kappa}\Big)+O\Big(\|\theta-\theta_0\|n^{-8\kappa+4}h^{-2\kappa d+d)}\Big)\end{aligned}$$

with Ξ_κ some positive constant. Using B6, we obtain the following rates for $\kappa = 1$: $O(\|\theta-\theta_0\|n^{-4}h^{-d}) + O(\|\theta-\theta_0\|n^{-4}h^{-2d+d}) = O(\|\theta-\theta_0\|n^{-4}h^{-d})$. By the bandwidth conditions in

B8, Markov's inequality thus implies $T_{12n}(\theta) = o_p(n^{-1})$ uniformly over $o_p(1)$ neighbourhoods of θ_0.

Next, we consider $T_{2n}(\theta)$, which can again be decomposed as:

$$
\begin{aligned}
T_{2n}(\theta) =& \frac{1}{n(n-1)(n-2)(n-3)} \sum_{k\neq l\neq i\neq j} \Big(F_V\Big(b + (\widehat{V}_i - V_i)\Big) + F_V\Big(b\Big)\Big) \mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big] \\
&+ \frac{1}{n(n-1)(n-2)(n-3)} \sum_{k\neq l\neq i\neq j} \Big(F_V\Big(b + (\widehat{V}_i - V_i)\Big) + F_V\Big(b\Big)\Big) \\
&\times \Big(\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta) - \mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big]\Big) \\
=& T_{21n}(\theta) + T_{22n}(\theta)
\end{aligned}
$$

We start with T_{21n}. Using B7, a mean value expansion around $(\widehat{V}_i - V_i) = 0$, and a simplification (since T_{21n} only depends on j) yield:

$$
\frac{1}{n}\sum_{i=1}^{n} F_V^{(1)}\Big(\overline{V}_b\Big)\Big(\widehat{V}_i - V_i\Big)\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big] = \frac{1}{n}\sum_{i=1}^{n} F_V^{(1)}\Big(\overline{V}_b\Big)\Big(\widehat{g}(Z_i) - g(Z_i)\Big)\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big] \tag{A-2.14}
$$

where $\overline{V}_b \in [b, b + (\widehat{V}_i - V_i)]$ and $F_V^{(1)}$ denotes the first derivative w.r.t. its argument. Using iterated expectations, change of variables, and a second order Taylor expansion of $\mathbb{E}\Big[A_{lk}(\theta)|X_k\Big]$ around θ_0, the expectation expression in (A-2.14) yields:

$$
\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big] = (\theta - \theta_0)' \int U_{jG}\nabla_\theta\psi_2(X_k, \theta_0)dF_{U_G,X}(U_G, X_k)(1 + O(h)) + o(\|\theta - \theta_0\|^2)
$$

For the random component in (A-2.14), recall that $\widehat{g}(Z_i) = (\sum_{j=1}^{n} X_{1j}\mathbf{k}_h(Z_i - Z_j))/\sum_{j=1}^{n}\mathbf{k}_h(Z_i - Z_j))$. We examine the following standard decomposition:

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{g}(Z_i) - g(Z_i)\widehat{f}_Z(Z_i)}{\widehat{f}_Z(Z_i)} \\
=& \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\Big(\widehat{g}(Z_i) - g(Z_i)\widehat{f}_Z(Z_i)\Big)}{f_Z(Z_i)} + \left(\frac{f_Z(Z_i) - \widehat{f}_Z(Z_i)}{f_Z(Z_i)\widehat{f}_Z(Z_i)}\right) \times \Big(\widehat{g}(Z_i) - g(Z_i)\widehat{f}_Z(Z_i)\Big)\right\}
\end{aligned}
$$

Since $\sup_{Z\in\mathcal{W}}|f_Z(Z_i) - \widehat{f}_Z(Z_i)| = o_p((\ln(n)/nh^d)^{\frac{1}{2}}) = o_p(1)$ by B3 and B8, the second term is of smaller order than the first one and will hence be neglected. Moreover, since $X_{1i} = g(Z_i) + V_i$, observe that the first term is can be restated as:

$$
\frac{1}{n^2h}\sum_{j=1}^{n}\sum_{i=1}^{n}\Big(\frac{g(Z_j) - g(Z_i)}{f_Z(Z_i)}\Big)\mathbf{k}_h(Z_i - Z_j) + \frac{1}{n^2h}\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{V_j}{f_Z(Z_i)}\mathbf{k}_h(Z_i - Z_j)
$$

Now, notice that omitting observations with $i = j$ results in a negligible error of order $o_p((nh)^{-1})$, while $\frac{1}{2}(V_j f_Z^{-1}(Z_i) - V_i f_Z^{-1}(Z_j))\mathbf{k}_h(Z_i - Z_j)$ is the 'symmetrized' version of the second term and:

$$\frac{1}{2h}\left(\frac{(g(Z_i)-g(Z_j))}{f_Z(Z_j)}-\frac{(g(Z_j)-g(Z_i))}{f_Z(Z_i)}\right)\mathbf{k}_h(Z_i-Z_j)=\frac{1}{2h}\Big(\varpi_{ij}-\varpi_{ji}\Big)\mathbf{k}_h(Z_i-Z_j)$$

is the 'symmetrized' version of the first term with $\varpi_{ij} = (g(Z_i) - g(Z_j))/f_Z(Z_j)$ and ϖ_{ji} defined accordingly. Hence, the above expressions can be rewritten as symmetric second order U-statistics:

$$\begin{aligned}&\binom{n}{2}^{-1}\sum_{i\neq j}\frac{1}{2h}(\varpi_{ij}-\varpi_{ji})\mathbf{k}_h(Z_i-Z_j)+\binom{n}{2}^{-1}\sum_{i\neq j}\frac{1}{2h}\left(\frac{V_i}{f_Z(Z_j)}-\frac{V_j}{f_Z(Z_i)}\right)\mathbf{k}_h(Z_i-Z_j)\\&=T_{211n}+T_{212n}\end{aligned}$$

By symmetry of the kernel function and the i.i.d. assumption on one hand, and by independence between V_i and Z_i on the other, one can straightforwardly verify that $\mathbb{E}\Big[T_{211n}\Big] = \mathbb{E}\Big[T_{212n}\Big] = 0$. Moreover, letting $r_n(Z_i, Z_j) = \frac{1}{2h}(\varpi_{ij} + \varpi_{ji})\mathbf{k}_h(Z_i - Z_j)$, one can use B2, B3, B8, and a change of variables to verify that:

$$\mathbb{E}\Big[\Big|r_n(\xi_i,\xi_j)\Big|^2\Big]=o(n)$$

Thus, since by change of variables $\mathbb{E}\Big[r_n(Z_i, Z_j)\Big|Z_i\Big] = \mathbb{E}\Big[r_n(Z_i, Z_j)\Big|Z_j\Big] = O(h)$, one can use Lemma 3.1 in Powell, Stock, and Stoker (1989) to infer that $T_{211n} = o_p\Big(\frac{1}{\sqrt{n}}\Big)$. Next, we examine the leading term T_{212n}. Let $p_n(\xi_i, \xi_j) = \frac{1}{2h}((V_i/f_Z(Z_j)) - (V_j/f_Z(Z_i)))\mathbf{k}_h(Z_i - Z_j)$ with $\xi_i = \{Z_i, V_i\}$ and $\xi_j = \{Z_j, V_j\}$. By B2, B3, B8, and change of variables one can verify that:

$$\mathbb{E}\Big[\Big|p_n(\xi_i,\xi_j)\Big|^2\Big]=o(n)$$

Using again Lemma 3.1 in Powell, Stock, and Stoker (1989), we have that:

$$\sqrt{n}T_{212n}=\sqrt{n}\frac{2}{n}\sum_{i=1}^{n}\mathbb{E}\Big[p_n(\xi_i,\xi_j)\Big|\xi_i\Big]+o_p(1)$$

After change of variables with $u_3 = (Z_i - Z_j)/h$, independence between Z_i and V_i, and $\mathbb{E}\Big[V_j\Big|V_i\Big] = \mathbb{E}\Big[V_j\Big] = 0$:

$$\mathbb{E}\Big[p_n(\xi_i,\xi_j)\Big|\xi_i\Big]=\int\frac{1}{2}\left(\frac{V_i}{f_Z(Z_i+hu_3)}-\frac{V_j}{f_Z(Z_i)}\right)f_Z(Z_i+hu_3)f_V(V_j)dv_j=\frac{1}{2}V_i=\mathbb{E}\Big[p(\xi_i,\xi_j)\Big|\xi_i\Big]$$

where $\mathbb{E}\Big[p(\xi_i, \xi_j)\Big|\xi_i\Big]$ denotes the limit expression. Thus, we have:

$$\sqrt{n}T_{212n}=\sqrt{n}\frac{2}{n}\sum_{i=1}^{n}\mathbb{E}\Big[p(\xi_i,\xi_j)\Big|\xi_i\Big]+o_p(1)\tag{A-2.15}$$

Applying Lindberg Levy's Central Limit Theorem (CLT) to equation (A-2.15), we have:

$$\sqrt{n}T_{212n} \xrightarrow{d} N\Big(0, \int V_i^2 dF_V(V_i)\Big)$$

where $F_V(\cdot)$ is the distribution function of V_i. Thus, for (A-2.14) we obtain (adding the neglected term $F_V^{(1)}\Big(\overline{V}_a\Big)$):

$$\sqrt{n}\frac{1}{n}\sum_{m=1}^{n}\Big(\widehat{g}(Z_m)-g(Z_m)\Big)\Big(F_V^{(1)}\Big(\overline{V}_a\Big)+F_V^{(1)}\Big(\overline{V}_b\Big)\Big)\int U_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X}(U_G,X_k)\Big) \xrightarrow{d} N\Big(0,\Omega_1\Big)$$

where Ω_1 was defined in the statement of the lemma.

It remains to show that T_{22n} is of smaller order than the previous term. Using B2, B3, B7, a mean values expansion and a similar decomposition as for T_{12n}, one can show that $\mathbb{E}\Big[\Big|T_{22n}\Big|^\kappa\Big] = O(\|\theta-\theta_0\|h^{-(\kappa d-d)}) + o(\|\theta-\theta_0\|h^{-(\kappa d-d)})$. Thus, application of Rosenthal's inequality (with $\kappa = 1$), followed by Markov's inequality, and the bandwidth conditions imply that $T_{22n} = o_p(n^{-1})$ uniformly over $o_p(1)$ neighbourhoods of θ_0. ■

Lemma B5. Under assumptions A1-A3, B1-B8, the term S_{222n} defined in the proof of Theorem 3 is equal to:

$$\begin{aligned}&(\theta-\theta_0)'\frac{1}{\sqrt{n}}W_{2n} + o_p\Big(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\Big) + o_p\Big(\frac{1}{n}\Big)\\ &= (\theta-\theta_0)'\frac{1}{n}\sum_{m=1}^{n}\Big(\widehat{g}(Z_m)-g(Z_m)\Big)\Big(-\int I_iU_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X,V}(U_G,X_k,V_i)\Big)+o_p\Big(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\Big)+o_p\Big(\frac{1}{n}\Big)\end{aligned}$$

where

$$\frac{1}{\sqrt{n}}\sum_{m=1}^{n}\Big(\widehat{g}(Z_m)-g(Z_m)\Big)\Big(-\int I_iU_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X,V}(U_G,X_k,V_i)\Big) \xrightarrow{d} N\Big(0,\Omega_2\Big)$$

with $\Omega_2 = E_2\Phi_2E_2'$ where

$$\Phi_2 = \Phi_1$$

and

$$E_2 = -\int I_iU_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X,V}(U_{jG},X_k,V_i)$$

.

Proof of Lemma B5

Next, we consider S_{222n}. First we replace again $\widehat{f}_{X,V}(X_k, V_i)$ by $f_{X,V}(X_k, V_i)$ using B2, B3, and B8. After a mean value expansion around $(V_i - V_j)$, we have:

$$S_{222n} = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{k \neq l \neq i \neq j} \frac{1}{h^{d+1}} f_{X,V}^{-1}(X_k, V_i) I_i U_{jG} \mathbf{K}_{h,j}^{(1)}(X_k, \overline{V}) \Big(\Big(\widehat{V}_i - V_i \Big) - \Big(\widehat{V}_j - V_j \Big) \Big) A_{lk}(\theta)$$

where $\mathbf{K}_{h,j}^{(1)}(x, \overline{V}_i)$ is defined in the proof of Theorem 2. As before $\Big(\widehat{V}_i - V_i\Big) = \Big(\widehat{g}(Z_i) - g(Z_i)\Big)$, while the term involving subscript j follows by an identical argument. Let $\mathbf{C}_{ijkl}(\theta) = f_{X,V}^{-1}(X_k, V_i) I_i U_{jG} \mathbf{K}_{h,j}^{(1)}(X_k, \overline{V}) \mathbf{A}_{lk}(\theta)$. Then, we have:

$$\begin{aligned}
S_{212n} =& \frac{1}{n(n-1)(n-2)(n-3)} \sum_{k \neq l \neq i \neq j} \Big(\widehat{g}(Z_i) - g(Z_i) \Big) \Big(\frac{1}{h^{d+1}} \mathbf{C}_{ijkl}(\theta) - \mathbb{E}\Big[\frac{1}{h^{d+1}} \mathbf{C}_{ijkl}(\theta) \Big] \Big) \\
&+ \frac{1}{n(n-1)(n-2)(n-3)} \sum_{k \neq l \neq i \neq j} \Big(\widehat{g}(Z_i) - g(Z_i) \Big) \mathbb{E}\Big[\frac{1}{h^{d+1}} \mathbf{C}_{ijkl}(\theta) \Big] \\
=& R_{1n}(\theta) + R_{2n}(\theta)
\end{aligned}$$

We start with $R_{1n}(\theta)$. Integration by parts and a similar line of argument to before can be used to show that $\mathbb{E}\Big[\Big|R_{1n}(\theta)\Big|\Big] = O\Big(\|\theta - \theta_0\| h^{-(d\kappa - d) - (\kappa - 1)}\Big) + o\Big(\|\theta - \theta_0\| h^{-(d\kappa - d) - (\kappa - 1)}\Big)$, while the leading term of $\mathbb{E}\Big[R_{1n}(\theta)^2\Big]$ is $O\Big(\|\theta - \theta_0\| h^{-(d+1)}\Big)$. Applying again Rosenthal's inequality yields:

$$\mathbb{E}\Big[\Big(R_{1n}(\theta) \Big)^{2\kappa} \Big] \leq n^{-8\kappa} \Xi_\kappa \Big(\|\theta - \theta_0\| \Big(n^{4\kappa} h^{-d\kappa - \kappa} + n^4 h^{-2d\kappa - 2\kappa + (d+1)} \Big) \Big)$$

For $\kappa = 1$, we have $O(\|\theta - \theta_0\| n^{-4} h^{-d-1})$. By Markov's inequality and the bandwidth conditions, we have that $R_{1n}(\theta) = o_p(n^{-1})$ uniformly over $o_p(1)$ neighbourhoods of θ_0. Next, consider $R_{2n}(\theta)$. This term only depends on i and can be shown to converge in distribution as claimed in the above lemma using the same arguments as for $T_{21n}(\theta)$ in Lemma B2. That is:

$$\sqrt{n} \frac{1}{n} \sum_{m=1}^{n} \Big(\widehat{g}(Z_m) - g(Z_m) \Big) \Big(- \int I_i U_{jG} \nabla_\theta \psi_2(X_k, \theta_0) dF_{U_G, X, V}(U_G, X_k, V_i) \Big) \xrightarrow{d} N\Big(0, \Omega_2\Big)$$

where Ω_2 was defined in the lemma. ■

Lemma B6. Under assumptions A1-A3, B1-B8, the term S_{223n} defined in the proof of Theorem 3 is equal to:

$$\begin{aligned}
&(\theta - \theta_0)' \frac{1}{\sqrt{n}} W_{3n} + o_p\Big(\frac{\|\theta - \theta_0\|}{\sqrt{n}} \Big) + o_p\Big(\frac{1}{n} \Big) \\
&= (\theta - \theta_0)' \frac{1}{n} \sum_{m=1}^{n} \Big(U_{m\widehat{G}} - U_{mG} \Big) \Big(\int I_i \nabla_\theta \psi_2(X_k, \theta_0) dF_{X,V}(X_k, V_i) \Big) + o_p\Big(\frac{\|\theta - \theta_0\|}{\sqrt{n}} \Big) + o_p\Big(\frac{1}{n} \Big)
\end{aligned}$$

where

$$\frac{1}{\sqrt{n}}\sum_{m=1}^{n}\Big(U_{m\widehat{G}}-U_{mG}\Big)\Big(\int I_i\nabla_\theta\psi_2(X_k,\theta_0)dF_{X,V}(X_k,V_i)\Big)\xrightarrow{d}N\Big(0,\Omega_3\Big)$$

with $\Omega_3 = E_3\Phi_3E_3^{'}$ where

$$\Phi_3=\int_0^{\phi_Y}\mathbb{E}\Big[U_{1G}I[s<U_1]\Big]H_{t1}(s)\frac{dG(s)}{(1-G(s-))}$$

and

$$E_3=\int I_i\nabla_\theta\psi_2(X_k,\theta_0)dF_{X,V}(X_k,V_i)$$

where $F_{X,V}(\cdot,\cdot)$ denotes the joint distribution function of X_k and V_i.

Proof of Lemma B6

$\widehat{f}_{X,V}(X_k,V_i)$ in the denominator is again tackled using B2, B3, and B8.

Let $\mathbf{D}_{ijkl}(\theta)=f^{-1}_{X,V}(X_k,V_i)I_i\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)$. Then, S_{213n} can be decomposed as:

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n}\Big(U_{j\widehat{G}}-U_{jG}\Big)\mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\\&+\frac{1}{n(n-1)(n-2)(n-3)}\sum_{k\neq l\neq i\neq j}\Big(U_{j\widehat{G}}-U_{jG}\Big)\Big(\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)-\mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\Big)\\=&U_{11n}(\theta)+U_{12n}(\theta)\end{aligned}$$

Consider $U_{11n}(\theta)$. We define the following notation, which we keep as close as possible to Lu and Burke (2005):

$$\begin{aligned}\Lambda^G(t)=&\int_{-\infty}^{t}\frac{1}{1-G(s-)}dG(s)\\N_j(t)=&I[U_j\le t,\Delta_j=0]\\M_j(t)=&N_j(t)-\int_0^t I[U_j\ge s]d\Lambda_j(s),\qquad\Lambda_j(s)=\Lambda^G(s)\\Y_n(t)=&\sum_{j=1}^{n}I[U_j\ge t],\qquad\overline{Y}_n(t)=\frac{1}{n}Y_n(t)\end{aligned}$$

Moreover, $F_Y(\cdot-)$ will in the following refer to the left-continuous distribution function of Y_j. Noting that U_j only has support on the positive real line and that:

$$\frac{\widehat{G}(U_j-)-G(U_j-)}{1-G(U_j-)}=\int_{s<U_j}\frac{1-\widehat{G}(s-)}{1-G(s-)}\frac{\sum_{j=1}^n dM_j(s)}{Y_n(s)}=\frac{1}{n}\int_{s<U_j}\frac{1-\widehat{G}(s-)}{1-G(s-)}\frac{\sum_{j=1}^n dM_j(s)}{\overline{Y}_n(s)}$$

$U_{11n}(\theta)$ can be rewritten as:

$$
\begin{aligned}
U_{11n}(\theta) =& \frac{1}{n}\sum_{j=1}^{n}\Big(U_{j\widehat{G}} - U_{jG}\Big)\,\mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big] \\
=& \frac{1}{n}\sum_{j=1}^{n} U_{j\widehat{G}}\frac{\widehat{G}(U_j-) - G(U_j-)}{1-G(U_j-)}\,\mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big] \\
=& \mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\Bigg\{\frac{1}{n^2}\sum_{m=1}^{n}\int_0^{\infty}\sum_{j=1}^{n}\frac{U_j\Delta_j}{1-G(U_j-)}I[s<U_j]\frac{1}{\overline{Y}_n(s)}\frac{1-\widehat{G}(s-)}{1-G(s-)}dM_m(s) \\
&+\frac{1}{n^2}\sum_{m=1}^{n}\int_0^{\infty}\sum_{j=1}^{n}U_j\Delta_j\left(\frac{1}{1-\widehat{G}(U_j-)}-\frac{1}{1-G(U_j-)}\right)I[s<U_j]\frac{1}{\overline{Y}_n(s)}\frac{1-\widehat{G}(s-)}{1-G(s-)}dM_m(s)\Bigg\} \\
=& U_{111n}(\theta) + U_{112n}(\theta)
\end{aligned}
$$

Using Lemma A.2 (ii) in Lopez (2009) and B1, $U_{112n}(\theta)$ is of smaller order than $U_{111n}(\theta)$ and can hence be neglected in the following. Letting $H_{nt}(s) = \frac{1}{n}\sum_{j=1}^{n} U_{jG}I[s<U_j]\frac{1}{\overline{Y}_n(s)}\frac{1-\widehat{G}(s-)}{1-G(s-)}$, we have for the first term:

$$
U_{111n}(\theta) = \mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\frac{1}{n}\sum_{m=1}^{n}\int_0^{\infty} H_{nt}(s)dM_m(s)
$$

Now for $0<\nu<\phi_Y$, let:

$$
U_{111n}^{\nu}(\theta) = \mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\frac{1}{n}\sum_{m=1}^{n}\int_0^{\nu} H_{nt}(s)dM_m(s)
$$

Then, uniformly for $s\in[0,\nu]$, we have:

$$
\begin{aligned}
H_{nt}(s) =& \frac{1}{n}\sum_{m=1}^{n} U_{jG}I[s<U_j]\frac{1}{\overline{Y}_n(s)}\frac{1-\widehat{G}(s-)}{1-G(s-)} \\
=& \frac{1}{n}\sum_{m=1}^{n} U_{jG}I[s<U_j]\frac{1}{(1-F_Y(s-))(1-G(s-))} + o_p(1) \\
=& \mathbb{E}\Big[U_{1G}I[s<U_1]\Big]\frac{1}{(1-F_Y(s-))(1-G(s-))} + o_p(1) \\
=& H_{1t}(s) + o_p(1)
\end{aligned}
$$

where the second and third equality follow by adding and subtracting $\frac{1}{(1-F_Y(s-))(1-G(s-))}$ and $\mathbb{E}\Big[U_{1G}I[s<U_1]\Big]$, respectively (see Lemma A.8 in Lu and Burke (2005) for details). Moreover, using the same lines of arguments as in the proof of statement (2.29) in Lai, Ying, and Zheng (1995, p.274) and B1, we have:

$$
\sqrt{n}\frac{1}{n}\sum_{m=1}^{n}\int_{\nu}^{\infty} H_{nt}(s)dM_m(s) \xrightarrow{p} 0
$$

as $\nu \longrightarrow \phi_Y$ and $n \longrightarrow \infty$. Therefore:

$$\sqrt{n}U_{11n} = M_{n2t} + o_p(1)$$

For $0 < \nu < \phi_Y$, $\{M_{n2t}\}$ is a local martingale with predictable variation process (Lu and Burke, 2005, p.198).

$$\begin{aligned}
\langle M_{2nt}(\nu)\rangle =& \frac{1}{n}\sum_{m=1}^{n}\int_0^{\nu} H_{t1}^2(s)I[U_m \geq s](1-\Delta\Lambda^G(s))d\Lambda^G(s) \\
\xrightarrow{p}& \int_0^{\nu} H_{t1}^2(s)\,\mathbb{P}[U_1 \geq s](1-\Delta\Lambda^G(s))d\Lambda^G(s) \\
=& \int_0^{\nu} H_{t1}^2(s)(1-G(s-))(1-F_Y(s-))\frac{(1-G(s-)dG(s)}{(1-G(s-))(1-G(s-))} \\
=& \int_0^{\nu} \mathbb{E}\Big[U_{1G}I[s < U_1]\Big]H_{t1}(s)\frac{dG(s)}{(1-G(s-))}
\end{aligned}$$

where the second equality follows because of $\mathbb{P}[U_1 \geq s] = (1-H(s-)) = (1-G(s-))(1-F_Y(s-))$ and the definition of $\Lambda^G(s)$ before. In addition, we have:

$$\frac{1}{\sqrt{n}}\sum_{m=1}^{n}\int_{\nu}^{\phi_Y} H_{t1}(s)dM_m(s) \xrightarrow{p} 0$$

as $\nu \longrightarrow \phi_Y$. By Rebelledo's martingale central limit theorem (CLT), we obtain:

$$M_{2nt} \xrightarrow{p} N\Big(0, \Phi_3\Big)$$

with

$$\Phi_3 = \int_0^{\phi_Y} \mathbb{E}\Big[U_{1G}I[s < U_1]\Big]H_{t1}(s)\frac{dG(s)}{(1-G(s-))}$$

Thus:

$$\sqrt{n}\frac{1}{n}\sum_{m=1}^{n}\Big(U_{m\widehat{G}} - U_{mG}\Big)\Big(\int I_i\nabla_\theta\psi_2(X_k,\theta_0)dF_{X,V}(X_k,V_i)\Big) \xrightarrow{d} N\Big(0,\Omega_3\Big)$$

where Ω_3 was defined in the statement of the lemma.

