Do Reservation Wages Decline Monotonically?
A Novel Statistical Test

Daniel Gutknecht

No 991

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF
WARWICK

Do Reservation Wages Decline Monotonically? A Novel Statistical Test

Daniel Gutknecht

May 21, 2012

Abstract

This paper develops a test for monotonicity of the 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 paper 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 paper 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.

Key-Words: Reservation Wages, Test for Montonicity, Endogeneity, Control Function, Unobservable Instruments.

JEL Classification: C14, C36, C54, J64

Acknowledgements: I would like to thank my supervisors Valentina Corradi and Wiji Arulampalam for various helpful discussions and comments on the topic as well as for their encouraging support. I would also like to express my gratitude to Jennifer Smith for her support and advice regarding the collection of the data and the construction of the data set. All remaining errors are my sole responsibility.

University of Warwick, Department of Economics, Coventry CV4 7AL, UK, email: D.gutknecht@warwick.ac.uk

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

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

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.

The main theoretical contribution of this 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 properties. The 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 (similar to Lee, Linton, and Wang (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' approach to construct a bootstrap

[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).

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 Wang (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 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

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

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

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 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. Section 4 extends the framework of the previous sections to the case of 'unobservable instruments' as outlined above. Section 5 examines the finite sample properties of the estimator in a Monte Carlo Simulation. Section 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 7 concludes. All proofs and tables are postponed to the appendix.

2 Setup

To gain a better understanding of 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:

$$W_i = \widetilde{m}\Big(X_i, U_i\Big) + \epsilon_i \tag{1}$$

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 compact 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

[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.

$t \in \mathcal{T}$. 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 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 5.

Suppose a D_z-dimensional vector of instruments $Z_i = \{X_i, Z_{1i}\}$ exists. Moreover, the subvector Z_{1i} is allowed to include a non-zero constant and is assumed to be of dimension $D_{z1} \geq 2$ with at least one (nonconstant) continuous component. The following reduced form equation is assumed:

$$U_i = g\Big(Z_i\Big) + V_i \tag{2}$$

where $g(\cdot)$ is a real-valued, differentiable with non-zero derivative in its continuous argument(s). Without loss of generality, it is assumed that $\mathbb{E}\Big[V_i\Big|Z_i\Big] = 0$. V_i is the so called control function, which is assumed to satisfy a conditional mean independence condition:

$$\mathbb{E}\Big[\epsilon_i\Big|Z_i, V_i\Big] = \mathbb{E}\Big[\epsilon_i\Big|V_i\Big] \tag{3}$$

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 function of

[6]Notice that under this independence condition identification could be achieved even if the setup in (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.

Z_i (see section 6 for details). Equations (1) and (2) together with the assumption in (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:

$$\begin{aligned}
\mathbb{E}\Big[W_i\Big|U_i=u,Z_i=z\Big] =&\widetilde{m}\Big(x,u\Big)+\mathbb{E}\Big[\epsilon_i\Big|U_i=u,Z_i=z\Big]\\
=&\widetilde{m}\Big(x,u\Big)+\mathbb{E}\Big[\epsilon_i\Big|U_i=u,V_i=v\Big]\\
=&\widetilde{m}\Big(x,u\Big)+\mathbb{E}\Big[\epsilon_i\Big|V_i=v\Big]\\
\equiv&\widetilde{m}\Big(x,u\Big)+\lambda(v)
\end{aligned} \tag{4}$$

where the third equality follows from the conditional mean independence assumption in (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 (1). Then, identification up to an additive constant can be accomplished by reverting to the following lemma:

Lemma 1. *Suppose that equations (1) and (2) and the conditional mean independence assumption in (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 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

[7]The exact definition of identification in Newey, Powell, and Vella (1999, p.567) is based on equation (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* $\mathbb{P}[\delta(x,u)=c_m]=1$.

