*ISSN 1471-0498*

**DEPARTMENT OF ECONOMICS**

**DISCUSSION PAPER SERIES**

# TESTING FOR MONOTONICITY UNDER ENDOGENEITY

# AN APPLICATION TO THE RESERVATION WAGE FUNCTION

**Daniel Gutknecht**

Number 673
September 2013

Manor Road Building, Manor Road, Oxford OX1 3UQ

# Testing for Monotonicity under Endogeneity

## An Application to the Reservation Wage Function

Daniel Gutknecht[1]

University of Oxford

**Abstract**

This paper develops a test for monotonicity of nonparametric regression models under endogeneity, which in its generality is novel in the literature. The test statistic, which is built upon a second order U-process, introduces 'correction terms' based on control functions that purge the endogeneity. The test has a non-standard asymptotic distribution from which asymptotic critical values can directly be derived. Furthermore, the test statistic is extended to accomodate multivariate (exogenous) regressors. Consistency against general alternatives is proved and the test's finite sample properties are examined in a Monte Carlo experiment. The test is used to formally assess the monotonicity of the reservation wage as a declining function of elapsed unemployment duration. This relationship is difficult to measure due to the simultaneity of both variables. Preliminary results indicate that reservation wage functions do in fact not decline monotonically.

**Key-Words:** Control Function, Endogeneity, Reservation Wages, Test for Monotonicity.

**JEL Classification:** C14, C36, C54, J64

[1]Corresponding Address: Department of Economics / Nuffield College, University of Oxford, Manor Road Building, Manor Road, OX1 3UQ Oxford, UK. Email: Daniel.Gutknecht@economics.ox.ac.uk

# 1 Introduction

Tests for monotonicity have been a long standing topic in the statistical and econometric literature as applied economists have often been concerned with determining if a functional relationship between two random variables is monotonic (i.e. increasing or decreasing). In many cases, knowledge about that relationship conveys information about the properties of these variables or even of the underlying economic model itself. In practice, however, many of these functional relationships are difficult to evaluate due to simultaneity in the causal relationship, to omitted variables in the reduced-form model, or to other factors such as measurement error.

This paper develops a novel nonparametric test for monotonicitiy of a regression model that can be applied when the continuous regressor of interest is endogenous. I argue that such a testing framework can be relevant for various setups, for instance in labour economics and consumer theory: the functional relationship between annual hours worked and the hourly wage rate (Vella, 1993), between disposable income and expenditure on goods and services (so called Engel curves), or between the reservation wage and unemployment duration. In all three examples, standard monotonicity tests from the literature fail to be applicable due to the simultaneity of the variables of interest. On the other hand, the ability to formally test for a monotonically declining (or increasing) relationship is of interest because the outcome conveys economic implications about the underlying properties of the variables, e.g. the nature of the good in the Engel curve example.

As mentioned before, this novel test may not only be limited to circumstances in which endogeneity arises due to dual causality, but may also be applied to cases where endogeneity is caused by other factors such as measurement error or omitted variables in the reduced-form model: for instance, treating schooling as a continuous choice variable as in Garen (1984) and considering ability as the omitted variable, a further application could be to test for increasing returns to schooling. Ellison and Ellison (2011) have recently proposed another example from game theory. They examine the monotonic relationship between firms' investment levels and market size in the context of strategic entry deterrence in the pharmaceutical industry: the incumbent's total revenue prior to patent expiration, an available proxy for market size, might be correlated with the incumbent's investment level (see Ellison and Ellison, 2011, p. 14f.). And although the authors discuss restrictions and conditions under which this endogeneity issue could potentially be by-passed, a less restrictive testing solution appears to be desirable in this context.

The main contributions of this paper are as follows: first, I develop a nonparametric test to evaluate the monotonicity of regression models when the continuous regressor of interest is endogenous. In this generality (see below), it is the first paper within a relatively large strand of literature on monotonoicity testing that formally develops a statistical test with this feature. The test is constructed on the basis of a test statistic introduced by Ghosal et al. (2000), but "correction factors" are incorporated into that statistic in order to purge the endogeneity bias: if a suitable instrument (vector) that meets a conditional mean independence assumption exists, these first-stage correction factors can be identified using standard arguments from the control function literature (e.g., Newey et al., 1999; Blundell and Powell, 2003). Moreover, subsequent use of kernel methods allows to estimate these correction terms nonparametrically and to incorporate them into the test statistic.

Second, the above test statistic is extended to accommodate multivariate (exogenous) regressors formally. This case has so far only been considered by Chetverikov (2012) in a non-stochastic setting and is useful if researchers want to control for additional covariates while evaluating the monotonicity of the regression model w.r.t. their variable of interest.

Finally, as an illustration of the test, I examine the monotonicity of the functional relationship between the reservation wage and elapsed unemployment duration. While 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), elapsed unemployment durations refer to the lengths of unemployment spells at the time the reservation wage information is being retrieved. The effect of unemployment duration on the reservation wage is generally ambiguous and difficult to measure because both variables are determined simultaneuosly if reservation wages are flexible (e.g., Kiefer and Neumann, 1979; Lancaster, 1985; Van den Berg, 1990). Moreover, 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 or not. Since, as will be argued in Section 6, knowledge about the monotonicity bears potential implications for underlying job search models, the paper provides a first insight into that question using self-reported hourly reservation wages and unemployment durations from the British Household Panel Survey (BHPS).

From a technical point of view, the derivation of the limiting distribution consists of two stages: the first step involves showing that, under suitable regularity and bandwidth conditions, the test statistic can be approximated by a stationary Gaussian process with continuous sample paths. While most of the proof here relies on results from Ghosal et al. (2000), the main additional step is to show that the bias arising from the nonparametric correction terms is of smaller order asymptotically. In a second step, I prove that the excursion probability of the maximum of that Gaussian process has a non-standard Gumbel distribution (a special case of the generalized extreme value distribution) from which critical values can directly be derived. The proof technique is similar to Lee et al. (2009) in using results from Piterbarg (1996), in particular Theorem G.1 thereof (Piterbarg, 1996, p. 32). As noted in Lee et al. (2009), since the poor quality of the asymptotic approximation is one of the main issues of extreme value limiting distributions (the error declines at a rate that is logarithmic in sample size), I take advantage of a higher order analytic approximation derived in Theorem G.1 that involves including the (known) logarithmic factor in the first-order error. This corrected distribution is closer to the actual distribution and might significantly improve the rate of convergence in finite samples. In fact, a Monte Carlo experiment in Section 5 demonstrates that this gain in power can be substantial in small samples. Finally, the test is shown to be consistent against fixed general alternatives.

Tests for monotonicity of the regression function have been a long-standing topic in the econometric and statistical literature: Bowman et al. (1998) for instance use Silvermans (1981) 'critical bandwidth' approach to construct a bootstrap test for monotonicity, while Gijbels et al. (2000) consider the length or runs of consecutive negative values of observation differences. Hall and Heckman (2000) suggest fitting straight lines through subsequent groups of consecutive points and rejecting monotonicity for too large negative values of the slopes. Other more recent tests include Durot (2003), Wang and Meyer (2011), and Birke and Neumeyer (2013). However, all these tests do require independence between the equation error and the regressor of interest and are thus not applicable to a wide range of economic setups that allow for dependence of those two.

As mentioned before, a notable exception is the paper by Ellison and Ellison (2011, p. 11 ff.), who investigate the problem of detecting strategic entry deterrence in the pharmaceutical industry. The authors propose various monotonicity tests that also allow for the endogenous case, but this case is restricted to monotonically increasing functions and it requires various monotone likelihood ratio assumptions to be satisfied. In an extension, Ellison and Ellison (2011) also mention a possible nonparametric IV approach to testing the monotonicity hypothesis under endogeneity, but no formal theory is developed.

Another test, which has recently been proposed by Chetverikov (2012), encompasses various tests from above and is based on an adaptive procedure for the construction of critical values. The adaptive choice of critical values results in a test that, unlike most of its competitors, is asymptotically non-conservative. The testing procedure has various appealing features such as consistency under conditional heteroscedasticity or the extendability to multivariate regressors, but requires those regressors to be non-stochastic, which again rules out various economic setups that are characterized by simultaneity.

A generalization of the above tests to monotonicity of nonparametric conditional distribution functions has recently been carried out by Lee et al. (2009) and Delgado and Escanciano (2012), where the former has a test statistic that is similar in nature to the one of Ghosal et al. (2000). But 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 that function. Finally, rather than testing for monotonicity of the regression function, Hoderlein et al. (2011) have developed a test for monotonicity of the scalar unobservable in nonparametric regression models, where monotonicity is often crucial as it conveys identifying power.