It remains to show that $U_{12n}(\theta)$ is of smaller order. Notice that by a similar argument to before and B1, one can show that uniformly for $0 < \nu < \phi_Y$:

$$\mathbb{E}\Big[\Big(\Big(U_{j\widehat{G}} - U_{jG}\Big)\Big(\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta) - \mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\Big)\Big)^2\Big] = O\Big(\frac{\|\theta-\theta_0\|}{h^d}\Big) + O\Big(\|\theta-\theta_0\|^2\Big) + o(1)$$

which is $O(\|\theta-\theta_0\|h^{-d})$ by B8. A similar line of argument and B1 can used to show that the leading term of

$$\mathbb{E}\Big[\Big|\Big(U_{j\widehat{G}} - U_{jG}\Big)\Big(\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta) - \mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\Big)\Big|^{\kappa}\Big]$$

is $O(\|\theta - \theta_0\|h^{-d(\kappa-1)})$ given B8. Thus, applying Rosenthal's inequality we obtain:

$$\begin{aligned}\mathbb{E}\Big[\Big(U_{12n}(\theta)\Big)^{2\kappa}\Big] \leq & n^{-8\kappa}\Xi_\kappa(\|\theta-\theta_0\|)\Big(n^{4\kappa}h^{-d\kappa} + n^4 h^{-d(2\kappa-1)}\Big)\Big) \\ =& O\Big(\|\theta-\theta_0\|n^{-4\kappa}h^{-d\kappa}\Big) + O\Big(\|\theta-\theta_0\|n^{-8\kappa+4}h^{-d(2\kappa-1)}\Big)\end{aligned}$$

By Markov's inequality, we have that $U_{12n}(\theta) = o_p(n^{-1})$ uniformly over $o_p(1)$ neighbourhoods of θ_0 for $\kappa = 1$. ■

Proof of Theorem 4

We denote by $\mathbb{E}_*$ and var_* the mean and variance operators of the bootstrapping sampling. In addition, let $O_p^*(1)$ and $o_p^*(1)$ be the orders of magnitude according to the bootstrapping distribution.

Using a similar argument to Goncalves and White (2005), the theorem follows once we show that:

$$\mathbb{E}_*\Big[\sqrt{m}(\theta^* - \widehat{\theta})\Big] = o_p(1) \qquad \text{(A-2.16)}$$

$$\text{var}_*\Big(\sqrt{m}(\theta^* - \widehat{\theta})\Big) = \text{var}\Big(\sqrt{n}(\widehat{\theta} - \theta_0)\Big) + O_p\Big(\frac{1}{\sqrt{n}}\Big) \qquad \text{(A-2.17)}$$

and for some $\delta > 0$:

$$\mathbb{E}_*\Big[\Big(\sqrt{m}\|\theta^* - \theta_0\|\Big)^{2+\delta}\Big] = O_p(1) \qquad \text{(A-2.18)}$$

Equations (A-2.16) and (A-2.17) follow automatically once we have verified that $\sqrt{m}(\theta^* - \widehat{\theta})$ has the same limiting distribution as $\sqrt{n}(\widehat{\theta} - \theta_0)$ up to an error of smaller order. Thus, we show that Lemma B1 and B2 are also applicable to the bootstrap estimator in (2.13) (with n being replaced by m in both lemmata). Since the proof is rather lengthy and in large parts identical to before, we will only sketch the one of asymptotic normality and $\sqrt{m}$-consistency paralleling the proof of Theorem 3. Consistency follows in fact by similar arguments to the proof of Theorem 2 and the ones presented in the following.

The equation in (2.13) can be decomposed as in the proof of Theorem 3. That is, we examine:

$$\frac{1}{m(m-1)}\sum_{k\neq l}\mu(X_k^*) \times A_{lk}^*(\theta) + \frac{1}{m(m-1)}\sum_{k\neq l}\Big(\widehat{\mu}^*(X_k^*) - \mu(X_k^*)\Big) \times A_{lk}^*(\theta)$$

with $A_{lk}^*(\theta) = I[X_k^* \in \mathcal{X}]\big\{I[X_k^{*\prime}\beta(\theta) > X_l^{*\prime}\beta(\theta)] - I[X_k^{*\prime}\beta(\widehat{\theta}) > X_l^{*\prime}\beta(\widehat{\theta})]\big\}$. In a first step we show that:

$$\frac{1}{m(m-1)}\sum_{k\neq l}\mu(X_k^*)A_{lk}^*(\theta) \tag{A-2.19}$$

behaves as

$$S_{1n}=\frac{1}{n(n-1)}\sum_{k\neq l}\mu(X_k)A_{lk}(\theta)$$

from the proof of Theorem 3. Since (A-2.19) is again a second order U-statistic for every $\theta\in\Theta$, the same Hoeffding decomposition argument as in Lemma A.3 of Khan (2001) used in the proof of Theorem 3 can be applied: first notice that the conditional expectation over bootstrap samples given X_k and X_l, respectively, is:

$$\begin{aligned}\psi^*(X_k^*,\theta)=&\frac{1}{2}\Big\{\mathbb{E}_*\Big[\mu(X_k^*)A_{lk}^*(\theta)\Big|X_k^*\Big]+\mathbb{E}_*\Big[\mu(X_l^*)A_{kl}^*(\theta)\Big|X_k^*\Big]\Big\}\\=&\frac{1}{2}\Big\{\frac{1}{n}\sum_{l=1}^{n}\mu(X_k^*)A_{lk^*}(\theta)+\frac{1}{n}\sum_{l=1}^{n}\mu(X_l)A_{lk^*}(\theta)\Big\}\end{aligned} \tag{A-2.20}$$

where the subscript without star in the second line indicates the summable variable. Hence:

$$\mathbb{E}_*\Big[\psi^*(X_k^*,\theta)\Big]=\frac{1}{2}\frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\Big\{\mu(X_k)A_{lk}(\theta)+\mu(X_l)A_{lk}(\theta)\Big\}$$

This term can be expanded further to give:

$$\mathbb{E}\Big[\psi(X_k,\theta)\Big]+\Big(\frac{1}{n}\sum_{k=1}^{n}\psi^*(X_k,\theta)-\mathbb{E}\Big[\psi(X_k,\theta)\Big]\Big)=T_{1n}^*+T_{2n}^* \tag{A-2.21}$$

T_{1n}^* can be expanded as in Lemma A.3 of Khan (2001). For T_{2n}^* on the other hand, let $\phi_{lk}(\theta)=\Big\{\mu(X_k)A_{lk}(\theta)+\mu(X_l)A_{lk}(\theta)\Big\}-\mathbb{E}\Big[\Big\{\mu(X_k)A_{lk}(\theta)+\mu(X_l)A_{lk}(\theta)\Big\}\Big]$. Notice that by B2, B5, and B6 we have $\mathbb{E}\Big[\Big|\phi_{lk}(\theta)\Big|^{\kappa}\Big]=\|\theta-\theta_0\|$ and $\mathbb{E}\Big[\phi_{lk}(\theta)^2\Big]=\|\theta-\theta_0\|$. Thus, by Rosenthal's inequality:

$$\mathbb{E}\Big[{T_{2n}^*}^{2\kappa}\Big]\leq n^{-4\kappa}\Xi_\kappa\Big(\|\theta-\theta_0\|n^{2\kappa}+\|\theta-\theta_0\|n^2\Big)$$

For $\kappa=1$, $\mathbb{E}\Big[{T_{2n}^*}^{2\kappa}\Big]=O(\|\theta-\theta_0\|n^{-2})$ and thus by Markov's inequality $T_{2n}(\theta)=o_p\Big(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\Big)$. Next, we show that 'm out of n' bootstrap is also able to mimic the random elements of the 'projection' of the U-statistic used in Lemma A.3 of Khan (2001). Notice that:

$$\begin{aligned}&\frac{1}{m}\sum_{i=1}^{m}\Big\{\psi^*(X_k^*=X_i^*,\theta)-\mathbb{E}_*\Big[\psi^*(X_k^*=X_i^*,\theta)\Big]\Big\}\\=&\frac{1}{m}\sum_{i=1}^{m}\Big\{\psi(X_k^*=X_i^*,\theta)-\mathbb{E}\Big[\psi(X_k^*=X_i^*,\theta)\Big]\Big\}+o_p(1)\end{aligned}$$

where the $o_p(1)$ term follows again by a subsequent application of Rosenthal's and Markov's inequality. The term in curley brackets can be dealt with by the same arguments as in Lemma A.3 of Khan (2001) since $m\longrightarrow\infty$ as $n\longrightarrow\infty$. Finally, the quadratic term of the Hoeffding decomposition can be shown to be $o_p\left(\frac{1}{m}\right)$ uniformly over $o_p(1)$ neighbourhoods of $\widehat{\theta}$ by expanding this term as above and subsequently applying Rosenthal's and Markov's inequality.

Next, we give a rough sketch of the steps to show that:

$$\frac{1}{m(m-1)}\sum_{k\neq l}\Big(\widehat{\mu}^*(X_k^*)-\mu(X_k^*)\Big)A_{lk}^*(\theta) \tag{A-2.22}$$

behaves as

$$S_{2n}=\frac{1}{n(n-1)}\sum_{k\neq l}\Big(\widehat{\mu}(X_k)-\mu(X_k)\Big)A_{lk}(\theta)$$

A similar decomposition as in the proof of Theorem 3 yields:

$$\begin{aligned}
&\frac{1}{m(m-1)}\sum_{k\neq l}\Big(\widetilde{\mu}^*(X_k^*)-\mu(X_k^*)\Big)A_{lk}^*(\theta)+\frac{1}{m(m-1)}\sum_{k\neq l}\Big(\widehat{\mu}^*(X_k^*)-\widetilde{\mu}^*(X_k^*)\Big)A_{lk}^*(\theta)\\
=&S_{21n}^*+S_{22n}^*
\end{aligned}$$

where $\widetilde{\mu}^*(\cdot)$ is defined analogously to the proof of Theorem 3. We start with S_{21n}^*, which can again be rewritten as:

$$S_{21n}^*=\frac{1}{m(m-1)}\sum_{k\neq l}\left\{\frac{1}{m}\sum_{i=1}^{m}\frac{\frac{1}{mh^{*d}}\sum\limits_{j=1}^{m}I_i^*\big(U_{jG}^*-\mu(X_k^*)\big)\mathbf{K}_{h^*,j}(X_k^*,V_i^*)}{\frac{1}{mh^{*d}}\sum\limits_{j=1}^{m}I_i^*\mathbf{K}_{h^*,j}(X_k^*,V_i^*)}\right\}\times A_{lk}^*(\theta)$$

Using B2, B3, B8, and the same argument as in the proof of Lemma B3, we can replace $\widehat{f}_{X,V}^*(X_k^*,V_i^*)$ by $f_{X,V}(X_k^*,V_i^*)$. After omitting terms with $k=l=j=i$, which, parallel to Theorem **??**, results in an error of order $o_p^*(\|\theta-\theta_0\|/mh^{*d})$), the numerator is given by:

$$\frac{1}{m(m-1)(m-2)(m-3)}\sum_{k\neq l\neq i\neq j}\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k^*,V_i^*)}I_i^*\big(U_{jG}^*-\mu(X_k^*)\big)\mathbf{K}_{h^*,j^*}(X_k^*,V_i^*)A_{lk}^*(\theta) \tag{A-2.23}$$

Using a similar decomposition as in (A-2.21), it is straightforward to show that this fourth order U-statistic is degenerate in ξ_k^*,ξ_l^*,ξ_i^* for each $\theta\in\Theta$, where ξ_k^*,ξ_l^*,ξ_i^* are defined as in the proof of Lemma B3. As before, this also implies that:

$$\begin{aligned}
&\mathbb{E}_*\Big[\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k^*,V_i^*)}I_i^*\big(U_{jG}^*-\mu(X_k^*)\big)\mathbf{K}_{h^*,j^*}(X_k^*,V_i^*)A_{lk}^*(\theta)\Big]\\
=&\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{X,V}(X_k,V_i)}I_i\big(U_{jG}-\mu(X_k)\big)\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)\Big]\\
&+\left\{\frac{1}{n^4h^{*d}}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{1}{f_{X,V}(X_k,V_i)}I_i\big(U_{jG}-\mu(X_k)\big)\mathbf{K}_{h^*,j}(X_k,V_i)A_{lk}(\theta)\right.\\
&\left.-\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{X,V}(X_k,V_i)}I_i\big(U_{jG}-\mu(X_k)\big)\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)\Big]\right\}\\
=&o_p\Big(\frac{1}{n}\Big)
\end{aligned}$$

where the last equality follows by B8, the Rosenthal's and Markov's inequalities, and the fact that $\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{X,V}(X_k,V_i)}I_i\big(U_{jG}-\mu(X_k)\big)\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)\Big]=0$. The second term of the Hoeffding projection of (A-2.23) yields for each $\theta\in\Theta$:

$$
\begin{aligned}
&\frac{1}{m}\sum_{p=1}^{m}\mathbb{E}_*\Big[\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k^*,V_i^*)}I_i^*\big(U_{jG}^*-\mu(X_k^*)\big)\mathbf{K}_{h^*,j^*}(X_k^*,V_i^*)A_{lk}^*(\theta)\Big|\xi_j^*=\xi_p^*\Big]\\
=&\frac{1}{m}\sum_{p=1}^{m}\int\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k,V_i)}I_i\big(U_{pG}^*-\mu(X_k)\big)\mathbf{K}_{h^*,p^*}(X_k,V_i)A_{lk}(\theta)dF_{X,V}(X_k,X_l,V_i)+o_p\Big(\frac{1}{m}\Big)
\end{aligned}
$$

where $o_p\left(\frac{1}{m}\right)$ follows by another application of Rosenthal's inequality. The term $\int\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k,V_i)}I_i\big(U_{pG}^*-\mu(X_k)\big)\mathbf{K}_{h^*,p^*}(X_k,V_i)A_{lk}(\theta)dF_{X,V}(X_k,X_l,V_i)$ can now be treated as in the proof of Theorem 3 using a second order Taylor expansion:

$$
(\theta-\widehat{\theta})'I_i\big(U_{iG}^*-\mu(X_k)\big)\nabla_\theta\psi_2(X_k,\widehat{\theta})+O(\|\theta-\widehat{\theta}\|^2)
$$

In view that:

$$
\begin{aligned}
&\mathrm{var}_*\Big(\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\mathbb{E}_*\Big[\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k^*,V_i^*)}I_i^*\big(U_{jG}^*-\mu(X_k^*)\big)\mathbf{K}_{h^*,j^*}(X_k^*,V_i^*)A_{lk}^*(\theta)\Big|\xi_j^*\Big]\Big)\\
=&\mathrm{var}_*\Big(\mathbb{E}_*\Big[\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k^*,V_i^*)}I_i^*\big(U_{jG}^*-\mu(X_k^*)\big)\mathbf{K}_{h^*,j^*}(X_k^*,V_i^*)A_{lk}^*(\theta)\Big|\xi_j^*\Big]\Big)\\
=&\Omega_0+o_p(1)
\end{aligned}
$$

and since all higher order degenerate U-statistics from the decomposition are of smaller order, the term in (A-2.23) weakly converges to $N(0,\Omega_0)$ as both m and n go to infinity thus mimicking the limiting distribution of (A-2.13). S_{22n}^* and the three leading terms arising in analogy to S_{22n} can be treated as in the proof of Lemmata B4 to B6 using similar arguments to above. That is, the same decomposition as in the proof of those lemmata yields the remaining variance pieces Ω_1, Ω_2, and Ω_3 as m and n go to infinity. It follows that $\sqrt{m}(\theta^*-\widehat{\theta})$ has the same limiting distribution as $\sqrt{n}(\widehat{\theta}-\theta_0)$.

To verify equation (A-2.18), notice that from Lemma B1 and B2, $(\theta^*-\widehat{\theta})$ can be restated as:

$$
(\theta^*-\widehat{\theta})=-J^{-1}\frac{1}{\sqrt{m}}W_m+o_p\Big(\frac{1}{m}\Big)
$$

where the $o_p(m^{-1})$ follows from $\sqrt{m}$-consistency. Thus, $\sqrt{m}(\theta^*-\widehat{\theta})=-J^{-1}W_m+o_p\Big(\frac{1}{\sqrt{m}}\Big)$. Recalling that $J=\frac{1}{2}\mathbb{E}\big[\nabla_{\theta\theta'}\psi_1(X_k,\widehat{\theta})\big]$ is bounded and negative definite by B6, we can bound equation (A-2.18) as follows:

$$
\begin{aligned}
\mathbb{E}_*\Big[\Big(\sqrt{m}\|\theta^*-\theta_0\|\Big)^{2+\delta}\Big]=&\mathbb{E}_*\Big[\Big(\|-J^{-1}W_m\|\Big)^{2+\delta}\Big)\Big]+o_p(1)\\
\leq&\Xi_J\|W_n\|^{2+\delta}=O_p(1)
\end{aligned}
$$

where Ξ_J is a generic constant and the last equality follows since W_n converges in distribution.

■

A2.3 Tables

Table 2.1: Monte Carlo Simulation

Design I: Linear Model & No Censoring					
No. of obs.	Estimator	Mean Bias[1]	Median Bias[1]	RMSE[1]	MAD[1]
$n = 50$	RankCF	0.0392	0.0291	0.4778	0.2036
	MRE	0.5200	0.3774	0.7878	0.5327
	OLS	0.4943	0.4179	0.6085	0.4949
	TSLS	−0.0193	−0.0103	0.5317	0.3170
$n = 100$	RankCF	0.0448	0.0125	0.1976	0.1372
	MRE	0.4874	0.4097	0.5778	0.4876
	OLS	0.4479	0.4209	0.4926	0.4479
	TSLS	0.0170	0.0179	0.2640	0.2039
$n = 200$	RankCF	−0.0017	−0.0191	0.1164	0.0925
	MRE	0.4404	0.4361	0.4702	0.4404
	OLS	0.4241	0.4010	0.4445	0.4241
	TSLS	−0.0085	−0.0093	0.1772	0.1402
$n = 400$	RankCF	−0.0241	−0.0359	0.0805	0.0665
	MRE	0.4130	0.4180	0.4233	0.4130
	OLS	0.4033	0.4016	0.4124	0.4033
	TSLS	0.0023	0.0127	0.1159	0.0929
$n = 800$	RankCF	−0.0260	−0.0303	0.0687	0.0552
	MRE	0.4004	0.4009	0.4061	0.4004
	OLS	0.3939	0.3928	0.3986	0.3939
	TSLS	−0.0036	−0.0104	0.0736	0.0577
Design II: Non-linear Model & No Censoring					
No. of obs.	Estimator	Mean Bias[1]	Median Bias[1]	RMSE[1]	MAD[1]
$n = 100$	RankCF	0.3911	0.2200	0.6632	0.4509
	MRE	0.9628	0.9950	1.0410	0.9628
	MRC	0.9697	1.0100	1.0530	0.9697
$n = 200$	RankCF	0.3378	0.2400	0.5415	0.3738
	MRE	1.0305	1.0300	1.0879	1.0305
	MRC	1.0031	0.9750	1.0655	1.0031
$n = 400$	RankCF	0.2948	0.2200	0.4325	0.3035
	MRE	1.0846	1.0850	1.1199	1.0846
	MRC	1.0710	1.0700	1.1054	1.0710
$n = 800$	RankCF	0.2499	0.2400	0.3106	0.2582
	MRE	1.1180	1.1200	1.1411	1.1180
	MRC	1.1150	1.0900	1.1394	1.1150

[1] The figures in the table represent the average of the corresponding bias measure (401 replications).