$f(v)$ satisfying $\int f(v)dv = 1$:

$$\begin{aligned}\int \mathbb{E}\Big[W_i\Big|U_i = u, X_i = x, V_i = v\Big]f(v)dv = &\int \mathbb{E}\Big[W_i\Big|U_i = u, X_i = x, V_i = v\Big]f(v)dv\\ &+\mathbb{E}\Big[\int \mathbb{E}\Big[W_i\Big|U_i = u, X_i = x, V_i = v\Big]f(v)dv\Big] - \mathbb{E}\Big[W_i\Big]\\ =&\widetilde{m}\Big(x,u\Big) + \mathbb{E}\Big[\widetilde{m}\Big(X_i,U_i\Big)\Big] - \mathbb{E}\Big[W_i\Big]\\ =&\widetilde{m}\Big(x,u\Big)\end{aligned} \tag{5}$$

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:

$$\mu(x,u,v) = -\mathbb{E}\Big[W_i\Big|X_i = x, U_i = u, V_i = v\Big] \tag{6}$$

so that:

$$\begin{aligned}\mu(x,u) =& -\int \mathbb{E}\Big[W_i\Big|X_i = x, U_i = u, V_i = v\Big]f(v)dv\\ =&m(x,u) - \int \mathbb{E}\Big[\epsilon_i\Big|V_i = v\Big]f(v)dv\\ =&m(x,u)\end{aligned} \tag{7}$$

for every $x,u \in \mathcal{X} \times \mathcal{U}$.[8] Equation (7) can be consistently estimated using the kernel estimator suggested in Gutknecht (2011):

$$\widehat{\mu}(x,u) = \frac{1}{n}\sum_{j=1}^{n}\widehat{\mu}(x,u,\widehat{V}_j) \tag{8}$$

where

$$\widehat{\mu}(x,u,\widehat{V}_j) = -\widehat{\mathbb{I}}_j\frac{\frac{1}{nh_n^3}\sum\limits_{i=1}^{n}W_iK_h(x-X_i)K_h(u-U_i)K_h(\widehat{V}_j-\widehat{V}_i)}{\frac{1}{nh_n^3}\sum\limits_{i=1}^{n}K_h(x-X_i)K_h(u-U_i)K_h(\widehat{V}_j-\widehat{V}_i)} \tag{9}$$

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

[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.

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 $\widehat{V}_i = U_i - g(Z_i)$, where $\widehat{g}(\cdot)$ can be estimated using for instance the Nadaraya-Watson kernel estimator. The modified test statistic is based on $\widehat{\mu}(\cdot,\cdot)$ in (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:

$$\widehat{U}_{n,x}(t) = \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} \Big(\widehat{\mu}(x,U_j) - \widehat{\mu}(x,U_i)\Big) \operatorname{sign}\Big(U_i - U_j\Big) \frac{1}{h_n^2} K_h(U_i - t) K_h(U_j - t) \quad (10)$$

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

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

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:

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

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 (of the interval $\mathcal{T}$) of a suitably scaled version of (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

[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.

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 (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:

$$S_n = \sup_{t \in \mathcal{T}} \left\{ \frac{\sqrt{n}\widehat{U}_{n,x}(t)}{\widehat{\sigma}_{n,x}(t)} \right\} \tag{11}$$

where

$$\begin{aligned}\widehat{\sigma}^2_{n,x}(t) = \frac{1}{n(n-1)(n-2)} \sum_{1 \le i,j,k \le n, i \ne j \ne k} &\Big(\widehat{\mu}(x, U_j) - \widehat{\mu}(x, U_i)\Big)\Big(\widehat{\mu}(x, U_k) - \widehat{\mu}(x, U_i)\Big) \\ &\times \operatorname{sign}\Big(U_i - U_j\Big) \operatorname{sign}\Big(U_i - U_k\Big) \frac{1}{h_n^4} K_h(U_j - t) K_h(U_k - t) K_h^2(U_i - t)\end{aligned} \tag{12}$$

is the estimated U-process for:

$$\begin{aligned}\sigma^2_{n,x}(t) &= \int \left(\int \Big(\mu(x,\omega) - \mu(x,U)\Big) \operatorname{sign}\Big(U - \omega\Big) \frac{1}{h_n} K_h(\omega - t) \right)^2 dF(\omega) \frac{1}{h_n^2} K_h^2(U - t) dF(U) \\ &= \int \left(\int \int \Big(\mu(x,\omega_1) - \mu(x,U)\Big)\Big(\mu(x,\omega_2) - \mu(x,U)\Big) \operatorname{sign}\Big(U - \omega_1\Big) \operatorname{sign}\Big(U - \omega_2\Big) \right. \\ &\quad \left. \times \frac{1}{h_n^2} K_h(\omega_1 - t) K_h(\omega_2 - t) dF(\omega_1) dF(\omega_2) \right) \frac{1}{h_n^2} K_h^2(U - t) dF(U)\end{aligned} \tag{13}$$

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 (11) can be approximated by a stationary Gaussian process with continuous sample paths, which will be achieved in Theorem 2 of the next section.