The paper is organised as follows: Section 2 outlines the main setup and the test statistic when a suitable control function is at hand. Section 3 derives the limiting distribution of the statistic under the null hypothesis, while Section 4 extends the framework to multivariate exogenous regressors. Section 5 examines the finite-sample properties of the estimator in terms of a Monte Carlo study. Section 6 then devotes attention to the reservation wage example, motivating why it is important to address endogeneity and to test for monotonicity in that context. Section 7 concludes. All proofs and tables are relegated to the appendix.

## 2 Setup

To understand the setup, consider the following model: let $W_i$ denote a continuous outcome variable (e.g., the reservation wage), $U_i$ a continuous, endogenous regressor (e.g., elapsed unemployment duration). Then, with $\epsilon_i$ denoting the unobservable error term, the equation of interest is given by:

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

where $m(\cdot)$ is a real-valued function. In Section 4, I will also consider an extension to the case of a (continuous or discrete) exogenous regressor vector $X_i$ in a model of the form $W_i = m(U_i, X_i) + \epsilon_i$, which affects the way the null hypothesis is set up.

Let $\mathcal{U}_n$ denote a compact subset of the support of $U_i$ with strictly positive density everywhere and further properties to be specified in the next section. The interest lies in testing whether $m(\cdot)$ is an increasing (or decreasing) function in $U_i$ on a certain compact interval $\mathcal{T} = [a, b]$ such that $\mathcal{T} \subset \mathcal{U}_n$. For instance, in Section 6, the empirical question to be investigated is whether the reservation wage declines as a function of elapsed unemployment duration for every elapsed duration $t \in \mathcal{T}$. More generally, however, the null hypothesis is stated as:

$$H_0: \quad m(\cdot) \text{ is an increasing (non-decreasing) function on } \mathcal{T}.$$

In principle, the methodology outlined in this and the next section does not need $m(\cdot)$ to be differentiable in its argument $U_i$. However, for illustrative purposes, the derivative is assumed to exist for all $U_i \in \mathcal{U}_n$ in the following and $\nabla^{(1)} m(\cdot)$ denotes the first-order derivative w.r.t. the function's argument. Thus, the null hypothesis from above can be restated as:

$$H_0: \quad \nabla^{(1)} m(t) \geq 0 \quad \text{for all} \quad t \in \mathcal{T}.$$

The alternative is the negation of this null hypothesis. The dependence of $U_i$ and $\epsilon_i$ prevents a direct application of various tests from the literature such as the one developed by Ghosal et al. (2000), which is based on the observed $W_i$ and $U_i$. In the following, I will outline an identification strategy based on control functions that the researcher can apply if she has instrumental variables at hand that satisfy a conditional mean independence and a support condition.

Suppose a $d_z$-dimensional vector of instruments $Z_i$ with $d_z \geq 1$ exists. For simplicity, it is assumed that all components of this vector are continuous and nonconstant. The former condition could, however, be relaxed at the expense of more notation. The reduced-form equation of $U_i$ is given by:

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

where $g(\cdot)$ is a real-valued (non-constant) function satisfying certain smoothness properties that will be specified in the next section. Since $g(\cdot)$ is nonparametric in its form, it is assumed, without loss of generality, that $\mathbb{E}\Big[V_i \Big| Z_i\Big] = 0$ holds. $V_i$ is the so called control function, which satisfies a conditional mean independence condition, namely:

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

This restriction is crucial for identification purposes and referred to as 'exclusion restriction' in the literature. A sufficient condition is independence of the instrument vector $Z_i$ and the model unobservables $\epsilon_i$ and $V_i$. In the example of reservation wages researchers have, for instance suggested unemployment benefits or benefit income other than the former as possible instruments. The validity of these instruments $Z_i$ depends on whether the variables affect the reservation wage only through elapsed unemployment duration, a claim that will be discussed further in Section 6.

Equations (1) and (2) together with the conditional mean independence assumption in (3) and the normalization $\mathbb{E}\left[V_i\middle|Z_i\right] = 0$ characterize a standard nonparametric control function setup for an additive regression function (e.g., Blundell and Powell, 2003). It is assumed that all structural simultaneous equation examples from the introductory Section 1 such as the reservation wage setup can be cast into this reduced-form framework, which requires that the underlying structural functions satisfy a property referred to as "control function separability" (see Blundell and Matzkin (2010) for details).

Then, identification of $m(\cdot)$ can be achieved using, for instance, Theorem 2.3 of Newey et al. (1999). That is, notice that:

$$\begin{aligned}\mathbb{E}\left[W_i\middle|U_i = u, Z_i = z\right] =& m\left(u\right) + \mathbb{E}\left[\epsilon_i\middle|U_i = u, Z_i = z\right] \\ =& m\left(u\right) + \mathbb{E}\left[\epsilon_i\middle|U_i = u, V_i = v = u - g(z)\right] \\ =& m\left(u\right) + \mathbb{E}\left[\epsilon_i\middle|V_i = v\right] \\ \equiv& m\left(u\right) + \lambda(v),\end{aligned} \tag{4}$$

where the third equality follows from (3). Newey et al. (1999) show that identification of the additive $m(\cdot)$ from the above conditional expectation is equivalent to the identification of $m(\cdot)$ in Equation (1) (see p. 567 and Theorem 2.1 of Newey et al. (1999) for details). Under the additional assumptions that $\mathbb{E}\left[V_i\middle|Z_i\right] = 0$ and differentiability of $\lambda(V_i) \equiv \mathbb{E}\left[\epsilon_i\middle|V_i\right]$, the authors show that this result can be used to identify $m(\cdot)$ up to an additive constant (cf. Theorem 2.3). Furthermore, imposing the normalization $\mathbb{E}\left[\epsilon_i\right] = 0$ on $\epsilon_i$ such that $\mathbb{E}\left[W_i\right] = \mathbb{E}\left[m\left(U_i\right)\right]$, identification of the level of $m(\cdot)$ can be accomplished by assuming the existence of a density function $f(v)$ satisfying $\int f(v)dv = 1$, since:

$$\int \mathbb{E}\Big[W_i\Big|U_i = u, V_i = v\Big] f(v)dv = m(u) + \int \lambda(v)f(v)dv = m(u) + 0. \tag{5}$$

In order to use these results for the test statistic developed in this paper, define:

$$\mu(u, v) \equiv \mathbb{E}\Big[W_i\Big|U_i = u, V_i = v\Big] = m(u) + \lambda(v), \tag{6}$$

and:

$$\mu(u) \equiv \int \mathbb{E}\Big[W_i\Big|U_i = u, V_i = v\Big] f(v)dv = m(u) \tag{7}$$

so that $\mu(u, v) - \mu(u) = \lambda(v)$, where $\lambda(v)$ is the correction factor that was introduced informally in Section 1 and whose empirical equivalent will be used in the test statistic below. To construct the estimator of $\lambda(\cdot)$, note that Equation (7) can be consistently estimated using for instance a kernel estimator:

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

where:

$$\widehat{\mu}(u, \widehat{V}_i) = \frac{\frac{1}{nh_{1n}^2}\sum_{j=1}^{n} \widehat{\tau}_{in} W_j \overline{K}_{h1}(u - U_j)\overline{K}_{h1}(\widehat{V}_i - \widehat{V}_j)}{\frac{1}{nh_{1n}^2}\sum_{j=1}^{n} \overline{K}_{h1}(u - U_j)\overline{K}_{h1}(\widehat{V}_i - \widehat{V}_j)} \tag{9}$$

is the empirical counterpart of Equation (6) with the empirical control function $\widehat{V}_i$ being constructed as $\widehat{V}_i \equiv \widehat{V}(U_i, Z_i) = U_i - \widehat{g}(Z_i)$ and:

$$\widehat{g}(Z_i) = \frac{\frac{1}{nh_{1n}^{d_z}}\sum_{j=1}^{n} U_j \overline{K}_{h1}(Z_{i1} - Z_{j1}) \times \ldots \times \overline{K}_{h1}(Z_{id_z} - Z_{jd_z})}{\frac{1}{nh_{1n}^{d_z}}\sum_{j=1}^{n} \overline{K}_{h1}(Z_{i1} - Z_{j1}) \times \ldots \times \overline{K}_{h1}(Z_{id_z} - Z_{jd_z})}.$$

Here, $\overline{K}_{h1}(u) = \overline{K}(u/h_{1,n})$ denotes a one dimensional, p-th order kernel function ($p$ is even) with compact support and $h_{1n} \longrightarrow 0$ as $n \longrightarrow \infty$. For simplicity, it is assumed that $h_{1n} = h_{11n} = \ldots = h_{1d_zn}$. The motivation for utilising a higher order kernel function will be outlined in the next section. Moreover, $\widehat{\tau}_{in} \equiv \tau(\widehat{f}(u, Z_i))$ is a stochastic trimming function that tends to one as $n \longrightarrow \infty$ (again, the reader is referred to the next section for technical details). The need for trimming arises to circumvent the so called "small denominator" problem and to trim away observations with densities too close to zero. On the other hand, to avoid trimming a positive fraction of observations in the limit, which would imply that the empirical equivalent of $\int \lambda(v)f(v)dv$ does not necessarily converge to $\mathbb{E}[\epsilon_i] = 0$, the trimming function tends to one as $n \longrightarrow \infty$.