Table 2.2: Monte Carlo Simulation - Censoring

No. of obs.	Estimator	Mean Bias[1]	Median Bias[1]	RMSE[1]	MAD[1]
Design III: Linear Model & No Censoring					
$n = 100$, Av. Censor. Ratio: 0	RankCF	0.0121	−0.0300	0.1667	0.1309
	MRE	0.4392	0.3800	0.5016	0.4394
	MRC	0.4422	0.3800	0.5186	0.4424
$n = 200$, Av. Censor. Ratio: 0	RankCF	−0.0027	−0.0200	0.1296	0.0972
	MRE	0.4426	0.4200	0.4733	0.4426
	MRC	0.4509	0.4200	0.4876	0.4509
$n = 400$, Av. Censor. Ratio: 0	RankCF	−0.0129	−0.0300	0.0885	0.0692
	MRE	0.4317	0.4200	0.4466	0.4317
	MRC	0.4298	0.4200	0.4445	0.4298
$n = 800$, Av. Censor. Ratio: 0	RankCF	−0.0216	−0.0200	0.0647	0.0523
	MRE	0.4192	0.4100	0.4257	0.4192
	MRC	0.4176	0.4100	0.4253	0.4176
Design IV: Linear Model & Censoring (light)					
No. of obs.	Estimator	Mean Bias[1]	Median Bias[1]	RMSE[1]	MAD[1]
$n = 100$, Av. Censor. Ratio: .25	RankCF	0.2438	0.0300	0.7721	0.5631
	MRE	0.5046	0.4050	0.6316	0.5108
	MRC	0.5087	0.4200	0.6221	0.5110
$n = 200$, Av. Censor. Ratio: .25	RankCF	0.1024	−0.0500	0.5392	0.3729
	MRE	0.4720	0.4200	0.5428	0.4720
	MRC	0.4570	0.4100	0.5151	0.4570
$n = 400$, Av. Censor. Ratio: .25	RankCF	0.0659	−0.0700	0.4417	0.3050
	MRE	0.4293	0.4100	0.4636	0.4293
	MRC	0.4306	0.4200	0.4648	0.4306
$n = 800$, Av. Censor. Ratio: .25	RankCF	0.0794	0.0000	0.3383	0.2247
	MRE	0.4384	0.4300	0.4529	0.4384
	MRC	0.4371	0.4300	0.4495	0.4371
Design V: Linear Model & Censoring (heavy)					
No. of obs.	Estimator	Mean Bias[1]	Median Bias[1]	RMSE[1]	MAD[1]
$n = 100$, Av. Censor. Ratio: .35	RankCF	0.2526	0.0450	0.9521	0.7395
	MRE	0.5184	0.4100	0.6619	0.5318
	MRC	0.5099	0.4100	0.6455	0.5251
$n = 200$, Av. Censor. Ratio: .35	RankCF	0.2296	−0.0600	0.7441	0.5388
	MRE	0.4864	0.4300	0.5827	0.4864
	MRC	0.4726	0.4100	0.5581	0.4726
$n = 400$, Av. Censor. Ratio: .35	RankCF	0.1431	−0.0550	0.6501	0.4520
	MRE	0.4423	0.4100	0.4943	0.4423
	MRC	0.4219	0.4000	0.4641	0.4219
$n = 800$, Av. Censor. Ratio: .35	RankCF	0.1322	0.0150	0.4783	0.3186
	MRE	0.4426	0.4300	0.4626	0.4426
	MRC	0.4409	0.4300	0.4594	0.4409

[1] The figures in the table represent the average of the corresponding bias measure (401 replications).

Table 2.3: Empirical Illustration - Earnings Study

Estimator	Coefficient[1]	Value	90% Bootstrap-CI
RankCF	Constant	–	
	Education	0.1181	[0.0608;0.1754]
	Age	0.5794	[0.3339;0.8249]
	Age^2	−0.0057	[−0.0079;−0.0036]
MRE	Constant	–	
	Education	0.2042	[0.1809;0.2275]
	Age	0.5393	[0.4326;0.6460]
	Age^2	−0.0054	[−0.0064;−0.0045]
MRC	Constant	–	
	Education	0.2192	[0.1945;0.2439]
	Age	0.6767	[0.5631;0.7902]
	Age^2	−0.0064	[−0.0074;−0.0054]
OLS	Constant	−2.3042	[−6.9826;2.3742]
	Education	0.2052	[0.1804;0.2299]
	Age	0.6534	[0.4837;0.8231]
	Age^2	−0.0063	[−0.0078;−0.0049]
LAD	Constant	−9.4916	[−14.0756;−4.9075]
	Education	0.2125	[0.1900;0.2349]
	Age	0.9227	[0.7519;1.0935]
	Age^2	−0.0086	[−0.0100;−0.0071]
TSLS	Constant	−2.9967	[−7.8251;1.8318]
	Education	0.2313	[0.1720;0.2906]
	Age	0.6764	[0.5010;0.8518]
	Age^2	−0.0065	[−0.0081;−0.0050]

[1] The gender coefficient has been normalized to one.

3 Do Reservation Wages Decline Monotonically? A Novel Statistical Test

3.1 Introduction

This paper develops a test for monotonicitiy of the regression function when the continuous regressor of interest is endogenous. To the best of the author's knowledge, this case has not yet been studied in the literature, but it is argued that such a testing framework is relevant for various setups in Labour Economics and Industrial Organization. In particular, the paper provides an application to formally evaluate the monotonoicity of the reservation wage as a function of elapsed unemployment duration, which refers to the length of an unemployment spell at the time the reservation wage information is being retrieved.

Reservation wages lie at the heart of many partial and general equilibrium job search models and are viewed as a key determinant for the length of unemployment (Mortensen, 1986). However, the effect of unemployment duration on the reservation wage is generally ambiguous and difficult to measure since both variables are determined simultaneuosly if reservation wages are flexible. Numerous papers have assessed the impact of unemployment duration on the reservation wage using either structural approaches (Kiefer and Neumann, 1979; Lancaster, 1985; van den Berg, 1990) or instrumental variable methods (Addison, Centeno, and Portugal, 2004; Brown and Taylor, 2009). Despite some evidence for an overall declining reservation wage function over the course of unemployment, it is not yet well understood whether this decline is monotonic and whether it holds across different subgroups of the (unemployed) population.

Using a standard partial equilibrium job search model, it is shown that monotonicity, a restriction that has been imposed by several empirical studies (Kiefer and Neumann, 1981; Lancaster, 1985; Addison, Centeno, and Portugal, 2004; Brown and Taylor, 2009), only holds under certain conditions on the variables in the model. The paper sheds light on this monotonicity aspect by developing a test that can evaluate the restriction while addressing endogeneity either through a nonparametric control function argument (e.g. Newey, Powell, and Vella, 1999; Blundell and Powell, 2003) or through unobservable exogenous variation in the endogenous variable of interest (Matzkin, 2004). The test is set up to detect whether reservation wages decline monotonically for certain subsets of the support of elapsed unemployment duration conditional on different characteristics. That is, the test aims to give an answer to questions such as: does the reservation wage of a male decrease over the first three months of unemployment?[1] Knowledge about this kind of questions has policy implications since interventions may be designed accordingly. For instance, an increase in the reservation wage after an initial decline, which might be due to individuals becoming more selective the longer search lasts, could suggest implementing policies that enforce search or intensify search assistance in particular from the point when reservation wages increase again. Alternatively, unemployed individuals below a certain age typically undergo a tighter benefit regime and face more drastic sanctions than their older counterparts, which might result in a reservation wage that is monotonically declining throughout their unemployment. On the other hand, these sanctions or changes in the benefit level are often absent or less pronounced for unemployed individuals above a certain age (in particular the ones close to retirement age). Thus, reservation wages could, by contrast, behave quite differently for older unemployed. As mentioned above, this has general implications for the design of policies addressing unemployment.

From a theoretical perspective, the contribution of the paper is to combine different conditional mean estimators for the regression function with a test for monotonicity based on Ghosal, Sen, and van der Vaart (2000) and to derive its asymptotic proper-

[1] Notice that the notion of 'decreasing' here and in the following is understood as non-increasing during the entire period considered.

ties. The first step estimator can be constructed in multiple ways:[2] if the researcher has a suitable instrument (vector) at his disposal that gives rise to a conditional mean independence assumption, a consistent two-step estimator as in Gutknecht (2011) can be used following standard arguments from the control function literature (e.g. Newey, Powell, and Vella, 1999; Blundell and Powell, 2003). In a first step, the mean of the dependent variable conditional on exogenous covariates and the estimated control function is estimated using standard kernel methods. Subsequently, the control function is averaged out to yield the empirical conditional mean function of interest. Alternatively, if no appropriate instrumental variables together with a conditional mean independence condition exist, but instead variables that represent an exogenous perturbation of the endogenous regressor are available, the concept of 'unobservable instruments' (Matzkin, 2004) can be applied. It is shown that by assuming the existence of such an exogenous perturbation that can be integrated into the conditioning set, one may still identify and estimate the nonparametric regression function of interest using additive separability conditions together with backfitting methods (Mammen, Linton, and Nielsen, 1999).

After having constructed the first stage, either estimator can be plugged into a modified test statistic that is taken to be the supremum of a suitably rescaled second order U-process. The asymptotic distribution of this statistic can be approximated by a stationary Gaussian process with a covariance that resembles the one in Ghosal, Sen, and van der Vaart (2000). The main difference w.r.t. the latter consists of the estimated regression function that forms part of the modified test statistic and that requires extra consideration in the derivation of the limiting distribution (see also Lee, Linton, and Whang (2009)). The test is shown to be consistent against fixed general alternatives and its finite sample performance is studied in a Monte Carlo Simulation.

Tests for monotonicity of the regression function have been a long-standing topic in the statistical literature and numerous other tests have been developed: Bowman, Jones, and Gijbels (1998) for instance use Silvermans (1981) 'critical bandwidth'

[2]It is the aim of the author to extend the current setup also to nonparametric instrumental variable regression if the researcher has a corresponding moment condition at hand to construct a suitable first stage estimator (e.g. Chen and Pouzo, 2009).

approach to construct a bootstrap test for monotonicity, while Gijbels, Hall, Jones, and Koch (2000) consider the length or runs of consecutive negative values of observation differences and Hall and Heckman (2000) suggest to fit straight lines through subsequent groups of consecutive points and reject monotonicity for too large negative values of the slopes. A more recent example are Birke and Neumeyer (2010), who base their test on different empirical processes of residuals. All these tests do, however, require independence between the equation error and the regressor of interest and are hence not applicable to a wide range of economic setups that allow for a correlation of the latter. A generalization of the above tests to monotonicity of nonparametric conditional distributions has recently been carried out by Lee, Linton, and Whang (2009). Their test statistic is similar to the one of Ghosal, Sen, and van der Vaart (2000), albeit the asymptotic distribution takes a different and more complicated form. Another test for monotonicity of conditional distributions and its moments has been proposed by Delgado and Escanciano (2012). Even though in both examples the null of stochastic monotonicity implies monotonicity of the regression function (if it exists), rejection of the null does clearly not imply a failure of monotonicity of the regression function.

Changing reservation wages raise a simultaneity issue since the reservation wage does not only influence unemployment, but is in turn also affected by the length of unemployment itself. This inter-relationship is well understood and has aptly been discussed in the job search literature (e.g. Lancaster, 1985; van den Berg, 1990). The identification approach of this paper uses instrumental variables suggested by the literature such as logarithm of benefit income other than unemployment benefits, logarithm pay in the last job, an indicator variable for having a working spouse, marital status, or the number of dependent children (Kiefer and Neumann, 1979; Addison, Centeno, and Portugal, 2004; Brown and Taylor, 2009) to construct control functions that are plugged into the conditional mean estiamtor.[3] To check robustness of the results, the paper also proposes an alternative method based on a recent study by Addison, Machado, and Portugal (2011), who address endogeneity by using longitudinal information on completed durations: assuming that endogeneity

[3]The crucial underlying assumption here will be that all these instrumental variables impact the reservation wage only through elapsed unemployment duration.

arises due to an omitted, endogenous 'fixed effect' that is constant throughout the unemployment spell, an additively separable nonparametric model can be fitted to the data controlling contemporaneously for elapsed and completed unemployment duration. It is shown that controlling for both durations, together with suitable additivity assumptions, allows to recover the regression function of interest.[4] The estimated reservation wage function can then be plugged into the test statistic described before.

The data for the empirical analysis stems from the British Household Panel Survey (BHPS), a nationally representative survey on individuals from more than 5,000 households in the UK. This data source provides sufficient information on (hourly) reservation wages, unemployment spells, and instrumental variables to conduct the monotonicity test for different population subgroups. More generally, however, the testing framework can also be applied to other fields in economics where a formal evaluation of monotonicity is of interest to the researcher and assumptions set out in this paper are met. Examples include the relationship between the hourly wage rate and the number of annual hours worked (Vella, 1993) or returns to years of schooling (Garen, 1984) if years of schooling is modelled as a continuous choice variable.

The paper is organised as follows: Section 3.2 outlines the main setup and the test statistic if a suitable instrumental variable is at hand, while large sample properties of the statistic are examined in Section 3.3. Section 3.4 extends the framework of the previous sections to the case of 'unobservable instruments' as outlined above. Section 3.5 examines the finite sample properties of the estimator in a Monte Carlo Simulation. Section 3.6 will then be devoted to the reservation wage example and will provide a motivation for the methods suggested to address endogeneity and for testing monotonicity in that context. It will also outline the results from an application to UK unemployment data. Section 3.7 concludes. All proofs and tables are postponed to the appendix.

[4] A theoretical model will demonstrate how such an endogenous fixed effect can be incorporated into a standard partial equilibrium job search model.

3.2 Setup

To better understand the setup, consider the following model: let W_i be the continuous outcome variable (e.g. the 'reservation wage' from the application example). U_i is a continuous, endogenous regressor (e.g. elapsed unemployment duration) and X_i is a D dimensional random vector of exogenous characteristics of the individual. The vector may contain both, continuous as well as discrete elements. Then, with ϵ_i denoting the unobservable, the equation of interest is given by:

$$W_i = \widetilde{m}\Big(X_i, U_i\Big) + \epsilon_i \tag{3.1}$$

where $\widetilde{m}(\cdot,\cdot)$ is a real-valued function, which is differentiable in its continuous arguments. Before the test statistic is outlined, a few remarks on the notation are required: let $\mathcal{X}, \mathcal{U}$ denote subsets of the support of X and U with strictly positive density everywhere.[5] Moreover, let $\nabla_U \widetilde{m}(\cdot,\cdot)$ denote the derivative of $\widetilde{m}(\cdot,\cdot)$ w.r.t. the argument U_i, and $\mathcal{T} = [a, b]$ be a compact interval s.t. $\mathcal{T} \subset \mathcal{U}$. Reverting to the application example of reservation wages from the introduction, suppose the interest lies in testing whether the reservation wage function (for an individual with characteristics x) is declining for every elapsed unemployment duration $t \in \mathcal{T}$. That is, for a specific $x \in \mathcal{X}$, the null hypothesis is given by:

$$H_0 : \quad \nabla_U \widetilde{m}(x, t) \leq 0 \quad \text{for all} \quad t \in \mathcal{T}$$

The alterative is the negation of this null hypothesis. Notice that the above hypothesis can be restated as $-\nabla_U \widetilde{m}(x, t) \equiv \nabla_U m(x, t) \geq 0$ for all $t \in \mathcal{T}$, which will turn out to be more convenient when setting up the test statistic. What prevents a direct application of the test for monotonicity of Ghosal, Sen, and van der Vaart (2000) based on observed W_i and U_i is that U_i and ϵ_i are correlated and thus inference based on the former test statistic will be misleading. As outlined in the

[5]For expositional motives, X_i is assumed to contain exclusively continuous components in this section. This does obviously restrict the applicability for many empirical studies with small samples where researchers are typically interested in a differentiation of the sample according to different (discrete) subgroups such as gender, race etc.. Hence, an extension to discrete components is considered in section 3.3.

introduction, the paper proposes different identification strategies that allow to implement the test despite this endogeneity problem. In the following, the paper will outline a control function procedure that can be applied if the researcher has instrumental variables at his disposal that give rise to a conditional mean independence condition. If instead of an instrumental variable, an exogenous perturbation of the endogenous regressor U_i exists, an identification strategy along the lines of Matzkin (2004) becomes applicable. The latter will be outlined as an extension in section 3.4.

Suppose a D_z-dimensional vector of instruments $Z_i = \{X_i, Z_{1i}\}$ exists. Moreover, the subvector Z_{1i} is assumed to be of dimension $D_{z1} \geq 1$ with at least one (non-constant) continuous component. The following reduced form equation is assumed:

$$U_i = g\Big(Z_i\Big) + V_i \tag{3.2}$$

where $g(\cdot)$ is a real-valued, differentiable with non-zero derivative in its continuous argument(s). V_i is the so called control function, which is assumed to satisfy a conditional mean independence condition:

$$\mathbb{E}\Big[\epsilon_i \Big| Z_i, V_i\Big] = \mathbb{E}\Big[\epsilon_i \Big| V_i\Big] \tag{3.3}$$

This restriction is crucial for identification purposes and referred to as 'exclusion restriction' in the literature. A sufficient condition is independence between the instrument vector Z_i and the model unobservables ϵ_i and V_i.[6] To revert to the example of reservation wages, instruments suggested by the literature might for instance be the logarithm of benefit income other than unemployment benefits, the pay of a previously held job, having a working spouse, marital status, household size, or the number of dependent children. In order for these instruments to be valid, one has to assume that the variables only affect the reservation wage through elapsed unemployment duration. That is, elapsed unemployment U_i is assumed to be a

[6] Notice that under this independence condition identification could be achieved even if the setup in (3.1) contained a nonseparable function $W_i = \widetilde{m}\Big(X_i, U_i, \epsilon_i\Big)$, where $\widetilde{m}(\cdot,\cdot,\cdot)$ is strictly increasing in its last argument. Such an extension would require a strengthening of the exclusion restriction to conditional independence of U_i and ϵ_i given V_i (see Blundell and Powell, 2003), which follows from independence of the instrument vector Z_i and the unobservables ϵ_i and V_i.

function of Z_i (see section 3.6 for details). Equations (3.1) and (3.2) together with the assumption in (3.3) and the normalization characterize a standard nonparametric control function setup for an additive regression function (e.g. Blundell and Powell, 2003).

Identification of $\widetilde{m}(\cdot,\cdot)$ can be achieved using for instance Theorem 2.3 of Newey, Powell, and Vella (1999). In order to understand their result, notice that:

$$\begin{aligned}
\mathbb{E}\Big[W_i\Big|U_i = u, Z_i = z\Big] =& \widetilde{m}\Big(x, u\Big) + \mathbb{E}\Big[\epsilon_i\Big|U_i = u, Z_i = z\Big] \\
=& \widetilde{m}\Big(x, u\Big) + \mathbb{E}\Big[\epsilon_i\Big|U_i = u, V_i = v\Big] \\
=& \widetilde{m}\Big(x, u\Big) + \mathbb{E}\Big[\epsilon_i\Big|V_i = v\Big] \\
\equiv& \widetilde{m}\Big(x, u\Big) + \lambda(v)
\end{aligned} \tag{3.4}$$

where the third equality follows from the conditional mean independence assumption in (3.3). Newey, Powell, and Vella (1999) show that identification of the additive $\widetilde{m}(\cdot,\cdot)$ from the conditional expectation above is the same as identification of $\widetilde{m}(\cdot,\cdot)$ in equation (3.1). Then, identification up to an additive constant can be accomplished by reverting to the following lemma:

Lemma 5. *Suppose that equations (3.1) and (3.2) and the conditional mean independence assumption in (3.3) as well as* $\mathbb{E}\Big[V_i\Big|Z_i\Big] = 0$ *hold. Moreover, assume that* $\mathbb{E}\Big[\epsilon_i\Big|V_i\Big]$ *is differentiable and that the boundary of the support of* (Z, V) *has zero probability. Then,* $\widetilde{m}(X, U)$ *is identified up to an additive constant.*[7]

The proof of the lemma follows directly from an application of Theorem 2.3 in Newey, Powell, and Vella (1999). The rank condition of that theorem is trivially satisfied

[7]The exact definition of identification in Newey, Powell, and Vella (1999, p.567) is based on equation (3.4):

$$\mathbb{E}\Big[W_i\Big|U_i = u, Z_i = z\Big] = \mathbb{E}\Big[W_i\Big|U_i = u, X_i = x, V_i = v\Big] \equiv \widetilde{m}\Big(x, u\Big) + \lambda(v)$$

Since conditional expectations are unique with probability one, any other additive function $\overline{m}(x, u) + \overline{\lambda}(v)$ satisfying the above equation must satisfy $\mathbb{P}[\overline{m}(x, u) + \overline{\lambda}(v) = \widetilde{m}(x, u) + \lambda(v)] = 1$. Identification is thus equivalent to equality of conditional expectations, which in turn implies equality of the additive components, up to a constant. Equivalently, working with the difference of two conditional expectations, identification is equivalent to the statement that a zero additive function must have only constant components. Hence, the authors obtain the following Theorem, which also provides their definition of identification: *Theorem 2.1 of Newey, Powell, and Vella (1999):* $\widetilde{m}(x, u)$ *is identified, up to an additive constant, if and only if* $\mathbb{P}[\delta(x, u) + \gamma(v) = 0] = 1$ *implies there is a constant* c_m *with* $\mathbb{P}[\delta(x, u) = c_m] = 1$.

here because our setup only contains one endogenous regressor and thus $\frac{\partial}{\partial Z} g(\cdot)$ has a vector format with rank one. Then, imposing the normalization $\mathbb{E}\Big[\epsilon_i\Big] = 0$ on ϵ_i such that $\mathbb{E}\Big[W_i\Big] = \mathbb{E}\Big[\widetilde{m}\Big(X_i, U_i\Big)\Big]$, identification of the level of $\widetilde{m}(\cdot,\cdot)$ can be accomplished by assuming the existence of a function $f(v)$ satisfying $\int f(v)dv = 1$:

$$\begin{aligned}
\int \mathbb{E}\Big[W_i\Big|U_i = u, X_i = x, V_i = v\Big] f(v)dv =& \int \mathbb{E}\Big[W_i\Big|U_i = u, X_i = x, V_i = v\Big] f(v)dv \\
&+ \mathbb{E}\Big[\int \mathbb{E}\Big[W_i\Big|U_i = u, X_i = x, V_i = v\Big] f(v)dv\Big] \\
&- \mathbb{E}\Big[W_i\Big] \\
=& \widetilde{m}\Big(x, u\Big) + \mathbb{E}\Big[\widetilde{m}\Big(X_i, U_i\Big)\Big] - \mathbb{E}\Big[W_i\Big] \\
=& \widetilde{m}\Big(x, u\Big)
\end{aligned} \tag{3.5}$$

where the second equality follows because $\int \mathbb{E}[W_i|U_i = u, X_i = x, V_i = v] f(v)dv = \widetilde{m}(x, u) + \int \lambda(v) f(v)dv$ and $\int \lambda(v) f(v)dv = \int \mathbb{E}[\epsilon_i|V_i = v] f(v)dv = \mathbb{E}[\epsilon_i] = 0$. This establishes identification.