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 (10) be denoted as (Hoeffding, 1948):

$$\widehat{U}^p_{n,x}(t) = \frac{2}{n}\sum_{i=1}^{n}\int\Big(\widehat{\mu}(x,\omega) - \widehat{\mu}(x,U_i)\Big)\operatorname{sign}\Big(U_i - \omega\Big)\frac{1}{h_n^2}K_h(\omega - t)dF(\omega)K_h(U_i - t) \tag{14}$$

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

$$q(r) = \int\Big(\omega - r\Big)\operatorname{sign}\Big(r - \omega\Big)K(\omega)d\omega \tag{15}$$

and

$$\rho(s) = \frac{\int q(r)q(r-s)K(r)K(r-s)dr}{\int q^2(r)K^2(r)dr} \tag{16}$$

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 $x^{(d)}$ 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 (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:

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

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:

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

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 (7) and (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 on the regression function $\widetilde{m}(\cdot)$ and allow in principle a generalization of the setup to nonseparable 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 2. *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 (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 $U^p_{n,x}(t)$. Lemma A3 gives an approximation of the empirical process $\sqrt{n}U^p_{n,x}(t)$ 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 $\widehat{\mu}(\cdot,\cdot)$, while, similar to the proof of Theorem A.1 in Lee, Linton, and Wang (2009), Lemma A6 shows that, given the bandwidth conditions, it holds 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 appies for the estimated $\widehat{\sigma}_{n,x}(t)$.

The next step is to determine the asymptotic distribution of the test statistic. From Theorem 2 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 3. *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:

$$\text{Reject } H_0 \text{ if} \quad F_\infty(a_n(S_n - b_n)) \geq 1 - \alpha \tag{17}$$

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 2: bandwidth sequences such as $h_n = 1/\log(n)^\gamma$, $\gamma > 1$, or $h_n = 1/n^\eta$, $\eta < (1/3)$ if $D_z \leq 2$ and $\eta < (1/(D_z))$ if $D_z > 2$, 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 $[-1, 1]$, this integral 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 4. *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 (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 $nh_n^3/\log(n) \longrightarrow \infty$.

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, $U_i = s(T_i, \delta_i)$, where δ_i is an unobservable that is assumed to be independent of ϵ_i from (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:

$$\epsilon_i = r(T_i) + \eta_i \tag{18}$$

where η_i is an unobservable variable that is assumed to be independent of the observables T_i. η_i in the above equation must not to be confounded with the control function from section 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.[10] Thus, controlling for observed T_i implies controlling for the unobserved effect.[11] Inserting (18) into the regression equation yields:

$$W_i = \widetilde{m}\Big(X_i, U_i\Big) + r\Big(T_i\Big) + \eta_i \tag{19}$$

Equation (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 (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 (19) for a specific x (see their paper for details of the estimator), one may construct the following modified test

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

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

statistic:

$$S_n^* = \sup_{t \in \mathcal{T}} \left\{ \frac{\sqrt{n}\widehat{U}_{n,x}^*(t)}{\widehat{\sigma}_{n,x}^*(t)} \right\} \tag{20}$$

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

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

and $\widehat{m}(x, \cdot)$ is the backfitting estimator of $m(x, \cdot)$. $\widehat{\sigma}_{n,x}^*(t)$ is the equivalent to (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:

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:

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

where $m(\cdot, \cdot)$ is defined in (7) and C is a generic finite constant.

The following theorem can be established in analogy to Theorem 2:

Theorem 5. *Assume that A1, A3 as well as A2*, A4*, and A5* hold and that $\mathcal{T} \subset (0,1)$. 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 (16) and

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

The subsequent theorems follow in accordance with Theorem 3 and 4 from section 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.