The original test statistic of Ghosal et al. (2000) is based on the following U-process (indexed by $t \in \mathcal{T}$):

$$\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \text{sign}\Big(W_j - W_i\Big) \, \text{sign}\Big(U_i - U_j\Big) \frac{1}{h_{2,n}^2} K_{h2}(U_i - t) K_{h2}(U_j - t),$$

where $K_{h2}(u) = K(u/h_{2n})$ denotes to a second-order kernel function and $h_{2n} \longrightarrow 0$ as $n \longrightarrow \infty$. The sign$(\cdot)$ function, on the other hand, is defined as:

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

Intuitively, using the observed $W_i$ and $W_j$ in this context leads to a bias since $U_i$ depends on $\epsilon_i$. Instead, a modified and feasible U-process is given by:

$$\widehat{U}_n(t) = \frac{1}{n(n-1)} \sum_{1 \leq i < j \leq n} \Big(\Big(W_j - \widehat{\lambda}(\widehat{V}_j)\Big) - \Big(W_i - \widehat{\lambda}(\widehat{V}_i)\Big)\Big) \, \text{sign}\Big(U_i - U_j\Big) \frac{1}{h_{2,n}^2} K_{h2}(U_i - t) K_{h2}(U_j - t), \tag{10}$$

where $\widehat{\lambda}(\widehat{V}_i) = \widehat{\mu}(U_i, \widehat{V}_i) - \widehat{\mu}(U_i)$ is the empirical equivalent of the aforementioned correction factor (notice that the different scaling of the U-process is important for technical purposes as will become clear from the proof of Theorem 1 in Appendix A1). Loosely speaking, this correction term $\widehat{\lambda}(\widehat{V}_i)$ controls for the bias introduced by the correlation between $U_i$ and $\epsilon_i$. That is, if $\nabla^{(1)} m(t) \geq 0$, $\widehat{U}_n(t)$ should, apart from random fluctuations, be less than or equal to 0. To see this, let $U_n(t)$ denote the U-process for unobserved $\lambda(V_i)$. Then, inserting $W_i = m(U_i) + \epsilon_i$

and taking expectations yields:

$$\mathbb{E}\Big[U_n(t)\Big] = \int\int\Big(m(U_j) - m(U_i)\Big)\,\text{sign}\Big(U_i - U_j\Big)\frac{1}{h_{2,n}^2}K_{h2}(U_i - t)K_{h2}(U_j - t)dF(U_i)dF(U_j)$$
$$+\int\int\Big(\epsilon_j - \lambda(V_j) - \epsilon_i + \lambda(V_i)\Big)\,\text{sign}\Big(U_i - U_j\Big)\frac{1}{h_{2,n}^2}K_{h2}(U_i - t)K_{h2}(U_j - t)dF(U_i, V_i, \epsilon_i)$$
$$\times\, dF(U_j, V_j, \epsilon_j),$$

where $F(\cdot)$ denotes a (joint) distribution function corresponding to a marginal (joint) density function $f(\cdot)$. The second term is zero using iterated expecations and conditional mean independence (recall that $\lambda(V_i) = \mathbb{E}[\epsilon_i|V_i]$). For the first term on the other hand, one obtains by change of variables with $\nu = ((U_j - t)/h_{2,n})$ and $u = ((U_i - t)/h_{2,n})$:

$$\int\int\Big(m(t + h_{2n}\nu) - m(t + h_{2n}u)\Big)\,\text{sign}\Big(u - \nu\Big)K(u)K(\nu)f(t + h_{2n}u)f(t + h_{2n}\nu)dud\nu.$$

Since:

$$\frac{1}{h_{2n}}\Big(m(t + h_{2n}\nu) - m(t + h_{2n}u)\Big) \longrightarrow \nabla^{(1)}m(t)(\nu - u),$$

an application of dominated convergence yields:

$$\frac{1}{h_{2n}}\,\mathbb{E}\Big[U_n(t)\Big] \longrightarrow -\nabla^{(1)}m(t)\int\int|u - \nu|K(u)K(\nu)f(t)^2dud\nu. \tag{11}$$

Thus, the limit is negative or zero if and only if $\nabla^{(1)}m(t) \geq 0$. So, in expectation, the U-process should be less than or equal to zero under $H_0$. Vice versa, under the alternative, it should yield a positive value.

The test statistic is given as the supremum (over 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. However, as pointed out by Ghosal et al. (2000), other functionals might be chosen depending on the specific interest of the researcher. Specifically, the statistic is chosen to be:

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

where $c_n(t)$ is a scaling factor that may depend on $(U_1, \ldots, U_n)$ 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 $c_n(t) = \widehat{\sigma}_n(t)/\sqrt{n}$. This yields:

$$S_n = \sup_{t \in \mathcal{T}} \left\{ \frac{\sqrt{n}\widehat{U}_n(t)}{\widehat{\sigma}_n(t)} \right\}, \tag{12}$$

where

$$\widehat{\sigma}_n^2(t) = \frac{1}{n(n-1)(n-2)} \sum_{1 \le i,j,k \le n, i \ne j \ne k} \Big(\widehat{\epsilon}_i - \widehat{\lambda}(\widehat{V}_i)\Big)^2 \operatorname{sign}\Big(U_i - U_j\Big) \times \operatorname{sign}\Big(U_i - U_k\Big) \frac{1}{h_{2,n}^4} K_{h2}(U_j - t) K_{h2}(U_k - t) K_{h2}(U_i - t)^2 \tag{13}$$

with $\widehat{\epsilon}_i - \widehat{\lambda}(\widehat{V}_i) = (W_i - \widehat{\mu}(U_i)) - (\widehat{\mu}(U_i, \widehat{V}_i) - \widehat{\mu}(U_i)) = W_i - \widehat{\mu}(U_i, \widehat{V}_i)$ being the sample counterpart of:

$$\begin{aligned}\sigma_n^2(t) &= \int \left( \int \Big(\epsilon - \lambda(V)\Big) \operatorname{sign}\Big(U - \omega\Big) \frac{1}{h_{2n}} K_{h2}(\omega - t) dF(\omega) \right)^2 \frac{1}{h_{2n}^2} K_{h2}(U - t)^2 dF(U, V, \epsilon) \\ &= \int \left( \int \int \Big(\epsilon - \lambda(V)\Big)^2 \operatorname{sign}\Big(U - \omega_1\Big) \operatorname{sign}\Big(U - \omega_2\Big) \frac{1}{h_{2n}^2} K_{h2}(\omega_1 - t) \right. \\ &\quad \left. \times K_{h2}(\omega_2 - t) dF(\omega_1) dF(\omega_2) \right) \frac{1}{h_{2n}^2} K_{h2}(U - t)^2 dF(U, V, \epsilon).\end{aligned} \tag{14}$$

$\sigma_n^2(t)$ is the variance of $U_n(t)$ for the case which maximizes the type-I error (see next section). The respective test for Equation (12) is given by:

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

where $\lim_{n \longrightarrow \infty} \mathbb{P}\{S_n > \kappa_{n,\alpha}\} = \alpha$. Thus, to approximate the critical values, the limiting distribution of $S_n$ is required, which is established in the next section.