In order to use the above result for the test statistic developed in this paper, recall that $m(\cdot,\cdot) = -\widetilde{m}(\cdot,\cdot)$ and let:

$$\mu(x, u, v) = -\mathbb{E}\Big[W_i\Big|X_i = x, U_i = u, V_i = v\Big] \tag{3.6}$$

so that:

$$\mu(x, u) = -\int \mathbb{E}\Big[W_i\Big|X_i = x, U_i = u, V_i = v\Big] f(v)dv = m(x, u) \tag{3.7}$$

for every $x, u \in \mathcal{X} \times \mathcal{U}$.[8] Equation (3.7) can be consistently estimated using the

[8] A similar argument could be made under conditional independence and non-separability of $\widetilde{m}(\cdot,\cdot,\cdot)$. That is: $\mu(x, u) = -\int \mathbb{E}\Big[W_i\Big|X_i = x, U_i = u, V_i = v\Big] f(v)dv = \mathbb{E}\Big[m(x, u, \epsilon_i)\Big|V_i = v\Big] f(v)dv = \mathbb{E}\Big[m(x, u, \epsilon_i)\Big]$. Notice that for this argument to hold for every $x, u \in \mathcal{X} \times \mathcal{U}$, a large support condition similar to the one in Imbens and Newey (2009) is required.

kernel estimator suggested in Gutknecht (2011):

$$\widehat{\mu}(x,u) = \frac{1}{n}\sum_{j=1}^{n}\widehat{\mu}(x,u,\widehat{V}_j) \tag{3.8}$$

where

$$\widehat{\mu}(x,u,\widehat{V}_j) = -\widehat{\mathbb{I}}_j \frac{\frac{1}{nh_n^3}\sum_{i=1}^{n} W_i K_h(x-X_i)K_h(u-U_i)K_h(\widehat{V}_j-\widehat{V}_i)}{\frac{1}{nh_n^3}\sum_{i=1}^{n} K_h(x-X_i)K_h(u-U_i)K_h(\widehat{V}_j-\widehat{V}_i)} \tag{3.9}$$

and $\widehat{\mathbb{I}}_j = \mathbb{I}[x \in \mathcal{X}, u \in \mathcal{U}, \widehat{V}_j \in \mathcal{V}]$ denotes an indicator function that is equal to one on the compact set of interest. $K_h(u) = K(u/h_n)$ is a kernel function with compact support on $[-1,1]$ and $h_n \longrightarrow 0$ as $n \longrightarrow \infty$. The empirical control function $\widehat{V}_i$ can be constructed using $\widehat{V}_i = U_i - \widehat{g}(Z_i)$, where $\widehat{g}(\cdot)$ is estimated using for instance the Nadaraya-Watson kernel estimator. The modified test statistic is based on $\widehat{\mu}(\cdot,\cdot)$ in (3.8): assuming that the observations are in ascending order $1 \leq i < j \leq n$, a suitable second order U-process based on $\widehat{\mu}(\cdot,\cdot)$ and U_i (and indexed by t) is given by:

$$\widehat{U}_{n,x}(t) = \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\Big(\widehat{\mu}(x,U_j)-\widehat{\mu}(x,U_i)\Big)\,\mathrm{sign}\Big(U_i-U_j\Big)\frac{1}{h_n^2}K_h(U_i-t)K_h(U_j-t) \tag{3.10}$$

where $t \in \mathcal{T}$ and

$$\mathrm{sign}(x) = \begin{cases} 1 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x = 0 \\ -1 & \text{if} \quad x < 0 \end{cases}$$

If $\nabla_U m(x,t) \geq 0$, $\widehat{U}_{n,x}(t)$ should, apart from random fluctuations due to estimation errors, be less than or equal to 0. To see this, replace $\widehat{\mu}(\cdot,\cdot)$ by $\mu(\cdot,\cdot)$ and recall that $\mu(x,u) = m(x,u)$. Hence, taking expectations of the modified $U_{n,x}(t)$ and letting

$\nu = ((U_j - t)/h_n)$ and $u = ((U_i - t)/h_n)$, by change of variables:

$$
\begin{aligned}
\mathbb{E}\Big[U_{n,x}(t)\Big] &= \int\int \Big(m(x,U_j) - m(x,U_i)\Big)\,\mathrm{sign}\Big(U_i - U_j\Big) \\
&\quad \times \frac{1}{h_n^2} K_h(U_i - t) K_h(U_j - t) f(U_i) f(U_j) du_i du_j \\
&= \int\int \Big(m(x, t + h_n\nu) - m(x, t + h_n u)\Big)\,\mathrm{sign}\Big(u - \nu\Big) \\
&\quad \times K(u)K(\nu) f(t + h_n u) f(t + h_n \nu) du d\nu
\end{aligned}
$$

where $f(\cdot)$ denotes a marginal density function. Notice that:

$$
\frac{1}{h_n}\Big(m(x, t + h_n\nu) - m(x, t + h_n u)\Big) \longrightarrow \nabla_U m(x,t)(\nu - u)
$$

and hence by dominated convergence:

$$
\frac{1}{h_n}\,\mathbb{E}\Big[U_{n,x}(t)\Big] \longrightarrow -\nabla_U m(x,t) \int\int |u - \nu| K(u)K(\nu) f(t)^2 du d\nu
$$

Thus, the limit is negative or zero if and only if $\nabla_U m(x,t) \geq 0$. So in expectation, the statistic should be less than or equal to zero under H_0. Vice versa, under the alternative the statistic should yield a positive value.

The test statistic is given as the supremum (on the interval $\mathcal{T}$) of a suitably scaled version of (3.10), which corresponds to the choice of similar tests in the literature rendering the test particularly sensitive to large positive outliers violating the null hypothesis.[9] Specifically, the statistic is chosen to be:

$$
S_n = \sup_{t \in \mathcal{T}} \left\{ \frac{\widehat{U}_{n,x}(t)}{c_n(t)} \right\}
$$

where $c_n(t)$ is a scaling factor that may depend on $(\{X_1, U_1\}, \ldots, \{X_n, U_n\})$ and is assumed to have continuous sample paths as a process of t. A suitable choice given the U-process structure of (3.10), which ensures that the variability of S_n is approximately the same over different t, is $c_n(t) = \widehat{\sigma}_{n,x}(t)/\sqrt{n}$ so that S_n becomes:

[9] However, as pointed out by Ghosal, Sen, and van der Vaart (2000), other functionals might be chosen depending on the specific interest of the researcher.

$$S_n = \sup_{t \in \mathcal{T}} \left\{ \frac{\sqrt{n}\widehat{U}_{n,x}(t)}{\widehat{\sigma}_{n,x}(t)} \right\} \tag{3.11}$$

where

$$\begin{aligned}\widehat{\sigma}^2_{n,x}(t) = \frac{1}{n(n-1)(n-2)} \sum_{1 \le i,j,k \le n, i \ne j \ne k} \Big(\widehat{\mu}(x,U_j) - \widehat{\mu}(x,U_i)\Big)\Big(\widehat{\mu}(x,U_k) - \widehat{\mu}(x,U_i)\Big) \\ \times \operatorname{sign}\Big(U_i - U_j\Big) \operatorname{sign}\Big(U_i - U_k\Big) \frac{1}{h_n^4} K_h(U_j - t) K_h(U_k - t) K_h^2(U_i - t)\end{aligned} \tag{3.12}$$

is the estimated U-process for:

$$\begin{aligned}\sigma^2_{n,x}(t) &= \int \left(\int \Big(\mu(x,\omega) - \mu(x,U)\Big) \operatorname{sign}\Big(U - \omega\Big) \frac{1}{h_n} K_h(\omega - t) \right)^2 dF(\omega) \frac{1}{h_n^2} K_h^2(U-t) dF(U) \\ &= \int \Bigg(\int \int \Big(\mu(x,\omega_1) - \mu(x,U)\Big)\Big(\mu(x,\omega_2) - \mu(x,U)\Big) \operatorname{sign}\Big(U - \omega_1\Big) \\ &\quad \times \operatorname{sign}\Big(U - \omega_2\Big) \frac{1}{h_n^2} K_h(\omega_1 - t) K_h(\omega_2 - t) dF(\omega_1) dF(\omega_2) \Bigg) \frac{1}{h_n^2} K_h^2(U-t) dF(U)\end{aligned} \tag{3.13}$$

with $F(\cdot)$ denoting a distribution function that corresponds to the marginal density function $f(\cdot)$. The respective test is given by:

$$\text{Reject } H_0 \text{ at level } \alpha \text{ if} \quad S_n > \tau_{n,\alpha}$$

where $\lim_{n \longrightarrow \infty} \mathbb{P}\{S_n > \tau_{n,\alpha}\} = \alpha$. Thus, to approximate the critical values, the limiting distribution of S_n is required. The first step towards this point is to show that (3.11) can be approximated by a stationary Gaussian process with continuous sample paths, which will be achieved in Theorem 6 of the next section.

3.3 Large Sample Theory

Before assumptions and asymptotic theory are outlined, the following terms need to be defined: let the projection of the degenerate U-process in (3.10) for $m(x, \cdot) = 0$

be denoted as (Hoeffding, 1948):

$$\widehat{U}^p_{n,x}(t) = \frac{2}{n}\sum_{i=1}^{n}\int\Big(\widehat{\mu}(x,\omega) - \widehat{\mu}(x,U_i)\Big)\,\mathrm{sign}\Big(U_i-\omega\Big)\frac{1}{h_n^2}K_h(\omega-t)dF(\omega)K_h(U_i-t) \tag{3.14}$$

where the projection term of higher order $O_p(h_n n^{-1})$ has been omitted. Moreover, define:

$$q(r) = \int\Big(\omega - r\Big)\,\mathrm{sign}\Big(r-\omega\Big)K(\omega)d\omega \tag{3.15}$$

and

$$\rho(s) = \frac{\int q(r)q(r-s)K(r)K(r-s)dr}{\int q^2(r)K^2(r)dr} \tag{3.16}$$

as well as $\mathcal{T}_n = [0,(b-a)/h_n]$. The following assumptions are made:

A1 Let $\{W_i, X_i^T, U_i, Z_i\}_{i=1}^n$ be i.i.d. data with finite second moments.

A2 $\mathcal{W} = \mathcal{U}\times\mathcal{X}\times\mathcal{V}$ is a non-empty set, where $\mathcal{U}$, $\mathcal{X}$, and $\mathcal{V}$ are subsets in the interior of the marginal support of X and U and V. The marginal distribution function of V is continuously differentiable on $\mathcal{V}$. The elements x in the support of X can be partitioned into subvectors of discrete $x^{(d)}$ with a finite number of points and continuous $x^{(c)}$ components of dimension D_c. Let $\mathcal{X}^{(d)}$ and $\mathcal{X}^{(c)}$ be the corresponding discrete and continuous parts of $\mathcal{X}\subset\mathcal{W}$. Assume that the conditional density (given $x^{(d)}\in\mathcal{X}^{(d)}$) on $\mathcal{W}$ is continuously differentiable and strictly bounded away from zero.

A3 $K(\cdot)$ is a bounded and symmetric second order kernel function with compact support $[-1,1]$. It is twice continuously differentiable.

A4 The function $\mu(\cdot,\cdot) = \mathbb{E}\Big[\mu(\cdot,\cdot,V_i)\Big]$ as defined in (3.7) satisfies, for every $u_1,u_2\in\mathcal{U}$ and $x\in\mathcal{X}$, $\mu(\cdot,\cdot) = \mathbb{E}\Big[\mu(\cdot,\cdot,V_i)^2\Big] < \infty$ and the following Lipschitz condition:

$$\Big|\mu(x,u_1) - \mu(x,u_2)\Big| \leq C\Big|u_1 - u_2\Big|$$

where C is a generic finite constant. Moreover, let $w = \{x^{(c)},u,v\}^T$ denote a vector of dimension D_c+2 and $\|\cdot\|$ the Euclidean norm. Assume that for every $x^{(d)}\in\mathcal{X}^{(d)}$, $\mu(\cdot,\cdot,\cdot)$ is twice continuously differentiable with derivatives satisfying the Lipschitz conditions:

$$\left\|\nabla_w \mu(x^{(d)}, x_1^{(c)}, u_1, v_1) - \nabla_w \mu(x^{(d)}, x_2^{(c)}, u_2, v_2)\right\| \leq C_D \left\|w_1 - w_2\right\|$$
$$\left\|\nabla_{ww'} \mu(x^{(d)}, x_1^{(c)}, u_1, v_1) - \nabla_{ww'} \mu(x^{(d)}, x_2^{(c)}, u_2, v_2)\right\| \leq C_{DD} \left\|w_1 - w_2\right\|$$

for every $w_1, w_2 \in \mathcal{X} \times \mathcal{U} \times \mathcal{V}$, where C_D, C_{DD} are again generic finite constants.

A5 For every $\{x, U, t : x \in \mathcal{X}, U \in \mathcal{U}, t \in \mathcal{T}\}$ and bandwidth $h_n \longrightarrow 0$, assume that:

$$\int \Big(\widehat{\mu}(x,\omega) - \mu(x,\omega)\Big) \operatorname{sign}\Big(U - \omega\Big) \frac{1}{h_n} K_h(\omega - t) dF(\omega) \leq C \sup_{x \in \mathcal{X}} \Big|\widehat{\mu}(x,U) - \mu(x,U)\Big|$$

where $\mu(\cdot,\cdot)$ and $\widehat{\mu}(\cdot,\cdot)$ are defined in (3.7) and (3.8), respectively, and C is a generic finite constant.

Assumption A2 allows in principle for continuous as well as discrete conditioning variables x in $\mu(\cdot,\cdot)$. In practice, the former will obviously depend on the nature of the data set requiring a rich enough data source if continuous variables are part of the conditioning vector (for instance, if age was treated as a continuous random variable and sufficient observations existed, one might be testing the monontonicity of the reservation wage function for different age levels). The underlying assumption with discrete x is that, with slight abuse of notation, $n \longrightarrow \infty$ also holds conditional on a specific value of x. A2 also incorporates the identification conditions from the previous section (see Newey, Powell, and Vella (1999) for details), while the absolutely continous distribution of V requires U to be continuous. Condition A3 is satisfied by many commonly used kernel functions such as the Epanechnikov kernel $K(v) = 0.75(1 - v^2)\,\mathbb{I}[|v| \leq 1]$ or the biweight kernel $K(v) = (15/16)(1 - v^2)^2\,\mathbb{I}[|v| \leq 1]$. The Lipschitz continuity assumptions on $\mu(\cdot,\cdot)$ and the derivatives of $\mu(\cdot,\cdot,\cdot)$ in A4 comlement the assumptions made about the regression function $\widetilde{m}(\cdot)$ and allow in principle a generalization of the setup to non-separable models $W_i = m(X_i, U_i, \epsilon_i)$. However, such a generalization requires also a 'tightening' of the identification conditions to full independence of the instrument Z_i and the unobservables (ϵ_i, V_i) as discussed in the previous section. The following theorem establishes consistency of the rescaled test statistic.

Theorem 6. *Assume that A1 to A6 hold. Let the bandwidth sequence satisfy* $h_n\sqrt{\log(n)} \longrightarrow 0$*,* $nh_n(\log(n))^{-2} \longrightarrow \infty$*, and* $nh_n^3 \longrightarrow \infty$*. Then there exists a sequence of stationary Gaussian processes* $\{\xi_n(s) : s \in \mathcal{T}_n\}$ *with continuous sample paths s.t.:*

$$\mathbb{E}\Big[\xi_n(s)\Big] = 0, \quad \mathbb{E}\Big[\xi_n(s_1)\xi_n(s_2)\Big] = \rho(s_1 - s_2), \quad s_1, s_2, s \in \mathcal{T}_n$$

where $\rho(\cdot)$ *was defined in (3.16) and*

$$\begin{aligned}&\sup_{t\in\mathcal{T}}\left|\frac{\sqrt{n}\widehat{U}_{n,x}(t)}{\widehat{\sigma}_{n,x}(t)} - \xi_n(h_n^{-1}(t-a))\right| \\ =&O_p\Big(h_n\sqrt{\log(n)} + h_n^{\frac{1}{2}} + n^{-\frac{1}{2}}h_n^{-\frac{1}{2}}\log(n)\Big) \\ =&o_p(1)\end{aligned}$$

The proof is carried out in several steps, which follow closely the proof of Theorem 3.1 in Ghosal, Sen, and van der Vaart (2000): Lemma A2 establishes the order of the error when approximating the U-statistic $U_{n,x}(t)$ by its projection $U_{n,x}^p(t)$. Lemma A3 gives an approximation of the empirical process $\sqrt{n}U_{n,x}^p(t)$ by a Gaussian process $G_n(t)$. Lemma A4 shows that $\widehat{\sigma}_{n,x}(t)$ converges uniformly to $\sigma_{n,x}(t)$. Finally, in Lemma A5 it is shown that the scaled Gaussian process $G_n(t)/\sigma_{n,x}(t)$ can be approximated by a stationary Gaussian process $\xi_n(t)$. The key difference to Theorem 3.1 consists in the fact that $\widehat{\mu}(\cdot,\cdot)$ and $\widehat{\sigma}_{n,x}(t)$ are estimated, which needs to be accounted for. This is carried out in Lemma A1 and A6: Lemma A1 establishes a parametric convergence rate for the averaged nonparametric estimator $\widehat{\mu}(\cdot,\cdot)$. Similar to Gutknecht (2011), this is accomplished by using the fact that $(\widehat{V}_i - V_i) = (\widehat{g}(Z_i) - g(Z_i))$ and that the first stage estimator averages over the $\widehat{V}_i$. The latter allows for an approximation of the statistic by a second order U-process, which in turn can be approximated by a suitable project. Subsequently, a standard Lindberg-Levy Central Limit Theorem can be applied to prove convergence in distribution. Finally, Lemma A6 then shows that, given the bandwidth conditions, a similar argument to the proof of Theorem A.1 in Lee, Linton, and Whang (2009) can be used to show that $\widehat{U}_{n,x}(t) - U^p_{n,x}(t) = O_p(n^{-\frac{1}{2}})$ uniformly over $x \in \mathcal{X}$ and $t \in \mathcal{T}$. That is, the asymptotic distribution of S_n can be treated as if $\mu(\cdot,\cdot)$ was observed. The same argument applies for the estimated $\widehat{\sigma}_{n,x}(t)$.

The next step is to determine the asymptotic distribution of the test statistic. From Theorem 6 above, it is clear that $S_n = \sup_{s\in\mathcal{T}_n} \xi_n(s) + O_p(\delta_n)$, where $\delta_n = h_n\sqrt{\log(n)} + h_n^{\frac{1}{2}} + n^{-\frac{1}{2}}h_n^{-\frac{1}{2}}\log(n)$. Therefore, for some positive a_n and some real number b_n, if:

$$a_n\Big(\sup_{s\in\mathcal{T}_n} \xi_n(s) - b_n\Big) \xrightarrow{d} L$$

holds for some random variable L, then it also holds that:

$$a_n\Big(S_n - b_n\Big) \xrightarrow{d} L$$

provided $a_n\delta_n = o(1)$. As mentioned in Ghosal, Sen, and van der Vaart (2000), since the interest lies in distributions only and the covariance of $\rho(\cdot)$ is free from n, one may assume that all the Gaussian processes are the same with:

$$\mathbb{E}\Big[\xi(s)\Big] = 0, \quad \mathbb{E}\Big[\xi(s_1)\xi(s_2)\Big] = \rho(s_1 - s_2), \quad s_1, s_2, s \in \mathcal{T}_n$$

The following theorem establishes the limiting distribution of this Gaussian process.