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 1 of Appendix B2. 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 $K(v) = 0.75(1-v^2)\,\mathbb{I}[|v|\le 1]$ and the first stage functions $g(\cdot)$ in (2) and $\mu(\cdot)$ in (7) are estimated using the Nadaraya Watson estimator as explained in section 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 (11) with the interval $\mathcal{T}$ chosen to be $\mathcal{T} = \{0.05, 0.1, \dots, 0.9, 0.95\}$. 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 3. The simulations use sample sizes of $n = 100, 200, 300$ and $1,500$ replications are conducted for each simulation.

Reverting to Table 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 Wang (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.

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.[12] 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 $U(Z)$ 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.[13] 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.[14] 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

[12]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.

[13]$F_\omega(\cdot; X)$ 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.

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

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 $\lim_{\omega \to 0} F_\omega(\omega; X) = 0$ 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 $b_1 > b_2$.[15]

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:

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

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 decreasing in $U(Z)$. Obvi-

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

ously, 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 major reforms of the British tax credit system in April 2003. It is conjectured that this legislation change might have affected reservation wages as 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.[16]

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

[16] This only applies to spells recorded after 1999 when the minimum wage became applicable.

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 $\{0.9, 0.925, \ldots, 1.175, 1.2\}$.[17] Turning to the summary statistics of Table 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 minimum 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 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 1, 2, 3, and 4 display $-\widehat{\mu}(x, \cdot)$ from equation (8) for the first 250 days of elapsed unemployment duration for the different specifications. Figures 1 and 2 concern the estimated hourly reservation wage functions for males ($x = 1$), while Figures 3 and 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 1 and 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

[17] The biweight kernel has been used in all specifications.

jobseeker's allowance after six months for people with sufficient contributions.[18] 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 4 displays the results for the null of a decreasing reservation wage function across all the different specifications. The compact interval T is chosen to be bounded by 2% and 50% quantile for each sample and contains nineteen equally spaced points within those bounds.[19] Estimator and kernel function are as described in Section 5. The bandwidth size has been determined by $1.25 \cdot n^{-\frac{1}{5}}$ in compliance with Theorem 2, which required an upfront log transformation of elapsed duration.[20] Examining the test results in Table 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, as outlined in the introduction, being able to better understand the impact of elapsed unemployment duration on (hourly) reservation wages is paramount for future research due to its implications for the design of appropriate policies.

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 difficult to measure due to the simultaneity of both variables. Using

[18] 180 days also marks the starting point of the gateway period for unemployed qualifying for the 'New Deal' programme.

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

[20] 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).

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 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 (1) is altered to $W_i = \widetilde{m}\Big(X_i, U_i\Big) + \sigma(X_i)\epsilon_i$. This specification can be of interest in various empirical contexts 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 $\mathbb{E}\Big[\epsilon_i\Big|Z_i\Big] = 0$, 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 (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 is key objective for future research.

Appendix A1

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:

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

so that $\widehat{M}_{n,\widehat{\mu}}(x,t) = \widehat{U}_{n,x}(t)$ from (10). Moreover, let:

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

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 (14). Then, under assumptions A1 to A4, it holds that:

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

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 $\Psi_{n,t}(U) = \int\Big(\widetilde{\mu}(\omega)-\widetilde{\mu}(U)\Big)\,\mathrm{sign}\Big(U-\omega\Big)\frac{1}{h_n^2}K_h(\omega-t)dF(\omega)K_h(U-t)$ 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 (12) and (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 (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:

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

Proof of Theorem 2. 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 $\sup_{x\in\mathcal{X}}\sup_{u\in\mathcal{U}}\left|\widehat{\mu}(x,u)-\mu(x,u)\right|=O_p(n^{-\frac{1}{2}})$. Given that $nh_n^3\longrightarrow\infty$ by Theorem 2, this is slower than the rate established in Lemma A2 and thus $\sup_{x\in\mathcal{X}}\sup_{t\in\mathcal{T}}\left|\widehat{U}_{n,x}(t)-U^{p}_{n,x}(t)\right|=O_p(n^{-\frac{1}{2}})$ 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 4. 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:

$$h_nU_{n,x}(t)\xrightarrow{p}-\nabla_U m(x,t)\int\int|u-\nu|K(u)K(\nu)f(t)^2dud\nu$$

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 5. The proof follows identical steps to the one of Theorem 2. 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:

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

■

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 (9), and $\mu(x,u)$ for $\mathbb{E}\left[\mu(x,u,V_j)\right]$, where $\mu(\cdot,\cdot,\cdot)$ defined in (6). Then, it holds that:

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

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

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

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, the second follows by identical arguments. 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:

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

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

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

The term I_{12} 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 again the steps of Lemma B5 in Gutknecht (2011) (adding and subtracting $\mathbb{E}[(1/h_n^4)I_iW_jK_h(x-X_j)\times K_h(u-U_j)\times K_h^{(1)}(\overline{V}_i-\overline{V}_j)]$, integration by parts, change of variables, application of Rosenthal's inequality, and finally approximation 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, using A2 and since $\mu(x,u,V_j)$ is continuous (and hence bounded) on W, 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:

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

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

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

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

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

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):

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

Notice that, similar to Lee, Linton, and Wang (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 2. 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:

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

■

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:

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

and

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

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 $\mathbb{E}\Big[B_n(g_1)B_n(g_2\Big] = \mathrm{cov}(g_1, g_2)$ 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 (16). ■

Proof of Lemma A6. Notice that:

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

The first term is bounded by:

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

where C is some generic finite constant independent of t and x. Using assumptions A1 to A3 and standard empirical process theory, it is straightforward to see 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 2, the second term can be bounded by:

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

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 $nh_n^3 \longrightarrow \infty$. ■

Appendix A2

Table 1: Monte Carlo Results

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 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: Descriptive Statistics (contd.)

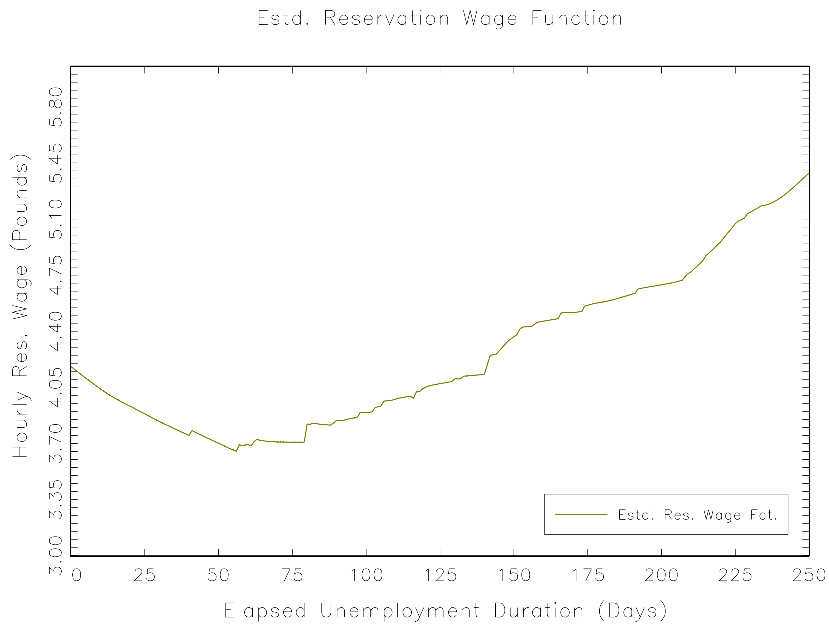
Min. Wage Correction: Males, 1996-2005		
Variable	Hourly Res. Wage (Pounds)	Elapsed Duration (Days)
Obs.	$1,203$	$1,203$
Mean	5.759774	206.8678
Std. Dev.	2.565025	259.0299
25% quantile	4.312823	51
50% quantile	5.208333	143
75% quantile	6.387328	304
Min. Wage Correction: Females, 1996-2005		
Variable	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 4: Test Results - 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 1: Men - Estimated Reservation Wage Function

(a) 1996 - 2002 Sample



(b) 1996 - 2002 Sample (“Min. Wage Correction”)