As a final remark, note that although the test is not consistent in the presence of conditional heteroscedasticity as the one of Chetverikov (2012), it is invariant to strictly monotonic transformations of the regressor $U_i$ provided that $\mathcal{T}$ is subject to the same transformation. Moreover, the test is also invariant to strictly monotonic transformations of the dependent variable $W_i$ provided that the conditional mean independence assumption of Equation (3) is strengthened to full independence between the instrument vector $Z_i$ and the unobservables $V_i$ and $\epsilon_i$. For instance, assuming full independence and letting $\Gamma(\cdot)$ denote a strictly increasing, differentiable transformation function of $W_i$ (e.g., a logarithmic transformation), one can show that $h_{2n}^{-1}\,\mathbb{E}\Big[U_n^*(t)\Big]$ with $U_n^*(t)$ denoting the U-process of the transformed $W_i$ converges to the limit:

$$-\,\mathbb{E}\Big[\nabla^{(1)}\Gamma(m(t)+\epsilon_i)\Big]\nabla^{(1)}m(t)\int\int|u-\nu|K(u)K(\nu)f(t)^2 du d\nu,$$

which is again negative or zero if and only if $\nabla^{(1)}m(t) \geq 0$.

# 3 Large Sample Theory

Before the asymptotic theory is outlined, notice that the type I error is maximized for the case where $\nabla^{(1)}m(t) = 0$ for all $t \in \mathcal{T}$. As for the test of Ghosal et al. (2000), it thus suffices to derive the limiting distribution for the boundary case of a constant function, which implies that the test is asymptotically conservative. The following assumptions are required for the derivation of the limiting distribution:

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

**A2** $U_i$, $\epsilon_i$, and $V_i = U_i - g(Z_i)$ are distributed with continuous joint density everywhere. The boundary of the support of $Z_i$ and $V_i$ is a set of measure zero. Moreover, let $\mathcal{W}_n = \mathcal{U}_n \times \mathcal{Z}_n$ denote a set that is the cartesian product of $d_z + 1$ connected, non-empty, compact intervals, where $\mathcal{U}_n$ and $\mathcal{Z}_n$ lie in the interior of the support of $U_i$ and $Z_i$. Assume that the joint density of $U_i$ and $Z_i$ on $\mathcal{W}_n$ is strictly bounded away from zero.

**A3** The joint density function $f(U_i, Z_i)$ and the real-valued function $g(Z_i)$ from (2) are p times continuously differentiable in their arguments with uniformly bounded derivatives, where $p$ is an even integer number satisfying $p > 3 + d_z$.

**A4** $K(\cdot)$ and $\overline{K}(\cdot)$ are bounded, symmetric functions with compact support on $[-1, 1]$ that are twice continuously differentiable. Moreover, assume that $\overline{K}(\cdot)$ is a p-th order kernel function ($p$ is even and satisfies $p > 3 + d_z$), while $K(\cdot)$ is a second order kernel function.

**A5** Let $V^m$ denote the support of the marginal distribution of $V_i = U_i - g(Z_i)$. Its distribution function is continuously differentiable and it holds that the support of $V_i$ conditional on $t \in \mathcal{T}$ equals $V^m$.

**A6** Let $c_n \longrightarrow 0$ as $n \longrightarrow \infty$. The trimming function $\tau(\cdot)$ is a continuous, twice differentiable function such that $\tau(x) = 0$ if $x < c_n$ and $\tau(x) = 1$ if $x > 2c_n$. The first and second order derivatives of $\tau$ satisfy $\sup_x|\nabla^{(1)}\tau(x)| = O(c_n^{-1})$ and $\sup_x|\nabla^{(2)}\tau(x)| = O(c_n^{-2})$.

The finite fourth moment assumption in A1 is a technical condition that is required for the application of uniform convergence results to kernel estimators on slowly expanding sets derived by Hansen (2008). It does, in addition, ensure that various limit expressions in the appendix exist and are finite.

Condition A2 consists of different parts: on one hand, it requires that $U_i$, $\epsilon_i$, and $V_i$ are jointly continuously distributed. While this is standard (and cannot be relaxed) for the control function $V_i$ and the endogenous regressor $U_i$, it is somewhat more restrictive on $\epsilon_i$ as it rules out various (partially) discrete distributions. However, while continuity is indispensable for the derivation of the limiting distribution of the test, it is worthwhile noting that the test statistic effectively only uses a subset of the data and that the condition could, in principle, be weakened for the estimation of $\widehat{\lambda}(\cdot)$. The boundary of the support of $Z_i$ and $V_i$ being a set of measure zero on the other hand is required for identification (cf. Newey et al., 1999). Finally, the exact rate $c_n$

at which the set $\mathcal{W}_n$ is expanding is specified in Theorem 1 below.

Condition A3 is an assumption on the smoothness of the functions of the auxiliary model $g(Z_i)$ and the joint density $f(U_i, Z_i)$. The degree of smoothness depends on the number of instrumental variables used and, loosely speaking, represents the "price to pay" to avoid an asymptotic bias from the use of nonparametric estimators. That is, condition A3 helps to ensure that replacing the estimated quantity $\widehat{\lambda}(\widehat{V}_i)$ in $\widehat{U}_n(t)$ and $\widehat{\sigma}_n(t)$ by the unknown term $\lambda(V_i)$ does not lead to an asymptotic bias, but to an error of smaller order (uniformly in $t$).

The assumptions in A4 on the second order kernel $K(\cdot)$ are 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 $K(v) = (15/16)(1-v^2)^2\,\mathbb{I}[|v| \leq 1]$, while the conditions on the first stage kernel function $\overline{K}(\cdot)$ are directly related to A3: if $d_z$ was for instance equal to two, a sixth order kernel function would be required at the first stage.

Condition A5 is a so called large support condition on the subset $\mathcal{T}$ and, in essence, represents a condition on the level of variation of the instrument (vector) (see Imbens and Newey, 2009). Even though it is required for technical purposes, notice that the condition only needs to apply to the subset $\mathcal{T}$ on which the test is carried out. Finally, the use of stochastic trimming in A6 has already been motivated in the previous section. The trimming has been adopted from Fermanian and Salanie (2004), who also provide an example for a function satisfying the assumptions outlined in A6.

Next, the asymptotic distribution of the test statistic is established when the null is true. This is carried out in two main steps (the reader is referred to the proof of Theorem 1 in Appendix A1 for explicit references and a more technical outline of these steps):

First of all, under the regularity conditions from above and the bandwidth conditions of Theorem 1 below, one can show that replacing $\widehat{\lambda}(\widehat{V}_i)$ by the unknown $\lambda(V_i)$ in $\widehat{U}_n(t)$ and $\widehat{\sigma}_n(t)$ results in an error of smaller order uniformly in $t \in \mathcal{T}$. More specifically, $\widehat{U}_n(t) = U_n(t) + O_p(n^{-\frac{1}{2}}) = U_n(t)[1 + o_p(1)]$ and $h_{2n}^{-\frac{1}{2}}\widehat{\sigma}_n(t) = h_{2n}^{-\frac{1}{2}}\widetilde{\sigma}_n(t)[1 + o_p(1)] = O_p(1)$ uniformly in $t \in \mathcal{T}$, where $\widetilde{\sigma}_n(t)$ is defined analogously to $U_n(t)$ containing the unobserved $\lambda(V_i)$ and $\lambda(V_j)$ in place of $\widehat{\lambda}(\widehat{V}_i)$

and $\widehat{\lambda}(\widehat{V}_j)$, respectively. Thus, since $h_{2n} \longrightarrow 0$ as $n \longrightarrow \infty$, one can show that $\sqrt{n}\widehat{U}_n(t)/\widehat{\sigma}_n(t)$ behaves asymptotically as $\sqrt{n}U_n(t)/\widetilde{\sigma}_n(t)$. The statistic $\sqrt{n}U_n(t)/\widetilde{\sigma}_n(t)$ is then approximated by a stationary, normalized Gaussian process $\xi_n(t)$ with continuous trajectories building largely on arguments from Ghosal et al. (2000).

In a second step, this approximation is used to derive the limiting behaviour of the maximum of that Gaussian process $\xi_n(t)$, which in turn yields the limiting distribution of the test statistic $S_n$ under the null hypothesis. That is, following Lee et al. (2009), I use asymptotic results of Piterbarg (1996) for Gaussian processes and fields since his limit theorems allow for a higher order analytic approximation that (potentially) outperform limit approximations from standard extreme value theory in finite samples (see discussion below and the Monte Carlo experiments in Section 5).