Theorem 7. *Let assumptions A1 to A5 hold. The bandwidth sequence satisfies* $h_n\log(n) \longrightarrow 0$, $nh_n(\log(n))^{-3} \longrightarrow \infty$, *and* $nh_n^3(\log(n))^{-1} \longrightarrow \infty$. *Then, for any x:*

$$\lim_{n\to\infty} \mathbb{P}\left(a_n(\sup_{t\in\mathcal{T}_n} \xi_t - b_n) \geq x \right) = exp(-e^{-x}) \equiv F_\infty(x)$$

where $a_n = \sqrt{2\log((b-a)/h_n)}$ *and*

$$b_n = \sqrt{2\log((b-a)/h_n)} + \frac{\log\left(\frac{\lambda^{\frac{1}{2}}}{2\pi}\right)}{\sqrt{2\log((b-a)/h_n)}}$$

with

$$\lambda = -\frac{\int q(v)q^{''}(v)K^2(v)dv + 2\int q(v)q^{'}(v)K(v)K^{'}(v)dv + \int q(v)^2K(v)K^{''}(v)dv}{\int q(v)^2K^2(v)dv}$$

and $q^{'}(\cdot), K^{'}(\cdot)$ *and* $q^{''}(\cdot), K^{''}(\cdot)$ *denote the first and second derivative of* $q(\cdot)$ *and the kernel function, respectively.*

Thus, as in Theorem 4.2 of Ghosal, Sen, and van der Vaart (2000), the asymptotic distribution of S_n follows straightforwardly from the above theorem:

$$\lim_{n\to\infty} \mathbb{P}\Bigg(a_n(S_n - b_n) \geq x \Bigg) = exp(-e^{-x}) \equiv F_\infty(x)$$

and one can construct a test with asymptotic level α:

$$\text{Reject } H_0 \text{ if } \quad F_\infty(a_n(S_n - b_n)) \geq 1 - \alpha \tag{3.17}$$

Notice that, in order to ensure $a_n\delta_n \longrightarrow 0$, the restrictions imposed on the bandwidth sequence are slightly stronger than in Theorem 6: bandwidth sequences such as $h_n = 1/\log(n)^\gamma$, $\gamma > 1$, or $h_n = 1/n^\eta$, $\eta < (1/3)$, satisfy the above requirements and provide a broad range of possible bandwidths for which the test has asymptotic level α. To compute the above statistic, one needs to calculate a_n and b_n, which depend on h_n and λ. Since $K(\cdot)$ is supported on $[-1, 1]$, the integrals can be computed analytically. The biweight kernel for instance yields $\lambda \approx 3.082$, while for the Epanechnikov kernel one obtains $\lambda \approx 4.493$.

Next, the consistency of the test against general alternatives is examined, which leads to the following theorem:

Theorem 8. *Assume that* $nh_n^3/\log(n) \longrightarrow \infty$. *Then, for a given* $x \in \mathcal{X}$, *if* $\nabla_U m(x,t) < 0$ *for some* $t \in [a,b]$, *the test in (3.17) is consistent at any level* α.

The theorem imposes a further restriction on the bandwidth sequence. This is because, under violation of the null for some $t \in [a,b]$, $h_nU_{n,x}(t)$ for that t can be shown to converge to a positive limit in probability. Since S_n is of order $O_p(nh_n^3)$, this exceeds the order of b_n only if the bandwidth sequence satisfies $nh_n^3/\log(n) \longrightarrow \infty$.

3.4 Extension

A drawback of the control function approach is that suitable instrumental variables are required that satisfy appropriate relevance and exogeneity conditions. In the context of the reservation wage application, Addison, Centeno, and Portugal (2010) suggest the concept of 'unobservable instruments' as an alternative to the former using completed duration as an exogenous perturbation of elapsed duration to infer about the effect of elapsed unemployment duration on reservation wages. The underlying rationale of completed duration as 'unobservable instrument' will be explained by a job search model outlined in the next section. To formalize the econometric concept introduced by Matzkin (2004), assume that U_i is an exogenous perturbation of another continuous random variable T_i (completed duration). That is, $U_i = s(T_i, \zeta_i)$, where ζ_i is an unobservable that is assumed to be independent of ϵ_i from (3.1) and $s(\cdot,\cdot)$ is some unknown function. Further assume that the error term ϵ_i can be characterized by the following additive reduced form equation:

$$\epsilon_i = r(T_i) + \eta_i \tag{3.18}$$

where η_i is an unobservable variable that is assumed to be independent of the observable T_i. η_i in the above equation must not to be confounded with the control function from section 3.2. Notice also that T_i is not an instrumental variable in the traditional sense as it is correlated with the unobservable ϵ_i by construction. Put differently, endogeneity in this framework can be addressed because it is caused by a 'fixed effect' (e.g. unobserved heterogeneity) that is present in U_i as well as T_i. Thus, controlling for observed T_i implies controlling for the unobserved effect.[10]

[10]That is, the roles of the observable T_i and the unobservable η_i have been exchanged.

Inserting (3.18) into the regression equation yields:

$$W_i = \widetilde{m}\Big(X_i, U_i\Big) + r\Big(T_i\Big) + \eta_i \tag{3.19}$$

Equation (3.19) represents an additive nonparametric regression model with the unobservable error term η_i that is assumed to be independent of X_i, U_i, and T_i. Given some additional regularity and identification conditions, one may, for a specific $x \in \mathcal{X}$ of interest, recover $m(x,\cdot) = -\widetilde{m}(x,\cdot)$ and $r(\cdot)$ using standard backfitting methods as proposed by e.g. Mammen, Linton, and Nielsen (1999). In order to apply this procedure, it is assumed that X only contains discrete elements. With slight abuse of notation, $n \longrightarrow \infty$ will hence represent the sample size for a specific value x in the following. Also, for identification purposes it is assumed that $m(x,.)$ and $r(\cdot)$ can be normalized to $\mathbb{E}\Big[m(x,U_i)\Big] = 0$ and $\mathbb{E}\Big[r(T_i)\Big] = 0$ for every $x \in \mathcal{X}$. Once $\widehat{m}(x,\cdot)$ is obtained, this function can be plugged into the second order U-process in (3.10) in lieu of $\widehat{\mu}(x,\cdot)$ to compute the test statistic. That is, using the smooth backfitting procedure of Mammen, Linton, and Nielsen (1999) to estimate $m(x,\cdot)$ and $r(\cdot)$ from (3.19) for a specific x (see their paper for details of the estimator), one may construct the following modified test statistic:

$$S_n^* = \sup_{t \in \mathcal{T}} \left\{ \frac{\sqrt{n}\widehat{U}^*_{n,x}(t)}{\widehat{\sigma}^*_{n,x}(t)} \right\} \tag{3.20}$$

where $\widehat{U}^*_{n,x}(t)$ is defined as:

$$\widehat{U}^*_{n,x}(t) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \Big(\widehat{m}(x,U_j) - \widehat{m}(x,U_i)\Big) \operatorname{sign}\Big(U_i - U_j\Big) \frac{1}{h_n^2} K_h(U_i - t) K_h(U_j - t)$$

and $\widehat{m}(x,\cdot)$ is the backfitting estimator of $m(x,\cdot)$. $\widehat{\sigma}^*_{n,x}(t)$ is the equivalent to (3.12) using $\widehat{m}(x,\cdot)$ instead of $\widehat{\mu}(x,\cdot)$. The following regularity conditions from Mammen, Linton, and Nielsen (1999) are imposed:

A2* For every $x \in \mathcal{X}$, assume that U and T have compact support $[0,1] \times [0,1]$ and that the joint density function is continuously differentiable in its arguments and strictly bounded away from zero everywhere on $[0,1] \times [0,1]$.

A4* For some $\theta > \frac{5}{2}$, assume that $\mathbb{E}\Big[\Big|W\Big|^{\theta}\Big] < \infty$. Moreover, for every $x \in \mathcal{X}$, the functions $m(x,\cdot)$ and $r(\cdot)$ are twice continuously differentiable in their (second) argument.

Since U_i and T_i are both supported on $[0,\infty)$, condition A2* implies that trimming (as for A2 in the previous section) has to be applied to restrict the support to a compact subset with strictly positive density. A suitable affine transformation of the data will then ensure that the normalization to $[0,1]$ in A2* is satisfied. Moreover, let the assumption below substitute condition A5 from section 3.3:

A5* For every $\{x, U, t : x \in \mathcal{X}, U \in \mathcal{U}, t \in \mathcal{T}\}$ and bandwidth $h_n \longrightarrow 0$, assume that:

$$\int \Big(\widehat{m}(x,\omega) - m(x,\omega)\Big)\,\mathrm{sign}\Big(U - \omega\Big)\frac{1}{h_n}K_h(\omega - t)dF(\omega) \leq C \sup_{x\in\mathcal{X}}\Big|\widehat{m}(x,U) - m(x,U)\Big|$$

where $m(\cdot,\cdot)$ is defined in (3.7) and C is a generic finite constant.

The following theorem can be established in analogy to Theorem 6:

Theorem 9. *Assume that A1, A3 as well as A2*, A4*, and A5* hold and that* $\mathcal{T} \subset (0,1)$. *Let the bandwidth sequence satisfy* $nh_n^3 \longrightarrow \infty$, $nh_n(\log(n))^{-2} \longrightarrow \infty$, *and* $nh_n^5 \longrightarrow 0$. *Then there exists a sequence of stationary Gaussian processes* $\{\xi_n(s) : s \in \mathcal{T}_n\}$ *with continuous sample paths s.t.:*

$$\mathbb{E}\Big[\xi_n(s)\Big] = 0, \quad \mathbb{E}\Big[\xi_n(s_1)\xi_n(s_2)\Big] = \rho(s_1 - s_2), \quad s_1, s_2, s \in \mathcal{T}_n$$

where $\rho(\cdot)$ *was defined in (3.16) and*

$$\begin{aligned}&\sup_{t\in\mathcal{T}}\left|\frac{\sqrt{n}U_{n,x}(t)}{\widehat{\sigma}_{n,x}(t)} - \xi_n(h_n^{-1}(t-a))\right| \\ =&O_p\Big(h_n\sqrt{\log(n)} + n^{\frac{1}{2}}h_n^{\frac{5}{2}} + n^{-\frac{1}{2}}h_n^{-\frac{1}{2}}\log(n)\Big) \\ =&o_p(1)\end{aligned}$$

The subsequent theorems follow in accordance with Theorem 7 and 8 from section 3.3 and have been omitted for brevity. Hence, this section has presented a viable alternative to the use of control functions in case where suitable instrumental variables do not exist, but instead unobserved exogenous variation is available.

3.5 Monte Carlo Simulation

In order to research the performance of the test in small samples, a Monte Carlo simulation is carried out. The results are displayed in Table 3.1 of Appendix A3.2. Firstly, the behaviour of the test in small samples is examined when the null hypothesis is true. The underlying model for this case is chosen to be the following monotonically increasing function:

$$W_i = 0.1 \cdot U_i + \epsilon_i$$

where the regressor U_i is constructed as $U_i = Z_i + V_i$: the instrument Z_i and the control function V_i are drawn from uniform distributions supported on $[.25, .75]$ and $[-.25, .25]$, respectively. Thus, U_i is supported on the compact interval $[0, 1]$. The unobservable ϵ_i is given by: $\epsilon_i = V_i + 0.1 \cdot \varpi_i$ with $\varpi_i \sim N(0, 0.1^2)$.

The kernel function is chosen to be the Epanechnikov kernel $K(v) = 0.75(1 - v^2)\,\mathbb{I}[|v| \leq 1]$ and the first stage functions $g(\cdot)$ in (3.2) and $\mu(\cdot)$ in (3.7) are estimated using the Nadaraya Watson estimator as explained in section 3.2. The bandwidth for these estimators are determined using the rule of thumb for nonparametric density estimators, i.e. $C \cdot sd(\cdot) \cdot n^{-\frac{1}{5}}$ with $sd(\cdot)$ the standard deviation and $C = 2.34$ a constant for the Epanechnikov kernel ($C = 2.78$ for the Biweight kernel). Finally, the test statistic is constructed as described in (3.11) with the interval $\mathcal{T}$ chosen to be $\mathcal{T} = \{0.05, 0.1, \ldots, 0.9, 0.95\}$. There are three different bandwidth parameters used for the construction of the test statistic at the final stage, each of which satisfies the requirements of Theorem 7. The simulations use sample sizes of $n = 100, 200, 300$ and $1,500$ replications are conducted for each simulation.

Reverting to Table 3.1 one can observe that, while a reasonable approximation to the nominal size of 5% is obtained for $h_n = 1 \cdot n^{-\frac{1}{5}}$, the proportion of rejections appears to be rather sensitive to modifications of the bandwidth for this chosen specification: surprisingly, the share increases with growing sample size (and thus decreasing bandwidth) to roughly 16% for $h_n = 0.8 \cdot n^{-\frac{1}{5}}$ and decreases to 0.002% for $h = 1.2 \cdot n^{-\frac{1}{5}}$. This is in contrast to a stabilization of the size at around 3% for $h_n = 1 \cdot n^{-\frac{1}{5}}$. In order to better understand this sensitive behaviour, further simulations are to be carried out in the future.

Next, the behaviour of the test is examined when the null hypothesis is false. The model is altered to:

$$W_i = U_i(1 - U_i) + \epsilon_i$$

with the variables themselves being generated as before. In this case, reasonable rejection levels are achieved across all bandwidth specifications: already at $n = 100$, rejection shares range from 94 to 98%. At $n = 200$, these proportions have reached or are very close to one, while for $n = 300$ the nominal level is reached throughout.

Thus, despite a somewhat sensitive behaviour of the test under the null, simulation results presented in Appendix A2 do overall provide a fairly positive and encouraging picture of the small sample propteries of the test. Still, other specifications and different sample sizes are yet to be examined in order to further understand its performance under different model specifications. Moreover, using an asymptotic expansion similar to the one in Lee, Linton, and Whang (2009) to construct the test statistic might substantially improve the results as in their paper. The parameters for this asymptotic expansion have yet to be derived.

3.6 An Application to Reservation Wages

Reservation wages have been the focal point of labour economists for many decades since they play a key role in modern job search theory. While early partial equilibrium job search models typically assumed constant reservation wages, later studies mostly relaxed this assumption and allowed for flexible reservation wages that could change with elapsed unemployment duration (Kiefer and Neumann, 1979). In fact, in instances where the hypothesis of constant reservation wages has been tested empirically, it has typicallly been rejected (Kiefer and Neumann, 1979; Brown and Taylor, 2009; Addison, Machado, and Portugal, 2011). That is, using linear regression techniques, most studies established a significantly negative regression coefficient for elapsed duration hinting at declining reservation wages over time. However, despite its potential policy implications, no information has yet been provided about whether this decline is montonic (throughout an unemployment spell) and whether it holds across different subgroups of the population (see introduction).

In the following, it will be examined under which conditions reservation wages decline monotonically using a standard job search model based on the one of van den Berg (1990). Moreover, the model developed in this paper will accomodate unobserved heterogeneity across agents and thus provide a theoretical underpinning of completed unemployment duration as an exogenous perturbation of elapsed duration.

Let U denote continuous calendar time, which starts at the moment an individual becomes unemployed and thus characterizes elapsed unemployment duration. The underlying hazard rate for such an elapsed duration $U(Z) \in [0, \infty)$ is given by $\theta(U(Z), X, V) = \Psi(U(Z), X) \cdot V$. The dependence of elapsed unemployment duration $U(\cdot)$ on the vector of instruments $Z = \{Z_1, X\}$ is assumed to ensure the validity of the latter. V is a random variable denoting unobserved heterogeneity, which is assumed to be independent of elapsed duration $U(Z)$. Thus, in line with standard mixed proportional hazard models, unobserved heterogeneity V is time invariant.[11] Independence and constancy are certainly strong and rather unrealistic assumptions. However, both are owed to the use of completed unemployment as an exogenous variation for elapsed unemployment duration, which relies on the existence of a fixed effect as the only source of endogeneity. As outlined in the introduction, the paper uses different methods to verify the robustness of results obtained from regressions with completed unemployment duration as an exogenous perturbation of elapsed unemployment duration. Hence, severe violations of the above assumption are likely to lead to differing results. The conditional distribution function of elapsed durations

[11]Notice that, despite assuming that unobserved heterogeneity enters multiplicatively into the hazard, other observed covariates must not necessarily do so. This provides a certain degree of flexibility to the approach as no proportional hazards in the observed covariates are imposed.

$U(Z)$ is given by $F(U(Z)|X,V) = \{1 - \exp(-\int_0^{U(Z)} \Psi(t,X)dt)\exp(-V)\}$. Notice that while the model imposes proportionality in the unobserved heterogeneity term V (thus allowing for fixed effects), no such assumption is imposed on X. The latter is also reflected by the non-separability of the reduced form function $m(\cdot,\cdot)$ in its arguments. The unemployed agent receives job offers that arrive with a fixed wage ω attached, which represent random draws from a distribution with known distribution function $F_\omega(\cdot;X)$. Notice that the wage offer distribution function depends on observed covariates X and not on Z_1, which is crucial for our instruments Z_1 to be valid.[12] The discount factor is given by ρ and the rate at which he receives these offers is $\lambda(U(Z),X,V)$. Every time the agent receives such an offer, he may decide whether to accept or reject it: if he accepts, the job will be held forever, while a rejection of the offer implies that he may not recall this job at a later stage anymore.[13] In this stylized version of the model, there are no costs to search and agents receive benefits $b(U(Z))$ over their course of unemployment. Individuals are assumed to maximize the expected present value of income (over an infinite horizon) and they are able to anticipate changes in the exogenous $b(U(Z))$ and $\lambda(U(Z),\cdot,\cdot)$. While the assumption of an infinite decision horizon is clearly unrealistic (but typically adopted by the literature to simplify matters and to gain better insight into the essentials of the problem), the anticipation condition appears fairly plausible in instances where individuals have for instance upfront information about future reductions in benefit payments (as in the case of contribution based jobseeker's allowance in the UK).

The following assumptions are sufficient for a strictly monotonically declining reservation wage:

J1 The discount factor satisfies $0 < \rho < \infty$, while $0 < b(U(Z)) \leq \Psi_b < \infty$ for all $U(Z) \in [0,\infty)$ and some fixed Ψ_b. Additionally, assume that for every X and V it holds that $0 < \lambda(U(Z),X,V) \leq \Psi_\lambda < \infty$ for some fixed Ψ_λ and all $U(Z) \in [0,\infty)$. Let $F_\omega(\cdot;X)$ be continuous and strictly monotonic in ω with

[12] $F_\omega(\cdot;X)$ denotes the wage offer distribution rather than the conditional distribution function of elapsed unemployment durations. A possible element of X might be gender if one expects wage offer distributions to differ between men and women.

[13] See van den Berg (1990) for a discussion of these assumptions.

$\lim_{\omega \to 0} F_\omega(\omega; X) = 0$ and $\lim_{\omega \to \infty} F_\omega(\omega; X) = 1$. Moreover, assume that $F_\omega(\cdot; X)$ has finite first moment for every X.

J2 There exists some finite point $\overline{U} \in [0, \infty)$ such that $b(U(Z)) = b$ and $\lambda(U(Z), X, V) = \lambda$ are constant for every $U(Z) \in [\overline{U}, \infty)$. For every $U(Z) \in [0, \overline{U})$, assume that $\partial\lambda(U(Z), X, V)/\partial U(Z)$ exists and is negative. In addition, there exists some point $U_b \in [0, \overline{U})$ s.t. $b(U(Z)) = b_1$ for $U \in [0, U_b)$ and $b(U(Z)) = b_2$ for $U(Z) \in [U_b, \infty)$, respectively. Without loss of generality it is imposed that $b_1 > b_2$.[14]

This is the simplified setup of van den Berg (1990). The constancy assumption in J2 after $\overline{U}$ together with J1 imply that the model has a unique solution. Moreover, the assumption of $b(U(Z))$ being a declining step function (which captures the effect of a possible benefit reduction after 182 days) allows us to split the time axis $[0, \overline{U})$ into two intervals on which $b(U(Z))$ is constant. Take for instance the first interval $[0, U_b)$. The reservation wage function for every $U(Z) \in [0, U_b)$ is characterized by the following differential equation:

$$\begin{aligned}\frac{\partial \omega^*(U(Z), X, V)}{\partial U(Z)} &= \rho\omega^*(U(Z), X, V) - \rho b_1 - \lambda(U(Z), X, V) \\ &\qquad \times \int_{\omega^*(U(Z),X,V)}^{\infty} (\omega - \omega^*(U(Z), X, V)) dF_\omega(\omega; X)\end{aligned}$$

Notice also that the hazard rate $\theta(U(Z), X, V)$ can be rewritten as $\theta(U(Z), Z, V) = \lambda(U(Z), X, V)(1 - F_\omega(\omega^*(U(Z), X, V); X)$. By Theorem 2 of van den Berg (1990), the reservation wage function $\omega^*(U(Z))$, which is continuous and differentiable in $U(Z)$ for all $U(Z) \in [0, U_b)$ by J1 and J2, satisfies:

(i) $\omega_0^*(X, V) > \omega^*(U(Z), X, V)$ for every $U(Z) \in (0, U_b)$, where $\omega_0^*(\cdot, \cdot)$ denotes the solutions under the assumption of constant parameters from time 0 onwards.

(ii) For every $U(Z) \in (0, U_b)$, it holds that $\partial\omega^*(U(Z), X, V)/\partial U(Z) < 0$.

In other words, the reservation wage $\omega^*(U(Z), X, V)$ is monotonically declining for all $U(Z) \in (0, U_b)$ only if the function $b(U(Z))$ and $\lambda(U(Z), X, V)$ are strictly de-

[14]Imposing $b_2 > b_1$, albeit not a very realistic scenario, would simply alter the direction of inequalities in the following theorem.

creasing in $U(Z)$. Obviously, these assumptions are fairly restrictive and might sometimes be violated.

The remainder of this section will be dedicated to assessing the monotonicity of the reservation wage function using unemployment data from the BHPS, a nationally representative survey on individuals from more than 5,000 households in the UK. It contains detailed questions on the current labour market situation of adults in each household. Unemployment spells and labour market states are constructed using the case-by-case correction method of inconsistencies (Method C) developed by Paull (2002). The starting point of the sampling period is chosen to be October 1996, which conincides with the introduction of jobseeker's allowance in the UK. To examine the robustness of the empirical results, three different endpoints have been selected for the analysis, namely 31st of December 2002, 2005 and 2007. The first date has been selected due to major reforms of the British tax credit system in April 2003. It is conjectured that this legislation change could affect reservation wages since a provisional system (Working Family Tax Credits) leading up to this reform has recently been found to have impacted the latter (Brown and Taylor, 2009). Moreover, the end year 2007 has been chosen to avoid censoring of unemployment spells at the end of the observation period.