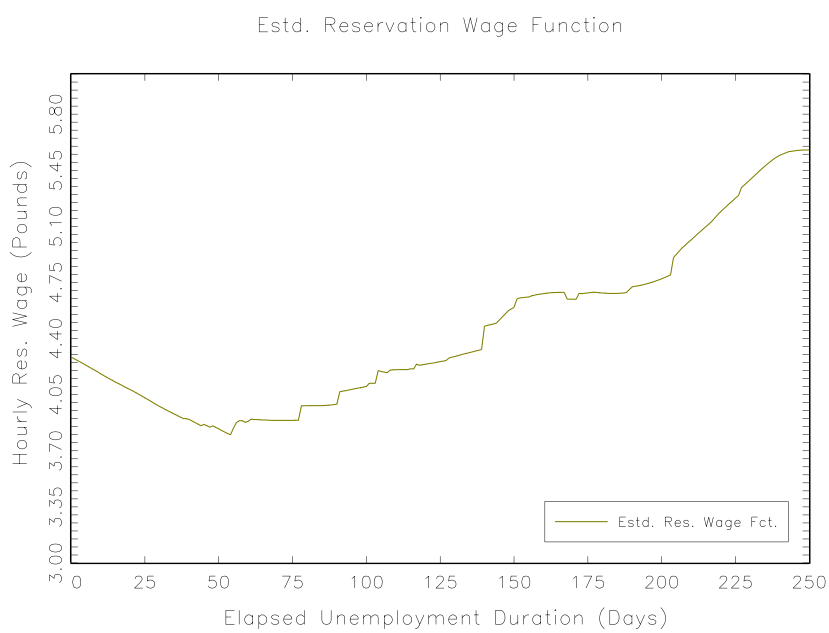
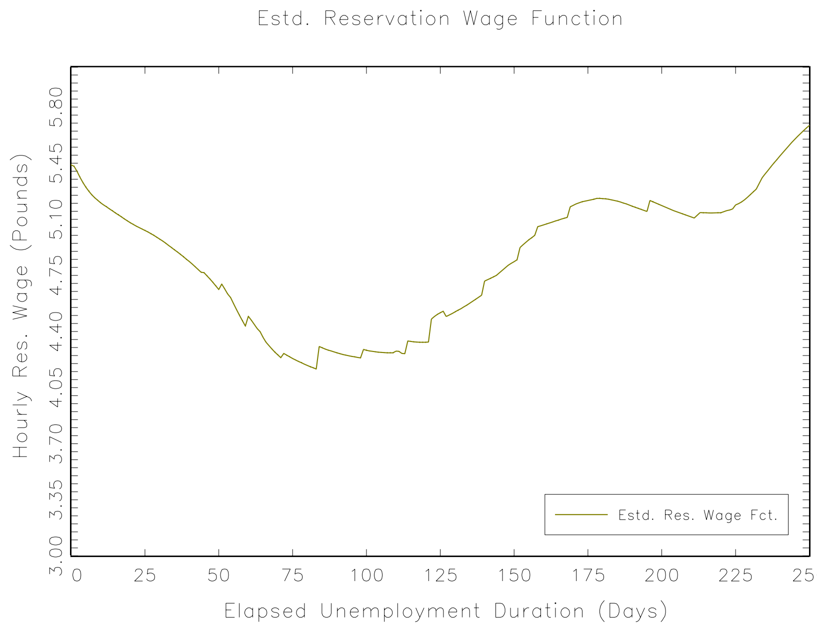


Figure 2: Men - Estimated Reservation Wage Function (contd.)