**Theorem 1.** *Let* $c_n = O(\log(n)^{\frac{1}{d_z+1}})$ *and* $\mathcal{T}_n = [0, (b-a)/h_{2n}]$, *where* $a$ *and* $b$ *have been defined in Section 2. Moreover, let* $\Delta_n$ *be a positive sequence of order* $O(1)$. *Assume that A1 to A6 hold and let the bandwidth sequences satisfy* $h_{2n}\log(n) \longrightarrow 0$, $n^{\frac{1}{2}}h_{2n}^{\frac{3}{2}}\log(n)^{-3} \longrightarrow \infty$, $n^{\frac{1}{2}}h_{2n}^{-1}h_{1n}^{d_z+3}c_n^{-2}\log(n)^{-1} \longrightarrow \infty$, $n^{\frac{1}{2}}h_{1n}^{d_z+1}c_n^{-2}\log(n)^{-1} \longrightarrow \infty$, *and* $n^{\frac{1}{2}}h_{1n}^{p}c_n \longrightarrow \Delta_n$ *with* $p > 3 + d_z$ *as* $n \longrightarrow \infty$. *Then, for any* $x$:

$$\mathbb{P}\Big(l_n(S_n - l_n) < x\Big) = \exp\Big(-\exp\Big(-x - \frac{x^2}{2l_n^2}\Big)\Big)\Big(1 + o(1)\Big)$$

*and*

$$\lim_{n\to\infty} \mathbb{P}\Big(l_n(S_n - l_n) < x\Big) = \exp\Big(-\exp\Big(-x\Big)\Big) \equiv F_\infty(x),$$

*where:*

$$l_n = \left(2\log\left(\frac{\theta_t^{\frac{1}{2}}(b-a)}{2\pi h_{2n}}\right)\right)^{\frac{1}{2}}$$

*with*

$$\theta_t = -\frac{6\int q(v)K(v)^2\nabla^{(1)}K(v)dv + 2\int q(v)^2K(v)\nabla^{(2)}K(v)dv}{\int q(v)^2K(v)^2dv}$$

*and* $\nabla^{(1)}K(\cdot)$ *and* $\nabla^{(2)}K(\cdot)$ *denote the first- and second-derivative of the second order kernel function and* $q(r) = \int \text{sign}\Big(r - \omega\Big)K(\omega)d\omega$.

Notice that the restrictions on the bandwidth sequences in Theorem 1 offer a relatively wide range of possible choices similar to the test introduced by Ghosal et al. (2000). For instance, $h_{2n}$ could be chosen as $h_{2n} = n^{-\delta_1}$ with $\delta_1 < \frac{1}{3}$ or as $h_{2n} = \log(n)^{-\delta_2}$ with $\delta_2 > 1$. Depending on the choice of the second stage bandwidth and the dimension of the instrument vector, $h_{1n}$ may be chosen at an according rate, but always slower than $\log(n)c_n^2/n^{\frac{1}{4}}$.

In practice, the researcher might also consider more data-driven bandwidth choices $\widehat{h}_{1n}$ and $\widehat{h}_{2n}$ (e.g., through cross-validation) if they satisfy the restrictions of Theorem 1, i.e. if $\widehat{h}_{1n}/h_{1n} \xrightarrow{p} 1$ and $\widehat{h}_{2n}/h_{2n} \xrightarrow{p} 1$. Since the test relies strongly on smoothing parameters and, in particular for $h_{2n}$, bandwidths can be selected from a relatively wide range, such data-driven selection mechanisms can be important as over- or under-smoothing at the estimation or the testing stage might result in a loss of power (e.g., small dips in functions may not be detected in the case of over-smoothing). And even though deriving theoretical results for data-driven bandwdiths is beyond the scope of this paper, one would conjecture that, as pointed out by Lee et al. (2009), regularity conditions similar to those in Einmahl and Mason (2005) could yield asymptotic results as in Theorem 1 for $\widehat{h}_{1n}$ and $\widehat{h}_{2n}$. In the illustration of Section 6, however, I use a more plug-in based approach for the bandwidth selection that appears to perform well in that specific empirical setup.

In order to compute the above statistic, $l_n$ needs to be calculated, which in turn requires values for the second stage bandwidth $h_{2,n}$ and $\theta$: since $K(\cdot)$ is supported on $[-1, 1]$, the integrals in $\theta$ can be computed analytically. For the biweight kernel, for instance, one obtains $\theta_t \approx 11.904$, while for the Epanechnikov kernel one obtains $\theta_t \approx 9.975$ (see Lee et al., 2009, p.591 f.).

Finally, note that stating Theorem 1 in a manner similar to Theorem 3.1 in Lee et al. (2009) allows not only to construct the test with asymptotic level $\alpha$, $0 < \alpha < 1$, in the form of:

$$\text{Reject } H_0 \text{ if:} \quad F_\infty(l_n(S_n - l_n)) \geq 1 - \alpha, \tag{15}$$

as in Theorem 4.2 in Ghosal et al. (2000), but also as:

$$\text{Reject } H_0 \text{ if:} \quad F_n(l_n(S_n - l_n)) \geq 1 - \alpha, \tag{16}$$

where for each $n$:

$$F_n(x) = \exp\Big(-\exp\Big(-x - \frac{x^2}{2l_n^2}\Big)\Big).$$

That is, even though the test in Equation (15) has the correct size $\alpha$ asymptotically, extreme value approximations are often known to perform poorly in finite samples. In fact, the Monte Carlo experiments in Section 5 indicate that these approximations can perform poorly even in samples of size $n = 400$. By contrast, higher order analytic approximations as the one in Theorem G.1 of Piterbarg (1996) perform potentially much better, as the rate at which the test converges to the correct asymptotic size $\alpha$ is polynomial with $O(n^{-q})$ for some $q > 0$ rather than $O(\log(n)^{-1})$ for the usual asymptotic approximation in extreme value theory.

Finally, examining the consistency of the test against fixed general alternatives, the following theorem can be obtained:

**Theorem 2.** *Assume that* $n^{\frac{1}{2}} h_{2n}^{\frac{3}{2}} \log(h_{2n}^{-1})^{-3} \longrightarrow \infty$. *Then, if* $\nabla_U m(t) < 0$ *for some* $t \in [a, b]$, *the test in (15) or (16) is consistent at any level* $\alpha$, $0 < \alpha < 1$.

The test achieves consistency against fixed alternatives if the lower bound on the second stage bandwidth $h_{2n}$ is set to the restriction outlined in Theorem 2. Notice also that $\log(h_{2n}^{-1}) = O(\log(n))$ so that the lower bound from Theorem 1 implies the bandwidth restriction here and hence also consistency against general alternatives.

# 4 Exogenous Regressors

In this section, I show that the test from Section 3 can be extended to a framework with additional multivariate exogenous regressors. The multivariate case has recently also been considered by Chetverikov (2012), albeit in a non-stochastic setting, and is relevant for setups where the researcher wants to control for additional covariates. The model that I examine is of the form:

$$W_i = m\Big(U_i, X_i\Big) + \epsilon_i,$$

where $X_i$ is a continuous (or discrete) $d_x$-dimensional vector of regressors with $d_x \geq 1$. The vector $X_i$ is assumed to be independent of $U_i$ and $\epsilon_i$ (notice that the remaining setup is defined as in Section 2). The interest lies again in testing for monotonicity of the function $m(\cdot,\cdot)$ in $U_i$ on a set to be specified below. For instance, recalling the motivating example from the introduction, the researcher could be interested in examining whether reservation wages decline monotonically with (elapsed) unemployment duration for all individuals of a specific age or gender group.

More generally, let $\mathcal{R}$ be a subset of $\mathcal{X}$, where $\mathcal{X}$ is itself a compact subset of the support of $X_i$ in $\mathbb{R}^{d_x}$ satisfying properties that are listed in assumption $A2^*$ below. Then, defining the set $\mathcal{T}$ as in Section 2, the null hypothesis becomes:

$$H_0: \quad \nabla_1^{(1)} m(t,r) \geq 0 \quad \text{for any} \quad t \in \mathcal{T} \quad \text{and} \quad r \equiv (r_1, \ldots, r_{d_x}) \in \mathcal{R},$$

where $\nabla_1^{(1)} m(\cdot,\cdot)$ denotes the first-order derivative w.r.t. the first argument of the function $m(\cdot,\cdot)$. The alternative is that there exists some $t \in \mathcal{T}$ and $r \in \mathcal{R}$ for which indeed $\nabla_1^{(1)} m(t,r) < 0$. The choice of the set $\mathcal{R}$ depends on the empirical analysis and can be selected by the researcher: for instance, if $X_i$ is dicrete as in the empirical example above and the researcher is only interested in testing for monotonicity at a specific point, the set $\mathcal{R}$ can be chosen to be $\mathcal{R} = r_0$ with $r_0 \equiv (r_{01}, \ldots, r_{0d_x})$.