The hourly reservation wage is constructed combining answers from the two questions "What is the lowest weekly take home pay you would consider accepting for a job?" and "About how many hours in a week would you expect to have to work for that pay?" that non-employed individuals are asked during the interview. The sample then includes all individuals of working age (16-65) who indicate such an hourly wage and who satisfy the rationality condition, which requires a reservation wage below the reported expected wage. Notice that even individuals who indicated to be economically inactive are included in the sample if they have a valid reservation and expected wage. The decision to incorporate these observations is based on recent advances in labour market research questioning the clear-cut distinction between inactive and labour-seeking agents and instead interpreting the indication of a reservation wage as a signal for labour market attachment (see Brown and Taylor (2009) and references therein). Finally, to test robustness, observations below the nationally binding minimum wage (which became applicable after 1999) have been dropped in all three but a basic 1996-2002 sample specification.[15]

For all empirical specifications, the sample is split into male and female subsamples. The continuous instrumental variables are unemployment benefits and other benefit income. In order to incorporate also discrete variables as instruments, principal component analysis was employed. The latter is a common statistical technique to aggregate multivariate data into a (smaller) set of linearly uncorrelated variables, the so called principal components. Despite the fact that the underlying asymptotic properties have been derived under the assumption of normally and continuously distributed variables, Kolenikov and Angeles (2009) point out in a recent simulation study on socioeconomic status measurements that the bias for using discrete variables in principal component analysis appears to be rather small if it contains mostly categorical data with several categories that is not transformed into binary indicators. Thus, ordinal and count variables such as number of dependent children, age, and education have been included in the analysis alongside variables such as having a companion (being married or living as a couple), having a working spouse, and regional dummy variables (only number of dependent children, having a companion, and having a working spouse are considered to be part of Z_1). Only the first two principal components have been retained. Moreover, as pointed out previously, the maintained assumption is that all instrumental variables impact reservation wages exclusively through elapsed unemployment duration. This is formalized by writing $U(\cdot)$ as a function of the instrument vector Z.

The bandwidth sequence is determined using the leave-one-out cross-validation method. That is, the bandwidth is determined according to $h_n = C \cdot \min\{sd(\cdot); 0.8 \cdot iqr\} \cdot n^{-\frac{1}{5}}$, where iqr denotes the interquartile range and C is chosen through cross-validation from the grid $\{0.9, 0.925, \ldots, 1.175, 1.2\}$.[16] Turning to the summary statistics of Table 3.2 in Appendix A2, which displays key features of the hourly reservation wage as well as the elapsed unemployment distribution for the 1996 to 2002 sample, one can observe that, irrespective of an elimination of reservation wages below the mini-

[15]This only applies to spells recorded after 1999 when the minimum wage became applicable.
[16]The biweight kernel has been used in all specifications.

mum wage ("Min. Wage Correction"), the reservation wage distribution for females has a slightly lower mean and higher variance than the one for males. By contrast, in both specifications, female subsamples display a larger average elapsed duration (and greater variance). The share of multiple spells is 20.6% (19.58%) for males and 15.71% (15.48%) for females and thus fairly evenly distributed across gender. Turning to Table 3.3 displaying statistics for the 1996-2005 and the 1996-2007 samples, one observes that, as expected, the means of the hourly reservation wage increase as more recent years are included. Likewise, average elapsed durations fall, which is again not surprising given the positive developments in the British labour market during the early 2000s.

The plots in Figures 3.1, 3.2, 3.3, and 3.4 display $-\widehat{\mu}(x, \cdot)$ from equation (3.8) for the first 250 days of elapsed unemployment duration for the different specifications. Figures 3.1 and 3.2 concern the estimated hourly reservation wage functions for males ($x = 1$), while Figures 3.3 and 3.4 show the equivalent figures for females ($x = 0$). Firstly, notice that the patterns are fairly robust across samples for both men and women: for men (Figures 3.1 and 3.2), one observes a fairly steep decline during the first 50-75 days of elapsed duration, which is followed by an increase that eventually exceeds the starting value. The initial fall seems to be more pronounced when observations after 2002 are taken into account, too, while the rise after the turning point at around 75 days is more marked for the 1996-2002 sample. Moreover, in particular for the extended 1996-2005 and 1996-2007 samples, there appears to be a 'dent' in the reservation wage curve after around 175-200 days, which conincides with the regime change from contribution to income based jobseeker's allowance after six months for people with sufficient contributions.[17] For women, a similar albeit much less pronounced pattern can be observed. A moderate initial decline during the first 50 days of elapsed duration is followed by a marked increase outweighing by far the initial loss.

Putting the observations from the graphs to the (formal) test, Table 3.4 displays the results for the null of a decreasing reservation wage function across all the different

[17] 180 days also marks the starting point of the gateway period for unemployed qualifying for the 'New Deal' programme.

specifications. The compact interval T is chosen to be bounded by the 2% and 50% quantile for each sample and contains nineteen equally spaced points within those bounds.[18] Estimator and kernel function are as described in Section 3.5. An upfront log transformation of elapsed duration has been performed and the bandwidth has been determined by $1.25 \cdot n^{-\frac{1}{5}}$ in compliance with Theorem 6.[19] Examining the test results in Table 3.4, one can observe a clear-cut rejection of monotonically declining reservation wages even at a 1% significance level for all specifications. Furthermore, with the exception of the 1996-2005 sample, larger test statistics are obtained for females throughout reflecting their upwards sloping reservation wage curves.

In summary, this section has demonstrated that the behaviour of the reservation wage function is more complicated than typically assumed, which might not be captured by standard linear estimation techniques. For instance, linear two stage least squares regressions for the different samples (with the exogenous and instrumental variables from before) yielded estimated coefficients for elapsed unemployment duration ranging from -0.0028 to -0.0045 (with t-statistics from $|t| = 2.00$ to $|t| = 2.20$) for males and from -0.0017 to 0.0016 (with t-statistics from $|t| = 0.48$ to $|t| = 0.78$) for females. Thus, being able to better understand the impact of elapsed unemployment duration on (hourly) reservation wages remains an important task for future research and has, as outlined in the introduction, implications for the design of appropriate labour market policies.

3.7 Conclusion

This paper proposes a test for monotonicity of the regression function when the (continuous) regressor of interest is endogenous. It is argued that this kind of test is relevant for various empirical setups. As an important application, the paper studies the behaviour of hourly reservation wages as a function of elapsed unemployment duration in the UK using the British Household Panel Survey as data source. The relationship between reservation wage and elapsed unemployment duration is diffi-

[18] The 50% quantile roughly corresponds to 200 days of elapsed unemployment duration.

[19] Results are fairly robust to changes in the bandwidth sequence. In fact, modifications of the multiplier to 1 and 1.5 yielded very similar test outcomes (available upon request).

cult to measure due to the simultaneity of both variables. Using instruments such as the number of dependent children, having a working spouse, the logarithm of unemployment benefit or other benefit income to construct control functions, it is shown that the reservation wage function does typically not decline monotonically. Rather, various specifications seem to suggest that, after an initial decline, the function increases again after 75 to 100 days of unemployment and even exceeds the initial reservation wage level after around 200 days. This finding is robust across different specifications for both men and women, albeit much less pronounced for the latter. This has policy implications and could, upon further and more detailed investigation, give new insights into the behaviour of unemployed individuals.

Technically, the paper combines different conditional mean estimators with a test based on Ghosal, Sen, and van der Vaart (2000) and derives its asymptotic properties: the conditional mean estimator(s) can be constructed using either estimated control functions or variables that represent exogenous perturbations of the endogenous regressor (Matzkin, 2004). Either nonparametric estimator can then be plugged into a modified test statistic, which is chosen to be the supremum of a suitably rescaled second order U-process. The asymptotic distribution of the test statistic can be approximated by a stationary Gaussian process. A Monte Carlo simulation study evaluates the finite sample behaviour of the test. It is shown that, even in small samples, the test behaves well if the null hypothesis is violated. If the null hypothesis is satisfied, the test appears to be rather sensitive to the bandwidth choice. Finally, the application in section 3.6 demonstrates the test's applicability to actual data with the test results being throughout in line with observations from the graphs.

A straightforward extension of the current paper is to the case of heteroscedasticity, where the model in (3.1) is altered to $W_i = \widetilde{m}\Big(X_i, U_i\Big) + \sigma(X_i)\epsilon_i$. This specification is of interest for various empirical setups with monotonicity still being testable by the procedures developed in this paper since the key identification condition $\mathbb{E}[\sigma(X_i)\epsilon_i|X_i = x, Z_{1i} = z_1, V_i = v] = \sigma(x) \cdot \mathbb{E}[\epsilon_i|V_i = v]$ remains satisfied. Another important extension concerns the amplification of the current setup to nonparametric IV estimators at the first stage: given a suitable moment condition of the form

$\mathbb{E}\left[\epsilon_i \middle| Z_i\right] = 0$, estimators similar to the one suggested in Chen and Pouzo (2009) could be used to recover $m(\cdot,\cdot)$. This estimator of $m(\cdot,\cdot)$ could then be plugged into (3.10) as a substitute for $\widehat{\mu}(\cdot,\cdot)$. Finally, to improve the external validity of the results, potential selection bias problems should also be taken into account in the future: in a recent paper on reservation wages collected from the Italian Labour Force Survey, Sestito and Viviano (2011) point out two selection biases that affect the reservation wage distribution and that require either a restriction of the data set or an appropriate adjustment mechanism. To address these three aspects remains the key objective for future research.

A3 Appendix

A3.1 Proofs

Lemma A1. Under assumptions A1 to A4, it holds that:

$$\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\widehat{\mu}(x,u)-\mu(x,u)\Big| = O_p(n^{-\frac{1}{2}})$$

Lemma A2. Define:

$$\widehat{M}_{n,\mu}(x,t) = \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\Big(\mu(x,U_j)-\mu(x,U_i)\Big)\operatorname{sign}\Big(U_i-U_j\Big)\frac{1}{h_n^2}K_h(U_i-t)K_h(U_j-t) \quad \text{(A-3.1)}$$

so that $\widehat{M}_{n,\widehat{\mu}}(x,t)=\widehat{U}_{n,x}(t)$ from (3.10). Moreover, let:

$$\widehat{M}^p_{n,\mu}(x,t) = \frac{2}{n}\sum_{i=1}^{n}\int\Big(\mu(x,\omega)-\mu(x,U_i)\Big)\operatorname{sign}\Big(U_i-\omega\Big)\frac{1}{h_n^2}K_h(\omega-t)dF(\omega)K_h(U_i-t) \quad \text{(A-3.2)}$$

be the projection of $\widehat{M}_{n,\mu}(x,t)$ with $\widehat{M}^p_{n,\widehat{\mu}}(x,t)=\widehat{U}^p_{n,x}(t)$, where $\widehat{U}^p_{n,x}(t)$ is defined in (3.14). Then, under assumptions A1 to A4, it holds that:

$$\sup_{t\in T}\sup_{x\in\mathcal{X}}\Big|\widehat{M}_{n,\mu}(x,t)-\widehat{M}^p_{n,\mu}(x,t)\Big| = O_p(n^{-1}h_n^{-2})$$

Lemma A3. There exists a sequence of Gaussian processes $G_n(\cdot)$ indexed by t, with continuous sample paths and with:

$$\mathbb{E}\Big[G_n(t)\Big]=0,\quad \mathbb{E}\Big[G_n(t_1)G_n(t_2)\Big]=\mathbb{E}\Big[\Psi_{n,t_1}(U)\Psi_{n,t_2}(U)\Big],\quad t,t_1,t_2\in\mathcal{T}$$

where $\Psi_{n,t}(U)=\int\Big(\widetilde{\mu}(\omega)-\widetilde{\mu}(U)\Big)\operatorname{sign}\Big(U-\omega\Big)\frac{1}{h_n^2}K_h(\omega-t)dF(\omega)K_h(U-t)$ with $\widetilde{\mu}(U)=\mu(x,U)$, such that:

$$\sup_{t\in\mathcal{T}}\Big|\sqrt{n}\widehat{M}^p_{n,\mu}(x,t)-G_n(t)\Big| = O_p(n^{-\frac{1}{2}}h_n^{-1}\log(n))$$

Lemma A4. Under assumptions A1 to A4, the following holds:

(i) $\sup_{t\in\mathcal{T}}\Big|h_n\sigma^2_{n,x}(t)-\sigma^2_x(t)\Big| = o(1)$

(ii) $\lim\inf_{n\longrightarrow\infty} h_n \inf_{t\in\mathcal{T}}\sigma^2_x(t)>0$

(iii) $\sup_{t\in\mathcal{T}}\Big|\widehat{\sigma}^2_{n,x}(t)-\sigma^2_{n,x}(t)\Big| = O_p(n^{-\frac{1}{2}}h_n^{-2})$

where $\widehat{\sigma}^2_{n,x}(t)$ and $\sigma^2_{n,x}(t)$ are defined in (3.12) and (3.13), while $\sigma^2_x(t)=f^3(t)\nabla_U\mu(x,t)\int q(v)K^2(v)dv$.

Lemma A5. For the sequence of Gaussian processes $\{G_n(t):t\in\mathcal{T}\}$ obtained in Lemma A3, there corresponds a sequence of stationary Gaussian processes $\{\xi_n(s):s\in\mathcal{T}_n\}$ with continuous sample paths s.t.:

$$\mathbb{E}\Big[\xi_n(s)\Big]=0,\quad \mathbb{E}\Big[\xi_n(s_1)\xi_n(s_2\Big]=\rho(s_1-s_2),\quad s_1,s_2,s\in\mathcal{T}_n$$

where $\rho(\cdot)$ was defined in (3.16) and:

$$\sup_{t\in\mathcal{T}}\Big|\frac{G_n(t)}{\sigma_{n,x}(t)} - \xi_n(h_n^{-1}(t-a))\Big| = O_p(h_n\sqrt{\log(h_n^{-1})})$$

with $\sigma_{n,x}(t)$ as in Lemma A4.

Lemma A6. Under assumptions A1 to A5, it holds that:

$$\sup_{t\in\mathcal{T}}\sup_{x\in\mathcal{X}}\Big|\widehat{M}_{n,\widehat{\mu}}(x,t) - \widehat{M}^p_{n,\mu}(x,t)\Big| = O_p(n^{-\frac{1}{2}})$$

Proof of Theorem 6. The proof consists of several steps, which follow closely the proof of Theorem 3.1 in Ghosal, Sen, and van der Vaart (2000). In particular, Lemma A2-A5 replace Lemma 3.1-3.4 in Ghosal, Sen, and van der Vaart (2000). Lemma A1 on the other hand establishes that $\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\widehat{\mu}(x,u) - \mu(x,u)\Big| = O_p(n^{-\frac{1}{2}})$. Given that $nh_n^3 \longrightarrow \infty$ by Theorem 6, this is slower than the rate established in Lemma A2 and thus $\sup_{x\in\mathcal{X}}\sup_{t\in\mathcal{T}}\Big|\widehat{U}_{n,x}(t) - U^p_{n,x}(t)\Big| = O_p(n^{-\frac{1}{2}})$ is the overall rate obtained in Lemma A6. Notice however that an identical adjustment is not required for Lemma A4 since its rate is slower than the parametric one. ■

Proof of Theorem 8. The proof follows the same steps as the one of Theorem 5.1 in Ghosal, Sen, and van der Vaart (2000). As in their proof, if the null is violated for a specific $t \in \mathcal{T}$ ($\nabla_U m(x,t) < 0$), one can straightforwardly show that:

$$h_n U_{n,x}(t) \xrightarrow{p} -\nabla_U m(x,t)\int\int|u-\nu|K(u)K(\nu)f(t)^2 du d\nu$$

which is positive. Since also $h_n^{\frac{1}{2}}\widehat{\sigma}_{n,x}(t)$ tends to a positive limit and S_n can be shown to be of order $O_p(n^{-\frac{1}{2}}h_n^{\frac{3}{2}})$, the test statistic only exceeds the order of b_n if the bandwidth condition satisfies $nh_n^3/\log(n) \longrightarrow \infty$. ■

Proof of Theorem 9. The proof follows identical steps to the one of Theorem 6. However, Lemma A6 is replaced by the uniform convergence result in Mammen, Linton, and Nielsen (1999). That is, using A1 and A3 as well as A2*, A4*, and A5*, for a specific $x \in \mathcal{X}$, Theorem 4 of Mammen, Linton, and Nielsen (1999) yields:

$$\sup_{U\in(0,1)}\Big|\widehat{m}(x,U) - m(x,U)\Big| = o_p(h_n^2)$$

■

Proof of Lemma A1. In the following, write $\widehat{\mu}(x,u) = \frac{1}{n}\sum_{j=1}^{n}\widehat{\mu}(x,u,\widehat{V}_j)$, where $\widehat{\mu}(\cdot,\cdot,\cdot)$ was defined in (3.9), and $\mu(x,u)$ for $\mathbb{E}\Big[\mu(x,u,V_j)\Big]$ with $\mu(\cdot,\cdot,\cdot)$ defined in (3.6). Then, it holds that:

$$\begin{aligned}\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n}\sum_{j=1}^{n}\widehat{\mu}(x,u,\widehat{V}_j) - \mu(x,u)\Big| \leq& \sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n}\sum_{j=1}^{n}\Big\{\widehat{\mu}(x,u,\widehat{V}_j) - \widehat{\mu}(x,u,V_j)\Big\}\Big| \\ &+ \sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n}\sum_{j=1}^{n}\widehat{\mu}(x,u,V_j) - \mu(x,u)\Big| \\ =& I_1 + I_2\end{aligned}$$

The first term can be addressed through a combination of arguments from Lemma A1 and Lemma B4, B5 in Gutknecht (2011). In the following, a sketch of the basic steps will be provided. First, notice that I_1 can be further decomposed into:

$$\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Bigg|\frac{1}{n}\sum_{i=1}^{n}\Bigg\{\frac{\widehat{s}(x,u,\widehat{V}_i)-\widehat{s}(x,u,V_i)}{\widehat{f}(x,V_i)}-\frac{\widehat{f}(x,u,V_i)-\widehat{f}(x,u,V_i)}{\widehat{f}(x,u,V_i)}\times\widehat{\mu}(x,u,\widehat{V}_i)\Bigg\}\Bigg|$$

where

$$\widehat{s}(x,u,\widehat{V}_i)=\frac{1}{nh^3}\sum_{j=1}^{n}\widehat{I}_iW_jK_h(x-X_j)\times K_h(u-U_j)\times K_h(\widehat{V}_i-\widehat{V}_j)$$

and

$$\widehat{f}(x,u,\widehat{V}_i)=\frac{1}{nh^3}\sum_{j=1}^{n}\widehat{I}_iK_h(x-X_j)\times K_h(u-U_j)\times K_h(\widehat{V}_i-\widehat{V}_j)$$

with $\widehat{f}(x,u,V_i)$ and $\widehat{s}(x,u,V_i)$ defined analoguously using I_i, V_j, respectively. Only the first term is examined, while the second follows by identical steps. Using standard arguments (see Lemma A1 in Gutknecht (2011) for details) and some algebra, one can show that the leading terms of the numerator are:

$$I_{11}=\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n^2h^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\big(\widehat{I}_i-I_i\big)W_jK_h(x-X_j)\times K_h(u-U_j)\times K_h(V_i-V_j)\Big|$$

and

$$I_{12}=\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n^2h^3}\sum_{i=1}^{n}\sum_{j=1}^{n}I_iW_jK_h(x-X_j)\times K_h(u-U_j)\times K_h(\widehat{V}_i-\widehat{V}_j)\Big|$$

The term I_{11} can be adressed using the same argument as in Lemma B4 of Gutknecht (2011). That is, omitting the dependence of the indicator function on x, u for notational simplicity, notice that $\widehat{I}_i - I_i = I\{v_a \leq \widehat{V}_i \leq v_b\} - I\{v_a \leq V_i \leq v_b\} = \big(I\{\widehat{V}_i \leq v_b\} - I\{V_i \leq v_b\}\big) + \big(I\{\widehat{V}_i \geq v_a\} - I\{V_i \geq v_a\}\big)$, where v_a, v_b denote the endpoints of the marginal support of $\mathcal{V}$ with $v_a < v_b$. Focusing on $\big(I\{\widehat{V}_i \leq v_b\} - I\{V_i \leq v_b\}\big)$ and reverting to condition A2, the term can be rewritten as:

$$\begin{aligned}
&\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n^2h^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\big(I\{V_i\leq v_b+V_i-\widehat{V}_i\}-I\{V_i\leq v_b\}\big)W_jK_h(x-X_j)\times K_h(u-U_j)\times K_h(V_i-V_j)\Big|\\
=&\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n^2h^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\big(F(v_b+V_i-\widehat{V}_i)-F(v_b)\big)W_jK_h(x-X_j)\times K_h(u-U_j)\times K_h(V_i-V_j)\Big|+o_p(1)\\
=&\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\big(F^{(1)}(\overline{V}_b)(\widehat{V}_i-V_i)\big)\,\mathbb{E}\Big[\frac{1}{h^3}W_jK_h(x-X_j)\times K_h(u-U_j)\times K_h(V_i-V_j)\Big]\Big|+o_p(1)
\end{aligned}$$

where the term of smaller order after the first equality follows by adding and subtracting $\big(F(v_b + V_i - \widehat{V}_i) - F(v_b)\big)$. The second equality on the other hand is obtained by a mean value expansion ($\overline{V}_b$ denotes the intermediate value) accompanied by addition and subtraction of $\mathbb{E}[(1/h^3)W_jK_h(x - X_j \times K_h(u - U_j \times K_h(V_i - V_j)]$, a change of variables, and an application of Rosenthal's inequality (using A1 and A3), which yields the $o_p(1)$ term. Since $(\widehat{V}_i - V_i) = (\widehat{g}(Z_i) - g(Z_i))$, one can show that the last expression can be approximated by a second order U-statistic (see Lemma B4 in Gutknecht (2011) for details). Then, applying Lemma 3.1 of Powell, Stock, and Stoker (1989) and deriving the projection of this degenerate second order U-statistic, gives the convergence rate for I_{11}:

$$\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n^2h^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\big(\widehat{I}_i-I_i\big)W_jK_h(x-X_j)\times K_h(u-U_j)\times K_h(V_i-V_j)\Big|=O_p(n^{-\frac{1}{2}})$$