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



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

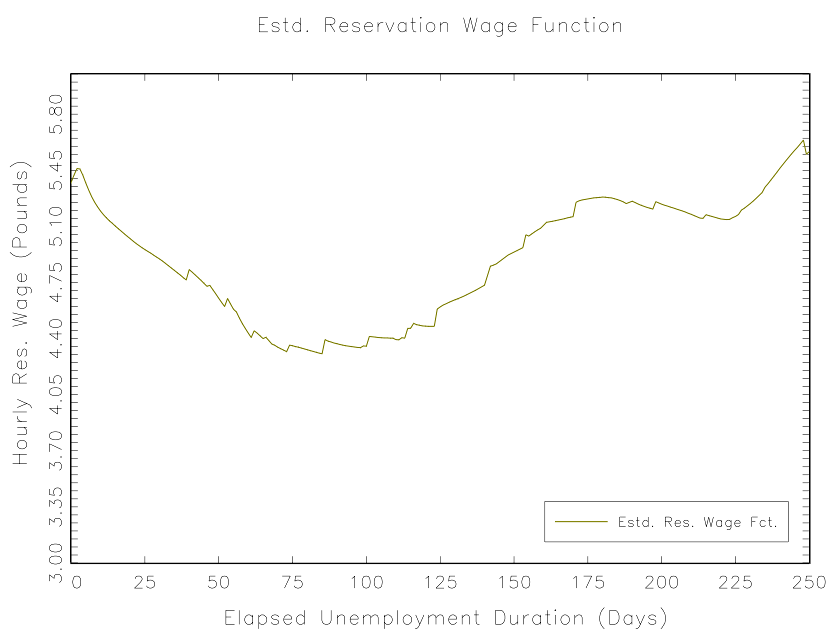
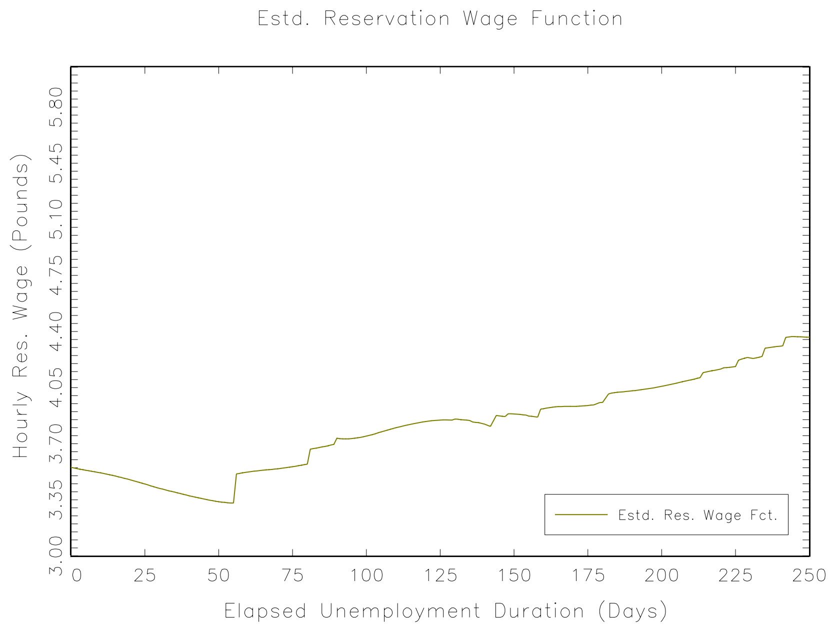


Figure 3: Women - Estimated Reservation Wage Function

(a) 1996 - 2002 Sample



(b) 1996 - 2002 Sample (“Min. Wage Correction”)