In analogy to Section 2, the test statistic is based on the following U-process indexed by $t$ and $r$:

$$\widehat{\widehat{U}}_n(t,r) = \frac{1}{n(n-1)} \sum_{1 \leq i < j \leq n} \Big( \Big(W_j - \widehat{\lambda}(\widehat{V}_j)\Big) - \Big(W_i - \widehat{\lambda}(\widehat{V}_i)\Big)\Big) \operatorname{sign}\Big(U_i - U_j\Big)$$
$$\times \frac{1}{h_{2n}^2 h_{3n}^{2d_x}} K_{h2}(U_i - t) K_{h2}(U_j - t)\, \mathbf{K}_{h3}^{d_x}(X_i - r)\, \mathbf{K}_{h3}^{d_x}(X_j - r), \quad (17)$$

where $\mathbf{K}_{h3}^{d_x}(X_i - r) \equiv K_{h3}(X_{i1} - r_1) \times \ldots \times K_{h3}(X_{id_x} - r_{d_x})$ denotes a $d_x$-dimensional (second order) product kernel and it is assumed that $h_{3n} = h_{31n} = \ldots = h_{3d_xn}$. The need for an additional smoothing parameter arises because the bias induced by comparing the function $m(\cdot,\cdot)$ at (potentially) different points $X_i$ and $X_j$ does not vanish asymptotically unless a bandwidth sequence $h_{3n}$ satisfying $h_{2n}^{-1}h_{3n} \longrightarrow 0$ is introduced. Thus, if $\overline{U}_n(t,r)$ denotes the U-process with the unknown $\lambda(V_i)$ and $\lambda(V_j)$ and $m(\cdot,\cdot)$ is differentiable in the direction of $U$ and $X$ on $\mathcal{T}$ and $\mathcal{R}$, respectively, it is straightforward to show that if $h_{3n}h_{2n}^{-1} \longrightarrow 0$, $h_{2n}^{-1}\,\mathbb{E}\Big[\overline{U}_n(t,r)\Big]$ converges to:

$$\frac{1}{h_{2n}}\,\mathbb{E}\Big[\overline{U}_n(t,r)\Big] \longrightarrow -\nabla_1^{(1)}m(t,r)\int\int |u_1-\nu_1|K(u_1)K(\nu_1)du_1d\nu_1 f(t)^2 f(r)^{2d_x},$$

which is less than or equal to zero under the null hypothesis and condition $A2^*$ below. The test statistic itself is given by:

$$\overline{S}_n = \sup_{(t,r)\in\mathcal{T}\times\mathcal{R}}\left\{\frac{\sqrt{n}\widehat{\overline{U}}_n(t,r)}{\widehat{\overline{\sigma}}_n(t,r)}\right\}, \tag{18}$$

where:

$$\begin{aligned}\widehat{\overline{\sigma}}_n^2(t,r) = &\frac{1}{n(n-1)(n-2)}\sum_{1\le i,j,k\le n, i\neq j\neq k}\Big(\widehat{\epsilon}_i - \widehat{\lambda}(\widehat{V}_i)\Big)^2\operatorname{sign}\Big(U_i-U_j\Big)\operatorname{sign}\Big(U_i-U_k\Big)\\ &\times\frac{1}{h_{2n}^4h_{3n}^{4d_x}}K_{h2}(U_j-t)K_{h2}(U_k-t)K_{h2}(U_i-t)^2\,\mathbf{K}_{h3}^{d_x}(X_j-r)\,\mathbf{K}_{h3}^{d_x}(X_k-r)\,\mathbf{K}_{h3}^{d_x}(X_i-r)^2,\end{aligned} \tag{19}$$

is the sample analogue of:

$$\begin{aligned}\overline{\sigma}_n^2(t,r) = &\int\int\left(\int\int\Big(\epsilon-\lambda(V)\Big)\operatorname{sign}\Big(U-\omega_1\Big)\frac{1}{h_{2n}h_{3n}^{d_x}}K_{h2}(\omega_1-t)\,\mathbf{K}_{h3}^{d_x}(\omega_2-r)dF(\omega_1)dF(\omega_2)\right)^2\\ &\times\frac{1}{h_{2n}^2h_{3n}^{2d_x}}K_{h2}(U-t)^2\,\mathbf{K}_{h3}^{d_x}(X-r)^2dF(U,V,\epsilon)dF(X).\end{aligned} \tag{20}$$

Before outlining the asymptotic results, I replace condition A2 from Section 3 by a straightforward extension to the multivariate case. For simplicity, I consider the case where $X_i$ contains continuous elements only:

**A2*** $U_i$, $X_i$, $\epsilon_i$, and $V_i = U_i - g(Z_i)$ are distributed with continuous joint density everywhere. The boundary of the support of $Z_i$ and $V_i$ is a set of measure zero. Moreover, let $\overline{\mathcal{W}}_n = \mathcal{U}_n \times \mathcal{Z}_n \times \mathcal{X}$ denote a set that is the cartesian product of $d_z + d_x + 1$ connected, non-empty, compact intervals, where $\mathcal{U}_n$, $\mathcal{Z}_n$, and $\mathcal{X}$ lie in the interior of the support of $U_i$, $Z_i$, and $X_i$. Assume that the joint density of $U_i$, $Z_i$, and $X_i$ on $\overline{\mathcal{W}}_n$ is strictly bounded away from zero.

Further to the above condition, I will assume, without loss of generality, that $\mathcal{R} = [0,1] \times \ldots \times [0,1] = [0,1]^{d_x}$ is normalized to a $d_x$ dimensional hypercube of side length one. Likewise, I set $(b-a) = 1$ so that $\mathcal{T}_n$ from Section 3 becomes $\mathcal{T}_n = [0, h_{2n}^{-1}]$. Both are mere normalizations that do not bear any consequences for the technical results in Theorem 3 below as the latter can be straightforwardly adapted to the case where these normalizations do not apply (simply replace $h_{2n}^{-1}$ and $h_{3n}^{-1}$ by $(b-a)h_{2n}^{-1}$ and $(b-a)h_{3n}^{-1}$ in the constant $\bar{l}_n$ in Theorem 3 below). Also note that the asymptotic distribution under the null hypothesis is, as in Section 3, derived for the least favourable case $\nabla_1^{(1)} m(t,r) = 0$ for all $t \in \mathcal{T}$ and $r \in \mathcal{R}$, which implies that the test is conservative.

**Theorem 3.** *Let $\overline{S}_n$ be defined as in Equation (18) and denote $\mathcal{R}_n = [0, h_{3n}^{-1}]^{d_x}$, while $\mathcal{T}_n = [0, h_{2n}^{-1}]$. Moreover, let $\Delta_n$ be a positive sequence of order $O(1)$. Assume that A1, A2*, A3, A4, A5, and A6 hold and let the bandwidth sequences satisfy $h_{2n}\log(n) \longrightarrow 0$, $n(h_{2n}h_3^{d_x})^{\frac{(d_x+3)}{(d_x+1)}} \times \log(n)^{-2(d_x+3)} \longrightarrow \infty$, $nh_{2n}^3 h_{3n}^{3d_x}\log(n)^{-1} \longrightarrow \infty$, $h_{2n}^{-1}h_{3n} \longrightarrow 0$, $nh_{2n}^{-2}h_{3n}^{-2d_x}h_{1n}^{2(d_z+3)} \times c_n^{-4}\log(n)^{-2} \longrightarrow \infty$, $nh_{1n}^{2(d_z+1)}c_n^{-4}\log(n)^{-2} \longrightarrow \infty$, and $nh_{1n}^{2p}c_n^2 \longrightarrow \Delta_n$ with $p > 3 + d_z$ as $n \longrightarrow \infty$. Then, for any $x$:*

$$\mathbb{P}\Big(\bar{l}_n(\overline{S}_n - \bar{l}_n) < x\Big) = \exp\Big(-\exp\Big(-x - \frac{x^2}{2\bar{l}_n^2}\Big)\Big)\Big(1 + \frac{x}{\bar{l}_n^2}\Big)^{d_x} + o(1)$$