The term I_{12}, on the other hand, can be adressed using arguments from Lemma B5 in Gutknecht

(2011). A mean value expansion around $(V_i - V_j)$ yields:

$$\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n^2h^4}\sum_{i=1}^{n}\sum_{j=1}^{n}I_iW_jK_h(x-X_j)\times K_h(u-U_j)\times K_h^{(1)}(\overline{V}_i-\overline{V}_j)\big\{(\widehat{V}_i-\widehat{V}_j)-(V_i-V_j)\big\}\Big|$$

where $K_h^{(1)}$ denotes the derivative of the kernel function and $(\overline{V}_i - \overline{V}_j)$ some intermediate value. Rewriting again $(\widehat{V}_i - V_i) = (\widehat{g}(Z_i) - g(Z_i))$, one obtains:

$$\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n^2h^4}\sum_{i=1}^{n}\sum_{j=1}^{n}I_iW_jK_h(x-X_j)\times K_h(u-U_j)\times K_h^{(1)}(\overline{V}_i-\overline{V}_j)\big\{(\widehat{g}(Z_i)-g(Z_i))+(\widehat{g}(Z_j)-g(Z_j))\big\}\Big|$$

Using the steps of Lemma B5 in Gutknecht (2011) (that is, adding and subtracting $\mathbb{E}[(1/h_n^4)I_iW_jK_h(x-X_j)\times K_h(u-U_j)\times K_h^{(1)}(\overline{V}_i-\overline{V}_j)]$, applying integration by parts, change of variables, and an application of Rosenthal's inequality (using A1 and A3), and finally approximating the statistic by a second order U-statistic) yields the same convergence rate for I_{12} as before:

$$\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n^2h^3}\sum_{i=1}^{n}\sum_{j=1}^{n}I_iW_jK_h(x-X_j)\times K_h(u-U_j)\times K_h(\widehat{V}_i-\widehat{V}_j)\Big| = O_p(n^{-\frac{1}{2}})$$

The second term I_2 on the other hand can be bounded by:

$$\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n}\sum_{j=1}^{n}\Big\{\widehat{\mu}(x,u,V_j)-\mu(x,u,V_j)\Big\}\Big| + \sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n}\sum_{j=1}^{n}\mu(x,u,V_j)-\mu(x,u)\Big| = I_{21}+I_{22}$$

Since $\widehat{\mu}(x,u,V_i)$ is a consistent estimator for $\mu(x,u,V_i)$ and $\mathbb{E}[\mu(x,u,V_i)^2] < \infty$ for every $x \in \mathcal{X}$ and $u \in \mathcal{U}$, one obtains by standard arguments:

$$\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\frac{1}{n}\sum_{j=1}^{n}\Big\{\widehat{\mu}(x,u,V_j)-\mu(x,u,V_j)\Big\}\Big| = O_p(n^{-\frac{1}{2}})$$

Likewise, since $\mu(x,u,V_j)$ is continuous (and hence bounded) on W and using A2, the same rate can be obtained for I_{22}. Hence the result of the lemma follows. ■

Proof of Lemma A2. The proof is similar to that of Lemma 3.1 in Ghosal, Sen, and van der Vaart (2000). Hence only differences will be pointed out. Consider the following class of functions $\mathcal{F} = \{f_{t,x} : (t,x) \in \mathcal{T}\times\mathcal{X}\}$, where:

$$f_{t,x}:(t,x) = \Big(\mu(x,U_1)-\mu(x,U_2)\Big)\,\mathrm{sign}\Big(U_2-U_1\Big)\frac{1}{h_n^2}K_h(U_2-t)K_h(U_1-t)$$

This class is contained in the product of the three classes:

$$\mathcal{F}_1 = \Big\{h_n^{-2}\Big(\mu(x,U_1)-\mu(x,U_2)\Big)\,\mathrm{sign}\Big(U_2-U_1\Big)\times\mathbb{I}\Big\{|U_2-U_1|\le 2h_n\Big\} : x\in\mathcal{X}\Big\}$$

$$\mathcal{F}_2 = \Big\{K_h(U_1-t) : t\in\mathcal{T}\Big\}$$

$$\mathcal{F}_3 = \Big\{K_h(U_1-t) : t\in\mathcal{T}\Big\}$$

with envelopes $h_n^{-2}C|U_1-U_2|\times\mathbb{I}\Big\{|U_2-U_1|\le 2h_n\Big\}$, $\|K_h\|_\infty$, and $\|K_h\|_\infty$, respectively. Since K_h is of bounded variation and $\mu(\cdot,\cdot)$ satisfies a Lipschitz condition, by Lemma 2.6.15, 2.6.16, and 2.6.18 of van der Vaart and Wellner (1996) $\mathcal{F}$ is a Vapnik-Cervonekis (VC) class with the envelope Ch_n^{-2}, where C is some generic finite constant. Then, applying Theorem 2.6.7 of van der Vaart and Wellner (1996) and following the steps of Lemma 3.1 in Ghosal, Sen, and van der Vaart (2000), one obtains (Theorem A.2):

$$\mathbb{E}\Big[\sup_{t\in T}\sup_{x\in\mathcal{X}}\Big|\widehat{M}_{n,\mu}(x,t)-\widehat{M}^{p}_{n,\mu}(x,t)\Big|\Big]\leq Cn^{-1}h_n^{-2}$$

Notice that, similar to Lee, Linton, and Whang (2009), the rate by which the term can be bounded slightly differs from the one in Ghosal, Sen, and van der Vaart (2000) since replacing the estimator by the true $\mu(x,U)$ results in error of order $o_p(1)$ by the convergence result of Lemma A1. ■

Proof of Lemma A3. Unlike in the bivariate case of Ghosal, Sen, and van der Vaart (2000), there is no need to show that Theorem 1.1 of Rio (1994) holds. Instead, one can directly appeal to Theorem 3 and the subsequent Corollary of Komlos, Major, and Tusnady (1975). In the following, write:

$$\Psi_{n,t}(U)=\int\Big(\widetilde{\mu}(\omega)-\widetilde{\mu}(U)\Big)\operatorname{sign}\Big(U-\omega\Big)\frac{1}{h_n^2}K_h(\omega-t)dF(\omega)K_h(U-t)$$

As pointed out by Ghosal, Sen, and van der Vaart (2000), since U is supported on a compact interval with positive and continuous density, a simple affine transformation can be used to normalize the empirical process and t, which is necessary to satisfy the formal setup of Komlos, Major, and Tusnady (1975), who require U to have a uniform distribution on $[0,1]$. The transformation can subsequently be reversed by an inverse transformation. Also notice that:

$$\sup_{t\in\mathcal{T}}\Big(h_n\cdot\Psi_{n,t}(U)\Big)=O(1)$$

Defining a sequence of centered Gaussian processes with covariance:

$$\mathbb{E}\Big[G_n(t_1)G_n(t_2)\Big]=\mathbb{E}\Big[\Psi_{n,t_1}(U)\Psi_{n,t_2}(U)\Big]$$

one can apply Theorem 3 and the subsequent Corollary of Komlos, Major, and Tusnady (1975) using $\{h_n\Psi_{n,t}(U)\}$ and the Brownian bridge just defined. The same arguments as in Ghosal, Sen, and van der Vaart (2000) hold and switching back to the original equation, the result follows. ■

Proof of Lemma A4. Assertions (i) and (ii) are identical to the proof of Lemma 3.3 in Ghosal, Sen, and van der Vaart (2000). In (iii), to deal with the fact that $\widehat{\sigma}_{n,x}(t)$ depends on the estimated $\mu(x,U)$, let $\widetilde{\sigma}^2_{n,\mu}(t,x)$ be identical to $\widehat{\sigma}^2_n(t)$ except for $\widehat{\mu}(x,U)$ being replaced by $\mu(x,U)$, which results in an error of smaller order by the uniform convergence result of Lemma A1 and the bandwidth conditions stated in Theorem 6. As in Lemma A2, this will lead to a slightly larger bound than in Ghosal, Sen, and van der Vaart's (2000) paper. Again, using the modified $\widetilde{\sigma}^2_{n,\mu}(t,x)$ and following the steps of Lemma 3.3 in Ghosal, Sen, and van der Vaart (2000) yields:

$$\sup_{t\in\mathcal{T}}\sup_{x\in\mathcal{X}}\Big|\widetilde{\sigma}^2_{n,\mu}(t,x)-\mathbb{E}\Big[\widetilde{\sigma}^2_{n,\mu}(t,x)\Big]\Big|=O_p(n^{-\frac{1}{2}}h_n^{-2}+n^{-1}h_n^{-3}+n^{-\frac{3}{2}}h_n^{-4})$$

■

Proof of Lemma A5. Let $\mathcal{G}_n$ denote the class of functions $\{g_{n,t}(U):t\in\mathcal{T}\}$ where $g_{n,t}(U)=\frac{\Psi_{n,t}(U)}{\sigma_n(t)}$. Let $\overline{\mathcal{G}}_n$ stand for the class of functions $\{\overline{g}_{n,t}:t\in\mathcal{T}\}$ with $\overline{g}_{n,t}(U)=\frac{\overline{\Psi}_{n,t}(U)}{\overline{\sigma}_n(t)}$ where:

$$\overline{\Psi}_{n,t}(U)=\int\Big(\widetilde{\mu}(\omega)-\widetilde{\mu}(U)\Big)\operatorname{sign}\Big(U-\omega\Big)\frac{1}{h_n^2}K_h(\omega-t)d\omega K_h(U-t)$$

and

$$\overline{\sigma}_n(t)=\Big(\int\Big(\int\Big(\widetilde{\mu}(\omega)-\widetilde{\mu}(Y)\Big)\operatorname{sign}\Big(Y-\omega\Big)\frac{1}{h_n^2}K_h(\omega-t)d\omega\Big)\frac{1}{h_n^2}K_h^2(Y-t)dY\Big)^{\frac{1}{2}}f(U)^{\frac{1}{2}}$$

As explained in Remark 8.3 of Ghosal, Sen, and van der Vaart (2000), it is possible to extend Lemma A3 to show that there is a sequence of Browian bridges $\{B_n(g):g\in\mathcal{G}_n\cup\overline{\mathcal{G}}_n\}$ with $\mathbb{E}\Big[B_n(g)\Big]=0$ and $\mathbb{E}\Big[B_n(g_1)B_n(g_2\Big]=\operatorname{cov}(g_1,g_2)$ for $g,g_1,g_2\in\mathcal{G}_n\cup\overline{\mathcal{G}}_n$ and with continuous sample paths w.r.t. the L_1 metric such that $G_n(t) = \sigma_n(t)B_n(\Psi_{n,t}(U))$, where $G_n(t)$ was defined in Lemma A3. Set $\overline{\xi}_n(t) = B_n(\overline{g}_{n,t})$ and note that $\gamma_n(t) = G_n(t)/\sigma_n(t) - \overline{\xi}_n(t)$ is also a mean zero Gaussian process with:

$$\mathbb{E}\Big[\gamma_n(t_1)\gamma_n(t_2)\Big] = \mathbb{E}\Big[\Big(g_{n,t_1} - \overline{g}_{n,t_1}\Big)\Big(g_{n,t_2} - \overline{g}_{n,t_2}\Big)\Big]$$

The rest of the proof follows as in the proof of Lemma 3.4 of Ghosal, Sen, and van der Vaart (2000). Notice that a mean value expansion in the numerator and denominator is required to obtain the form of the covariance $\rho(\cdot)$ in (3.16). ■

Proof of Lemma A6. Notice that:

$$\begin{aligned}
\Big|\widehat{M}^p_{n,\widehat{\mu}}(x,t) - \widehat{M}^p_{n,\mu}(x,t)\Big| =& \Big|\frac{2}{n}\sum_{i=1}^{n}\Big\{\int\Big(\widehat{\mu}(x,\omega) - \widehat{\mu}(x,U_i)\Big)\operatorname{sign}\Big(U_i-\omega\Big)\frac{1}{h_n^2}K_h(\omega-t)dF(\omega)K_h(U_i-t)\\
&-\int\Big(\mu(x,\omega) - \mu(x,U_i)\Big)\operatorname{sign}\Big(U_i-\omega\Big)\frac{1}{h_n^2}K_h(\omega-t)dF(\omega)K_h(U_i-t)\Big|\\
\leq&\Big|\frac{2}{n}\sum_{i=1}^{n}\Big\{\int\Big(\widehat{\mu}(x,\omega) - \mu(x,\omega)\Big)\operatorname{sign}\Big(U_i-\omega\Big)\frac{1}{h_n^2}K_h(\omega-t)dF(\omega)K_h(U_i-t)\Big|\\
&+\Big|\frac{2}{n}\sum_{i=1}^{n}\Big(\widehat{\mu}(x,U_i) - \mu(x,U_i)\Big)\int\operatorname{sign}\Big(U_i-\omega\Big)\frac{1}{h_n^2}K_h(\omega-t)dF(\omega)K_h(U_i-t)\Big|
\end{aligned}$$

Using A5, the first term is bounded by:

$$C\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\widehat{\mu}(x,u) - \mu(x,u)\Big|\frac{1}{nh_n}\sum_{i=1}^{n}K_h(U_i-t)$$

where C is some generic finite constant independent of t and x. Using assumptions A1 to A3 and standard empirical process theory, one sees that:

$$\sup_{t\in\mathcal{T}}\Big|\frac{1}{nh_n}\sum_{i=1}^{n}K_h(U_i-t)\Big| = O_p(1)$$

Moreover, by the uniform convergence result of Lemma A1 and the bandwidth conditions of Theorem 6, the second term can be bounded by:

$$\begin{aligned}
&C\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\widehat{\mu}(x,u) - \mu(x,u)\Big|\frac{1}{nh_n^2}\sum_{i=1}^{n}\int\operatorname{sign}\Big(U_i-\omega\Big)K_h(\omega-t)dF(\omega)K_h(U_i-t)\\
\leq&C\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\Big|\widehat{\mu}(x,u) - \mu(x,u)\Big|\sup_{t\in\mathcal{T}}\Big|\frac{1}{nh_n^2}\sum_{i=1}^{n}\int\operatorname{sign}\Big(U_i-\omega\Big)K_h(\omega-t)dF(\omega)K_h(U_i-t)\Big|\\
=&O_p(n^{-\frac{1}{2}})
\end{aligned}$$

where the inequality follows from assumptions A2, A3, while the convergence rate is taken from Lemma A1. C is again a generic, finite constant. Combining this result with the result of Lemma A2 yields the claim of the lemma since $n^{-1}h_n^{-2}$ is of smaller order given that $nh_n^3 \longrightarrow \infty$. ■

A3.2 Tables & Figures

Table 3.1: Monte Carlo Simulation

H_0 is true, $\alpha = .5$		
$h = 0.8 \cdot n^{-\frac{1}{5}}$	$n = 100$	0.10733
	$n = 200$	0.13467
	$n = 300$	0.16067
$h = 1 \cdot n^{-\frac{1}{5}}$	$n = 100$	0.03677
	$n = 200$	0.03133
	$n = 300$	0.03400
$h = 1.2 \cdot n^{-\frac{1}{5}}$	$n = 100$	0.02067
	$n = 200$	0.00733
	$n = 300$	0.00267
H_0 is false, $\alpha = .5$		
$h = 0.8 \cdot n^{-\frac{1}{5}}$	$n = 100$	0.98333
	$n = 200$	1.00000
	$n = 300$	1.00000
$h = 1 \cdot n^{-\frac{1}{5}}$	$n = 100$	0.96733
	$n = 200$	0.99933
	$n = 300$	1.00000
$h = 1.2 \cdot n^{-\frac{1}{5}}$	$n = 100$	0.94333
	$n = 200$	1.00000
	$n = 300$	1.00000

Table 3.2: Descriptive Statistics

Males, 1996-2002		
Variable	Hourly Res. Wage (Pounds)	Elapsed Duration (Days)
Obs.	$1,034$	$1,034$
Mean	5.169163	224.824
Std. Dev.	2.156003	291.4457
25% quantile	3.953016	56
50% quantile	4.751848	155
75% quantile	5.636979	316
Females, 1996-2002		
Variable	Hourly Res. Wage (Pounds)	Elapsed Duration (Days)
Obs.	974	974
Mean	4.828762	288.6828
Std. Dev.	2.211663	317.3773
25% quantile	3.690456	82
50% quantile	4.349572	212
75% quantile	5.279831	345
Min. Wage Correction: Males, 1996-2002		
Variable	Hourly Res. Wage (Pounds)	Elapsed Duration (Days)
Obs.	914	914
Mean	5.427778	222.7144
Std. Dev.	2.147863	284.8766
25% quantile	4.152823	55
50% quantile	5.099003	158
75% quantile	5.901098	316
Min. Wage Correction: Females, 1996-2002		
Variable	Hourly Res. Wage (Pounds)	Elapsed Duration (Days)
Obs.	846	846
Mean	5.070153	281.9728
Std. Dev.	2.262114	326.456
25% quantile	3.945885	81
50% quantile	4.514527	210.5
75% quantile	5.399568	344

Table 3.3: Descriptive Statistics (contd.)

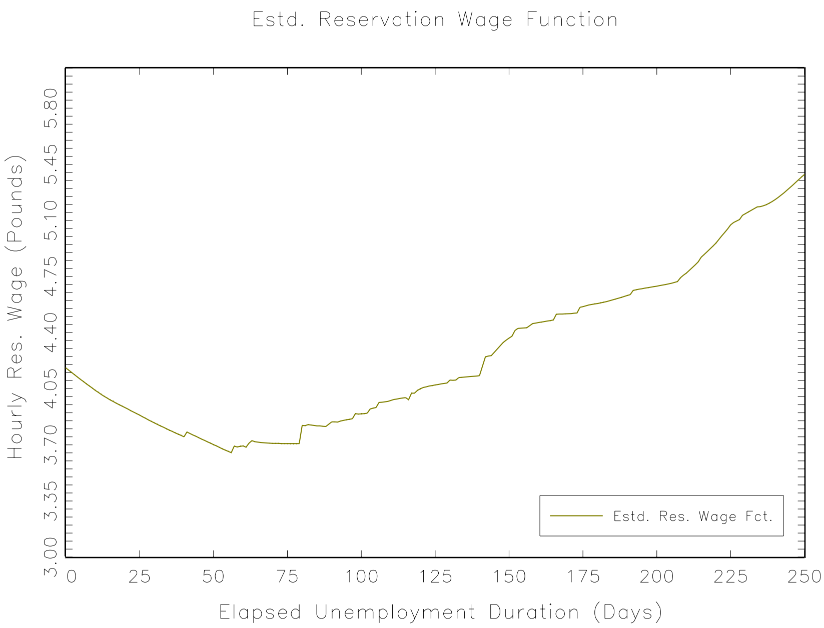
Min. Wage Correction: Males, 1996-2005		
Variable	Hourly Res. Wage (Pounds)	Elapsed Duration (Days)
Obs.	1,203	1,203
Mean	5.759774	206.8678
Std. Dev.	2.565025	259.0299
25% quantile	4.312823	51
50% quantile	5.208333	143
75% quantile	6.387328	304
Min. Wage Correction: Females, 1996-2005		
Variable	Hourly Res. Wage (Pounds)	Elapsed Duration (Days)
Obs.	1,105	1,105
Mean	5.323397	260.9674
Std. Dev.	2.412956	295.9899
25% quantile	4.174159	74
50% quantile	4.887904	205
75% quantile	5.636979	338
Min. Wage Correction: Males, 1996-2007		
Variable	Hourly Res. Wage (Pounds)	Elapsed Duration (Days)
Obs.	1,307	1,307
Mean	5.842997	201.1446
Std. Dev.	2.642724	251.3508
25% quantile	4.36205	50
50% quantile	5.274261	137
75% quantile	6.436663	296
Min. Wage Correction: Females, 1996-2007		
Variable	Hourly Res. Wage (Pounds)	Elapsed Duration (Days)
Obs.	1,203	1,203
Mean	5.456375	253.9609
Std. Dev.	2.497814	287.4156
25% quantile	4.223865	73
50% quantile	4.970179	201
75% quantile	5.958292	335

Table 3.4: Test Outcomes - H_0: Montonically Declining Rerservation Wages

	CV 1%	CV 5%	CV 10%	Test Statistic
Men: 1996-2002	3.7730	3.0322	2.7050	11.5773
Men: 1996-2002, Min. Wage	3.7661	3.0167	2.6858	10.8199
Women: 1996-2002	3.7679	3.0208	2.6909	13.1156
Women: 1996-2002, Min. Wage	3.7642	3.0123	2.6803	12.3965
Men: 1996-2005, Min. Wage	3.7719	3.0297	2.7019	11.5756
Women: 1996-2005, Min. Wage	3.7699	3.0253	2.6965	8.6792
Men: 1996-2007, Min. Wage	3.7805	3.0482	2.7248	12.9244
Women: 1996-2007, Min. Wage	3.7772	3.0412	2.7161	14.5392

Figure 3.1: Men - Estimated Reservation Wage Function

(a) 1996 - 2002 Sample



(b) 1996 - 2002 Sample ("Min. Wage Correction")

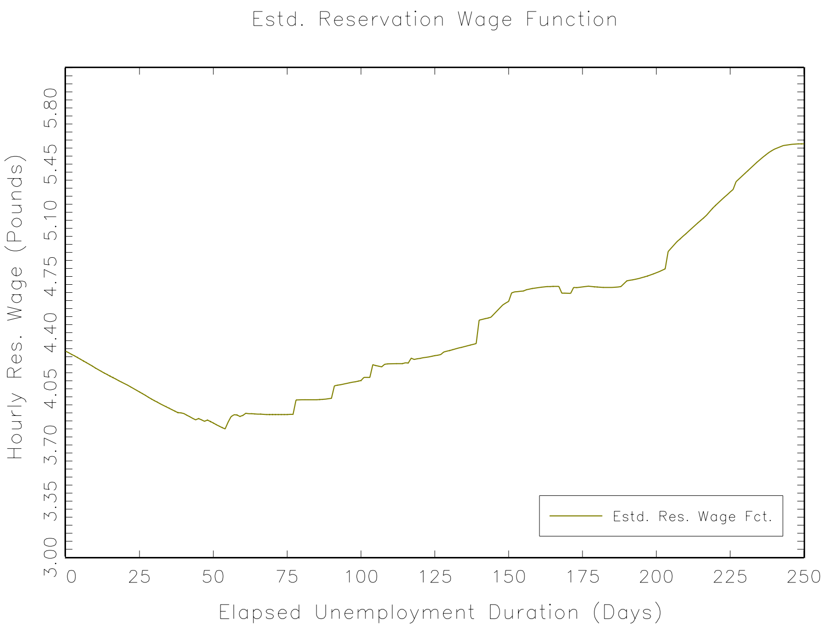
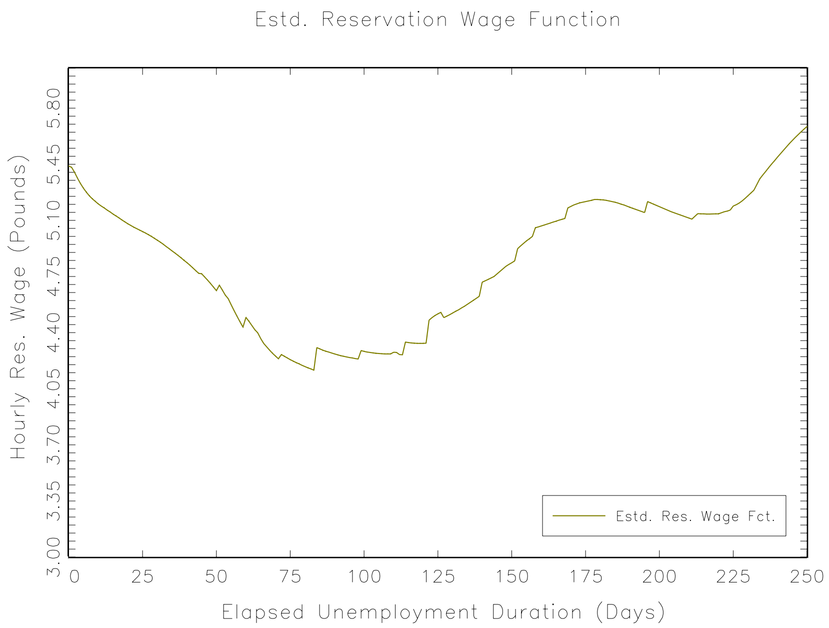


Figure 3.2: Men - Estimated Reservation Wage Function (contd.)