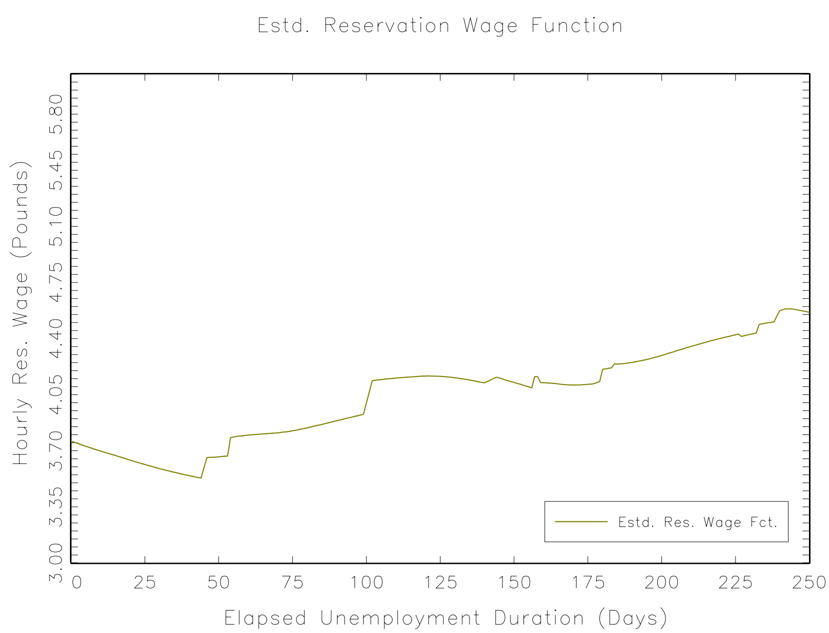
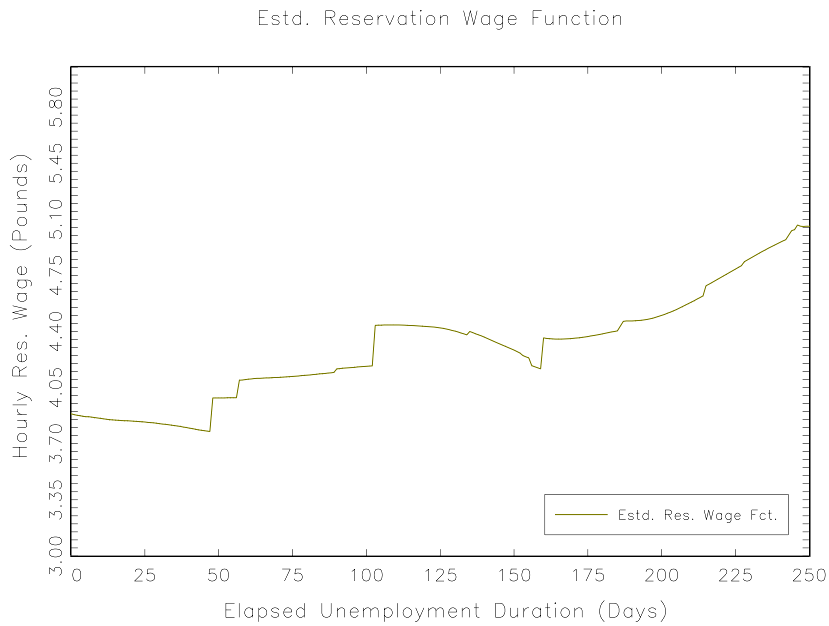
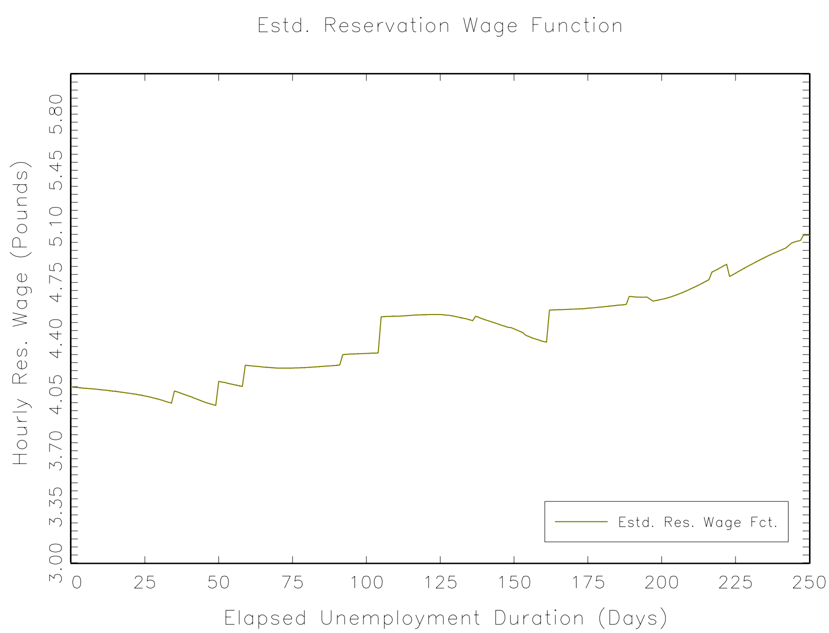


Figure 4: Women - Estimated Reservation Wage Function (contd.)

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



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



References

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.

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.

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

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

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

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

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.

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

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

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

Imbens, G. W., and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica*, 77(5), 1481–1512.

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.

Lancaster, T. (1985): "Simultaneous equations models in applied search theory," *Journal of Econometrics*, 28(1), 113–126.

Lee, S., O. Linton, and Y.-J. Wang (2009): "Testing for Stochastic Monotonicity," *Econometrica*, 77, 585–602.

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.

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., 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.

Sestito, P., and E. Viviano (2011): "Reservation Wages: Explaining Some Puzzling Regional Patterns," *Labour*, 25, 63–88.

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.