*and*

$$\lim_{n\to\infty} \mathbb{P}\Big(\bar{l}_n(\overline{S}_n - \bar{l}_n) < x\Big) = \exp\Big(-\exp\Big(-x\Big)\Big),$$

*where $\bar{l}_n$ is the largest solution to:*

$$h_{2n}^{-1} h_{3n}^{-d_x} 2^{-d_x+1} \left(\frac{\theta_t}{2}\right)^{\frac{1}{2}} \frac{\theta_r^{\frac{d_x}{2}}}{\pi^{\frac{3}{2}+\frac{d_x}{2}}} \bar{l}_n^{d_x} \exp\left(-\frac{\bar{l}_n^2}{2}\right) = 1$$

*with*

$$\theta_t = -\frac{6\int q(v)K(v)^2\nabla^{(1)}K(v)dv + 2\int q(v)^2K(v)\nabla^{(2)}K(v)dv}{\int q(v)^2K(v)^2dv}$$

*and*

$$\theta_r = -\frac{\int K(v)\nabla^{(2)}K(v)dv}{\int K(v)^2dv}.$$

Since the leading term of $\bar{l}_n$ can be shown to be of order $O(h_{3n})$, an analogous result to Theorem 2 yields that the above test is consistent at any level $\alpha$ against fixed alternatives provided that $nh_{2n}^3 h_{3n}^{3d_x} \log((h_{3n}^{-1})^{-1} \longrightarrow \infty$, which is implied by the bandwidth conditions of Theorem 3. More generally, however, the rather lengthy and complicated restriction on the bandwidth are owed to the need of controlling for three different smoothing parameters simultaneously. If $d_x = 1$, possible rates for the bandwidth sequences $h_{2n}$ and $h_{3n}$ include, for instance, $h_{3n} = n^{-\delta_3}$ and $h_{2n} = \log(n)^{\frac{1}{3}} n^{-\delta_3}$ with $\delta_3 < \frac{1}{6}$. Thus, aside from issues of data availability, it becomes clear that a completely nonparametric test as the one in Theorem 3 may suffer from the curse of dimensionality if $X_i$ is of higher dimension, which could in turn lead to a test with low power. The constant $\theta_r$, on the other hand, can be straightforwardly determined and computed analytically as $\theta_t$ in Section 3. It takes the value $\theta_r = 3$ for the Biweight and $\theta_r = 1.5$ for the Epanechnikov kernel function.

Finally, note that by using high-level conditions similar to Assumption 3.1 (f) and (g) in Lee et al. (2009), one could, as in their paper, also allow for a generalization to unobserved regressors of the form $X_i = \phi(E_i, \beta_0)$ of which only an estimate $\widehat{X}_i = \phi(E_i, \widehat{\beta})$ exists. In such a setup, $\phi(\cdot,\cdot)$ is a known function, while $E_i$ and $\beta_0$ denote some finite dimensional regressor and parameter vectors, respectively, and $\beta_0$ can be estimated by a square root n consistent estimator $\widehat{\beta}$. Then, mimicing the proofs of Lemma A.6 and Lemma A.7 in the supplementary material of Lee et al. (2009), one can show that the same limit results as in Theorem 3 hold.

# 5 Monte Carlo Experiments

This section describes various Monte Carlo experiments that have been carried out to assess the performance of the test in small samples. Firstly, I examine the behaviour of the test when the null hypothesis is true. Since the test under investigation is asymptotically conservative, the model is chosen to be the boundary case of a constant function (M1):

$$W_i = 0 + \epsilon_i.$$

The regressor $U_i$ is constructed as $U_i = Z_i + V_i$. In order to ensure that the large support assumption on the control function $V_i$ from A5 is satisfied, I proceed in two steps: first, I sample the instrumental variable $Z_i$ and the control function $V_i$ from two uniform distributions supported on $[0, 1]$ and $[-1, 1]$, respectively. Since for $0 \leq u \leq 1$, the conditional support of $V_i$ is full, i.e. $[-1, 1]$, I only retain observations of $U_i$ that lie in the interval $[0, 1]$ in a second step. The unobservable $\epsilon_i$ is given by: $\epsilon_i = V_i + 0.1 \cdot \varpi_i$, with $\varpi_i \sim N(0, 0.1^2)$.

The kernel function is chosen to be the Epanechnikov kernel $K(v) = 0.75(1 - v^2)\,\mathbb{I}[|v| \leq 1]$ for the estimation as well as the testing stage. This choice, albeit theoretically questionable since the first-stage requires, in principle, the use of a sixth order kernel function if $d_z = 1$, is motivated by the frequently made observation that in many applications the resulting estimators often tend to have inferior finite sample properties compared to those based on standard kernel functions (e.g., Jones and Signorini, 1997).

The functions $g(\cdot)$ from Equation (2) and $\mu(\cdot)$ from Equation (7) are estimated using the Nadaraya-Watson estimator as explained in Section 2 and the trimming function $\tau(\cdot)$ is set equal to one for all observations in line with most of the literature. The test statistic is constructed as in Equation (12) with the interval $\mathcal{T}$ chosen to be $\mathcal{T} = \{0.05, 0.1, \ldots, 0.9, 0.95\}$. The sample sizes are $n = 200, 300, 400$ and $1,500$ replications are carried out for each simulation.

Since the test relies on smoothing parameters, which may render the power susceptible to variations in these parameters, I evaluate the test's performance under different bandwidth regimes: firstly, I consider a regime where first- and second-stage bandwidth are set at the same rate, i.e. $h_{1n} = h_{2n} = 0.8n^{-\frac{1}{4}}$. Notice that while $h_{2n}$ is in compliance with Theorem 1, $h_{1n}$ is slightly faster than required by Theorem 1, which postulates that $h_{1n}$ has to be of order $n^{-\delta_4}$ with $\delta_4 > \frac{1}{4}$ in case of $d_z = 1$. However, unreported results for $h_{1n}$ (available upon request) show that slightly slower bandwidth choices lead to results of a similar order of magnitude. The second bandwidth choice is $h_{1n} = 0.9n^{-\frac{1}{4}}$ and $h_{2n} = 0.9n^{-\frac{7}{24}}$, where $\frac{7}{24} = \left(\frac{1}{3} + \frac{1}{4}\right)/2$ is slightly larger than $\frac{1}{4}$ and hence faster than before. Finally, in view of the illustration in Section 6, in a third specification I select $h_{2n} = 0.8n^{-\frac{1}{4}}$ for the second-stage bandwidth, but use a plug-in choice $h_{1n} = 2.34\,\mathrm{sd}(\cdot)n^{-\frac{1}{5}}$ for the first-stage, where $\mathrm{sd}(\cdot)$ denotes the sample standard deviation. The choice for $h_{1n}$ is a rule-of-thumb bandwidth for kernel density estimators when using the Epanechnikov kernel (e.g., Hansen, 2012, p. 292)

Turning to the results of Table 5 in Appendix A2 and starting with the empirical size when using critical values from the asymptotic expansion $F_n$, the column labelled M1 reveals that rejection levels are somewhat heterogenous across the choices of bandwidths with the second specification with $h_{1n} = 0.9n^{-\frac{1}{4}}$ and $h_{2n} = 0.9n^{-\frac{7}{24}}$ outperforming the other two significantly as its rejections levels lie closest to the nominal level of 5% for all sample sizes. Moreover, notice the substantial difference in rejection probabilities between using critical values from the asymptotic expansion $F_n$ and from the extreme value distribution $F_\infty$: while rejections probabilities in the former case lie above 2% throughout, they are nowhere close to the 5% level in the latter case.

Next, I examine the test's behaviour when the null hypothesis is false. As Ghosal et al. (2000), I examine three different functions that violate the null hypothesis. The models are similar in nature to the functions considered by Ghosal et al. (2000) and can be viewed in Figure 1 of Appendix A2.

The first model violating the null (M2) takes the form:

$$W_i = 1.5 \cdot U_i(1 - U_i) + \epsilon_i$$

with the variables $U_i$ and $\epsilon_i$ being generated as before. The function $U_i(1 - U_i)$ is a standard non-monotonic function with a turning point at 0.5. Similar to before, one observes again that, when using critical values from the asymptotic expansion, rejection probabilities lie significantly above the ones when using instead the asymptotic distribution, with the latter being below 50% throughout. Moreover, unlike in the previous case, notice that rejection levels are somewhat more homogenous than before, albeit these levels do not reach the ones of the remaining two cases that violate $H_0$.