(a) 1996 - 2005 Sample ("Min. Wage Correction")



(b) 1996 - 2007 Sample ("Min. Wage Correction")

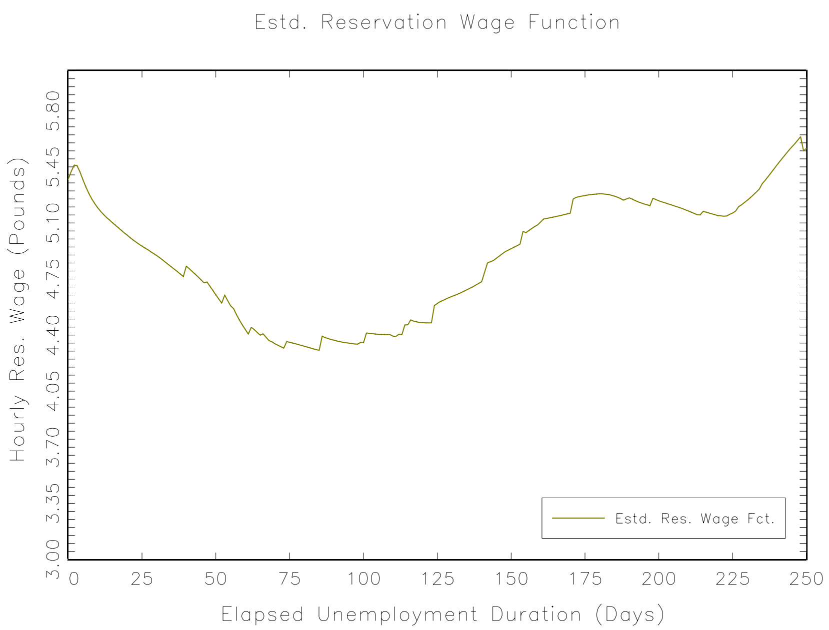
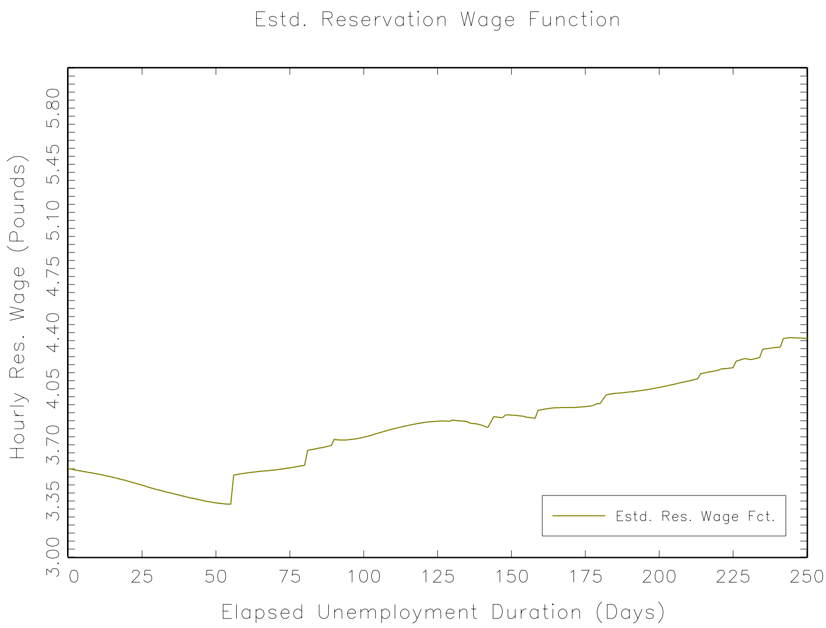


Figure 3.3: Women - Estimated Reservation Wage Function

(a) 1996 - 2002 Sample



(b) 1996 - 2002 Sample ("Min. Wage Correction")

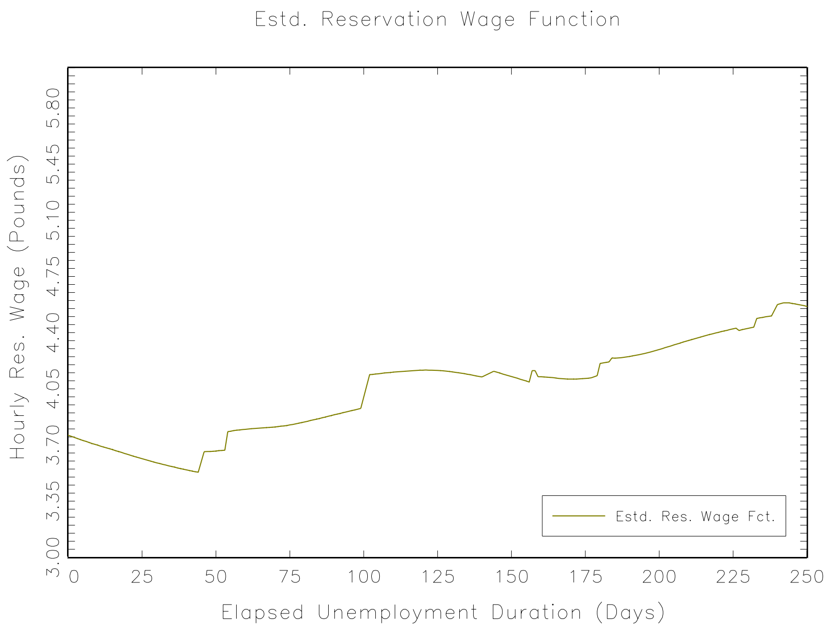
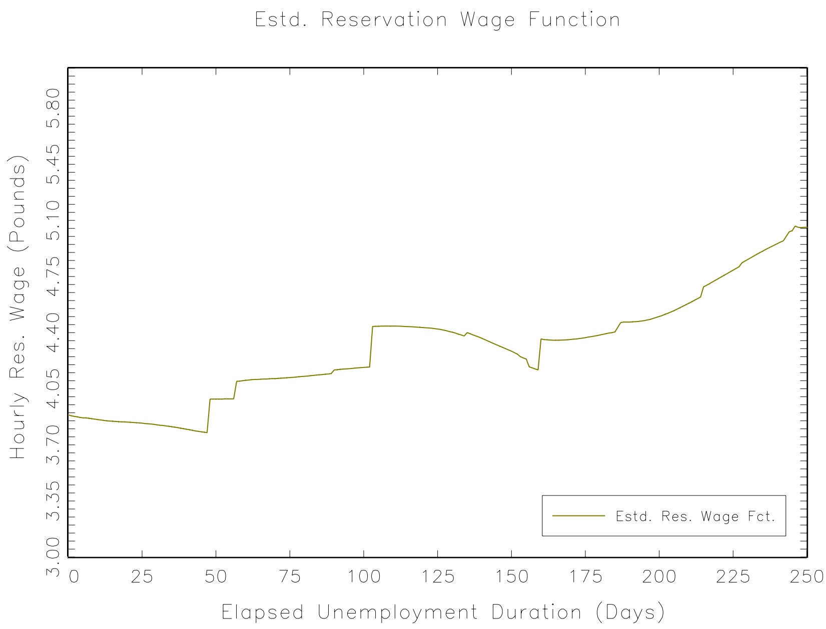


Figure 3.4: Women - Estimated Reservation Wage Function (contd.)

(a) 1996 - 2005 Sample ("Min. Wage Correction")



(b) 1996 - 2007 Sample ("Min. Wage Correction")

Bibliography

Abrevaya, J. (1999): "Rank Regression for Current-status data: Asymptotic normality," *Statistics and Probability Letters*, 43, 275–287.

——— (2000): "Rank estimation of a generalized fixed-effects regression model," *Journal of Econometrics*, 95(1), 1–23.

Abrevaya, J., and J. A. Hausman (1999): "Semiparametric estimation with mismeasured dependent variables: An application to duration models for unemployment spells.," *Annales d'Economie et de Statistique*, 55/56, 243–275.

——— (2004): "Response error in a transformation model with an application to earnings-equation estimation," *Econometrics Journal*, 7, 366–388.

Addison, J., M. Centeno, and P. Portugal (2004): "Reservation Wages, Search Duration, and Accepted Wages in Europe," Discussion Paper 1252, IZA Discussion Paper.

Addison, J., J. Machado, and P. Portugal (2011): "The Reservation Wage Unemployment Duration Nexus," Revised Version of the IZA Discussion Paper No. 5077.

Addison, J. T., M. Centeno, and P. Portugal (2010): "Unemployment benefits and reservation wages: key elasticities from a stripped-down job search approach," *Economica*, 77(305), 46–59.

Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1843.

Ameniya, T. (1974): "The Nonlinear Two-Stage Least Squares Estimator," *Jounal of Econometrics*, 2, 105–110.

Arcones, M., and E. Gine (1992): "On the bootstrap of U and V statistics," *Annals of Statistics*, 20, 655–674.

Augustin, T. (1999): "Correcting for measurement error in parametric duration models by quasi-likelihood," Discussion Paper 157, Sonderforschungsbereich 386, Universität München.

Birke, M., and N. Neumeyer (2010): "Testing monotonicity of regression functions an empirical process approach," Discussion paper, University of Bochum.

Blundell, R., and J. L. Powell (2003): "Endogeneity in nonparametric and semiparametric regression models," in *Advances in Economics and Econometrics*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky, vol. 2, pp. 312–357. Cambridge University Press, Cambridge, UK, Eigth World Congress.

——— (2004): "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71, 581–913.

Bound, B., C. Brown, and N. Mathiowetz (2001): "Measurement Error in Survey Data," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 59, pp. 3705–3843. Elsevier.

Bowman, A., M. Jones, and I. Gijbels (1998): "Testing monotonicity of regression," *Journal of Computational and Graphical Statistics*, 7, 489–500.

Bricker, J., and G. V. Engelhardt (2007): "Measurement Error In Earnings Data In The Health and Retirement Study," Working Paper 16, Centre for Retirement Research at Boston College.

Brown, S., and K. Taylor (2009): "Reservation Wages, Expected Wages and the Duration of Unemployment: Evidence from British Panel Data," 3981, IZA Discussion Paper.

Butcher, K., and A. Case (1994): "The effect of sibling sex composition on women's education and earnings," *Quarterly Journal of Economics*, 109(3), 531–563.

Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–1160.

Cavanagh, C., and R. Sherman (1998): "Rank estimation for monotonic index models," *Journal of Econometrics*, 84, 351–381.

Chalak, K., and H. White (2011): "An Extended Class of Instrumental Variables for the Estimation of Causal Effects," *Canadian Journal of Economics*, 44, 1–51.

Chen, K., and S.-H. Lo (1997): "On the rate of uniform convergence of the product-limit estimator: strong and weak laws," *Annals of Statistics*, 25, 1050–1087.

Chen, X., H. Hong, and E. Tamer (2005): "Measurement Error Models with Auxiliary Data," *Review of Economic Studies*, 72(2), 343–366.

Chen, X., and D. Pouzo (2009): "Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals," *Journal of Econometrics*, 152, 46–60.

Chesher, A., M. Dumangane, and R. J. Smith (2002): "Duration response measurement error," *Journal of Econometrics*, 111, 169–1994.

Corradi, V., W. Distaso, and N. Swanson (2011): "Predictive Infererence for Integrated Volatility," *Journal of the American Statistical Association*, 106(496), 1496–1512.

Cristia, J., and J. Schwabish (2007): "Measurement Error in the SIPP: Evidence from matched administrative records," Working paper, Congressional Budget Office.

Darolles, S., Y. Fan, J. P. Florens, and E. Renault (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79(5), 1541–1565.

Delgado, M. A., and J. C. Escanciano (2012): "Distribution-free tests of stochastic monotonicity," Journal of Econometrics (In Press, Accepted Manuscript).

Dumangane, M. (2007): "Measurement Error Bias Reduction in Unemployment Durations," Discussion paper, Cemmap Working Paper.

Garen, J. (1984): "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable," *Econometrica*, 52, 1199–1218.

Ghosal, S., A. Sen, and A. W. van der Vaart (2000): "Testing Monotonicity of Regression," *Annals of Statistics*, 28(4), 1054–1092.

Gijbels, I., P. Hall, M. Jones, and I. Koch (2000): "Tests for monotonicity of a regression mean with guaranteed level," *Biometrika*, 87, 663–673.

Glewwe, P., and H. Patrinos (1999): "The Role of the Private Sector in Education in Vietnam: Evidence From the Vietnam Living Standards Survey," *World Development*, 27(5), 887–902.

Goncalves, S., and H. White (2004): "Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models," *Journal of Econometrics*, 119, 199–219.

——— (2005): "Bootstrap Standard Error Estimates for Linear Regression," *Journal of the American Statistical Association*, 100(471), 970–979.

Gutknecht, D. (2011): "Nonclassical Measurement Error in a Nonlinear (Duration) Model," Working Paper 961, University of Warwick.

Hall, P., and J. Heckman (2000): "Testing for monotonicity of a regression mean by calibrating for linear functions," *Annals of Statistics*, 28, 20–39.

HAN, A. (1987): "Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator.," *Journal of Econometrics*, 35, 303–316.

HANSEN, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HÄRDLE, W. (1992): *Applied Nonparametric Regression*, no. 19 in Econometric Society Monographs. Cambridge University Press.

HAUSMAN, J. A., J. ABREVAYA, AND F. M. SCOTT-MORTON (1998): "Misclassification of the dependent variable in a discrete-response setting," *Journal of Econometrics*, 87(2), 239–269.

HODERLEIN, S., AND J. WINTER (2010): "Structural Measurement Error in Nonseparable Models," *Journal of Econometrics*, 157, 432–440.

HOEFFDING, W. (1948): "A Class of Statistics with Asymptotically Normal Distribution," *Annals of Mathematical Statistics*, 19, 293–325.

HU, Y., AND S. M. SCHENNACH (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76(1), 195–216.

ICHINO, A., AND R. WINTER-EBMER (1999): "Lower and Upper Bounds of Returns to Schooling: An Exercice in IV Estimation with Dierent Instruments," *European Economic Review*, 43(6), 889–902.

IMBENS, G. W., AND W. K. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica*, 77(5), 1481–1512.

JÄCKLE, A. (2008): "Measurement Error and Data Collection Methods: Effects on Estimates from Event History Data," Working Paper 2008-13, Institute for Social and Economic Research (ISER).

JOCHMANS, K. (2010): "Estimating Monotone Index Models with Nonparametric Controls," Job-Market Paper.

KAPLAN, E., AND P. MEIER (1958): "Nonparametric Estimation for Incomplete Observations," *Journal of American Statistical Association*, 53, 457–481.

KHAN, S. (2001): "Two-stage rank estimation of quantile index models," *Journal of Econometrics*, 100(2), 319–355.

KIEFER, N., AND G. NEUMANN (1979): "A Empirical Job-Search Model, with a Test of the Constant Reservation Wage Hypothesis," *The Journal of Political Economy*, 87, 89–107.

——— (1981): "Individual Effects in a Non-Linear Model: Explicit Treatment of Heterogeneity in the Empirical Job Search Model," *Econometrica*, 49, 965–979.

KOLENIKOV, S., AND G. ANGELES (2009): "Socioeconomic Status Measurement with Discrete Proxy Variables: Is Principal Component Analysis a Reliable Answer?," *The Review of Income and Wealth*, 55, 128–165.

KOMLOS, J., P. MAJOR, AND G. TUSNADY (1975): "An Approximation of Partial Sums of Independent RV's, and the Sample DF I," *Zeitschrift fuer Wahrscheinlichkeitstheorie*, 32, 111–131.

KOUL, H., V. SUSARLA, AND J. VAN RYZIN (1981): "Regression analysis with randomly right-censored data," *Annals of Statistics*, 9, 1276–1288.

LAI, T., Z. YING, AND Z. ZHENG (1995): "Asymptotic Normality of a class of adaptive statistics with applications to synthetic data methods for censored regression," *Journal of Multivariate Analysis*, 52, 259–279.

LANCASTER, T. (1985): "Simultaneous equations models in applied search theory," *Journal of Econometrics*, 28(1), 113–126.

LEE, S., O. LINTON, AND Y.-J. WHANG (2009): "Testing for Stochastic Monotonicity," *Econometrica*, 77, 585–602.

Lewbel, A. (1998): "Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors," *Econometrica*, 66(1), 105–121.

Lopez, O. (2009): "Single-index regression models with right-censored responses," *Journal of Statitical Planning and Inference*, 139, 1082–1097.

Lu, X., and M. Burke (2005): "Censored multiple regression by the method of average derivatives," *Journal of Multivariate Analysis*, 95, 182–205.

Lu, X., and T. Cheng (2007): "Randomly censored partially linear single-index models," *Journal of Multivariate Analysis*, 98, 1895–1922.

Mammen, E., O. Linton, and J. Nielsen (1999): "The Existence and Asymptotic Properties of a Backfitting Projection Algorithm under Weak Conditions," *Annals of Statistics*, 27, 1443–1490.

Matzkin, R. L. (2004): "Unobservable Instruments," Mimeo, Northwestern University.

——— (2007): "Nonparametric Survey Response Errors," *International Economic Review*, 48(4), 1411–1427.

Mortensen, D. (1986): "Job Search and Labor Market Analysis," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and R. Layard. North-Holland, Amsterdam.

Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

Newey, W. K., J. L. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67(3), 565–603.

Paull, G. (2002): "Biases in the Reporting of Labour Market Dynamics," Working Paper 02/10, Institute for Fiscal Studies, London.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57(6), 1403–1430.

Rio, E. (1994): "Local invariance principles and their application to density estimation," *Probability Theory and Related Fields*, 98, 21–45.

Rothe, C. (2009): "Semiparametric estimation of binary response models with endogenous regressors," *Journal of Econometrics*, 153, 51–64.

Serfling, R. (1980): *Approimation Theorems of Mathematical Statistics*, Wiley Series in Probability and Statistics. John Wiley, New York.

Sestito, P., and E. Viviano (2011): "Reservation Wages: Explaining Some Puzzling Regional Patterns," *Labour*, 25, 63–88.

Sherman, R. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61(1), 123–137.

Skinner, C., and K. Humphreys (1999): "Weibull Regression for Lifetimes measured with Error," *Lifetime Data Analysis*, 5, 23–37.

Srinivasan, C., and M. r. w. c. Zhou (1994): "Linear regression with censoring," *Journal of Multivariate Analysis*, 49, 179–201.

Stock, J. H., and F. Trebbi (2003): "Who Invented Instrumental Variable Regression?," *Journal of Economic Perspectives*, 17, 177–194.

Subbotin, V. (2008): "Essays on the econometric theory of rank regressions," Working Paper 14086, Northwestern University.

van den Berg, G. (1990): "Nonstationarity in Job Search Theory," *Review of Economic Studies*, 57, 255–277.

van der Vaart, A. W., and J. Wellner (1996): *Weak Convergence and Empirical Processes.* Springer, New York.

Vella, F. (1993): "Nonwage Benefits in a Simultaneous Model of Wages and Hours: Labor Supply Functions of Young Females," *Journal of Labour Economics*, 11, 704–723.