The next model (M3), which fails again the null hypothesis, is taking the form:

$$W_i = U_i + \exp(-50 \cdot U_i^2) + \epsilon_i.$$

As displayed in Figure 1, the function is steadily upwards sloping after an initial slump. Unlike in the case of M2, note that, already at $n = 200$, rejection rates start off from levels that lie much above the ones from before. These rejection rates rise steadily with sample size across all bandwidth specifications reaching levels between 91% and 95% for critical values from the asymptotic expansion and 77% to 80% for critical values from the extreme value distribution.

Finally, the third model violating the null hypothesis (M4) is given by:

$$W_i = \begin{cases} 10 \cdot (U_i - 0.5)^3 - 1.5\exp(-100 \cdot (U_i - 0.25)^2) + \epsilon_i, & \text{if} \quad U_i < 0.5 \\ 0.1 \cdot (U_i - 0.5) - 1.5\exp(-100 \cdot (U_i - 0.25)^2) + \epsilon_i, & \text{if} \quad U_i \geq 0.5 \end{cases},$$

it exhibits a dip at $U_i = 0.25$ and a relatively flat portion on $U_i > 0.5$ (see Figure 1). From the start, rejection rates are high across all bandwidth specifications. This outcome holds when using critical values from the asymptotic expansion $F_n$ as well as when using critical values from $F_\infty$. At $n = 400$, the rejection probability has reached one throughout.

In summary, this small simulation study provided first insights into the small sample properties of the test statistic for specific bandwidth choices. The findings suggest that the performance of this multi-stage nonparametric test improves substantially for larger sample sizes when the null hypothesis is violated, but gains in empirical size are rather small when considering the boundary case of a constant function. Furthermore, one observes a large discrepancy between rejection levels when using critical from the asymptotic approximation $F_n$ and those when using the asymptotic distribution $F_\infty$: while results were relatively satisfactory in the fomer case, they were rather poor in the latter case. This suggests that particularly for small samples it is advisable to use critical values from the asymptotic expansion rather than from the asymptotic distribution.

# 6 Empirical Illustration

As an illustration of the test, I examine the nonparametric functional relationship between the hourly reservation wage $W_i$ and weekly elapsed unemployment duration $U_i$. The latter refers to the length of an unemployment spell at the time when the reservation wage information is being retrieved.

For many decades, reservation wages have been the focal point of interest of labour economists since they play an important role in modern job search theory and are viewed as a key determinant for the length of unemployment. Nevertheless, the effect of elapsed unemployment duration on the reservation wage is generally ambiguous and difficult to measure because both variables are determined simultaneuosly when reservation wages are flexible (e.g., Lancaster, 1985, p. 113f.): for instance, supposing that reservation wages are a deterministic, decreasing function of time, for a randomly selected sample from a pool of unemployed (stock sampling), it should hold that those with longer (elapsed) unemployment spells have lower reservation wages. On the other hand, assuming that people exit the pool of unemployed on the first occasion they receive a random wage offer above their reservation wage, a person with a long elapsed duration is likely to have been using a relatively high reservation wage as she would have already exited the pool of unemployed otherwise. This illustrates the duality of the setup since the observed (elapsed) duration is both, (elapsed) time as well as a realisation of the random (elapsed) duration. Potential causes of this duality are, for instance, unobserved heterogeneity or other omitted factors that enter standard job search models.

Numerous papers have assessed the impact of elapsed 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 et al., 2004; Brown and Taylor, 2009). However, 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 or not. This is in contrast to various job search models that go beyond the sole prediction of (on average) declining reservation wages and also make predictions about the nature of this decline. For instance, Mortensen (1986) develops a model where (known) credit constraints lead to a monotonically declining reservation wage since the agent anticipates the decrease in means to fund further search. Likewise, in another model with anticipation, Van den Berg (1990) shows that foreseen reductions in unemployment benefit payments after a certain period (e.g., due to exhaustion or regime switch) may lead to a similar outcome. In the same paper, however, the author also demonstrates that the shape of the reservation wage function crucially hinges on the functional form of the job offer arrival rate, of the wage offer distribution function, and of other exogenous variables that determine the reservation wage. That is, if at least one of these exogenous variables declines (and the others remain constant), monotonicity holds. Thus, testing for monotonicity of the functional relationship between the reservation wage and elapsed unemployment duration not only allows to further evaluate the validity of these models, but, in some cases, even allows to indirectly infer about the evolution of their exogenous driving factors over time (as rejecting monotonicity would imply rejecting the last set of assumptions).

The aim of this Section is (i) to provide a first insight into the nonparametric functional relationship of the hourly reservation wage $W_i$ and elapsed unemployment duration $U_i$ by estimating a reduced-form reservation wage function nonparametrically and (ii) to demonstrate the test's applicability by testing for monotonicity of the reservation wage function over a range of (elapsed) durations $\mathcal{T}$. As instruments $Z_i$, I select (individual-level) weekly unemployment benefits and (household-level) other weekly benefit income. As has been argued in the literature (e.g., Kiefer and Neumann, 1979; Addison et al., 2004; Brown and Taylor, 2009), the validity of these instrumental variables relies on the assumption that they affect the level of the hourly reservation wage $W_i$ only through elapsed duration $U_i$, which seems plausible if one assumes that the instrumental variables affect only the costs of unemployment.

The unemployment data in this empirical illustration stems from the BHPS. This survey is a nationally representative annual survey on individuals from more than 5,000 households in the UK. It asks 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 October 1996 coinciding with the introduction of jobseeker's allowance in the UK. The choice of December 2005 as sampling endpoint is based on Smith (2011), who, using unemployment data from the BHPS, finds that flow rates into and out of unemployment are relatively stable during this period.

The hourly reservation wage $W_i$ is measured in pounds and constructed combining answers from 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 subsequently includes all individuals of working age (16-65) have indicated such an hourly wage and who satisfy the rationality condition, which requires a reservation wage below the reported expected wage. Notice that individuals who indicated to be economically inactive are also included in the sample if they report a valid reservation and expected wage, as well as a strictly positive household benefit income. These conditions applied to 88% of the "economically inactive" in the data. The decision to include this group of individuals 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).

After removing outliers, the full sample comprises 1,921 spells of which a majority of over 90% are single ones. Summary statistics for the sample can be found in Table 2 of Appendix A2.

Mean and standard deviation of the hourly reservation wage are 4.98 and 2.55 pounds, with the (unreported) difference in means between males and females being 0.47 pounds. Elapsed durations, which are measured in weeks, have a mean of 19.15 weeks, while completed durations have a mean of 59.06 weeks. As expected, the latter is higher for females with 67.45 weeks than for males with 51.91 weeks (withouth table). An even larger difference can be observed between individuals that were considered to be out of the labour force (mean: 107.86 weeks) as opposed to those being unemployed (mean: 41.04 weeks), albeit no real difference in the mean hourly reservation wage exists (the mean difference in completed durations is partially also explained by the fact that 73% of all "economically inactive" agents in the sample are female).

In the following, I consider three different (sub-)samples: a sample containing (i) all individuals (1,921 observations - Sample I), (ii) men only (1,038 observations - labelled Sample II), and (iii) people who indicated to be unemployed rather than to be economically inactive only (1,403 observations - Sample III). Since the last two cases are considered separately (and not as a setup of the form $W_i = m(U_i, X_i) + \epsilon_i$ with $X_i$ equal to a point $x_0$), I will carry out the test on a model of the form:

$$W_i = m(U_i) + \epsilon_i.$$

To illustrate the shape of this function $m(\cdot)$, I estimate that function for each (sub-)sample using the Nadaraya-Watson estimator with the Epanechnikov kernel and the first-stage bandwidth sequence set according to the rule of thumb $2.34 \cdot \mathrm{sd}(\cdot) \cdot n^{-\frac{1}{5}}$ for density estimators as in the last specification of Section 5. Notice that this rate is in compliance with the restrictions of Theorems 1 and 2 since the second-stage bandwidth sequence $h_{2n}$ for the test is set sufficiently fast (see below). Finally, the trimming function is again set equal to one for all observations. The plots in Figures 2, 3 and 4 in Appendix A2 display $\widehat{\mu}(\cdot) = \widehat{m}(\cdot)$ from Equation (8) for the first 22 weeks of elapsed unemployment duration for the three different sample specifications. Pointwise 95% confidence intervals have been plotted around the curves, which have been constructed using the 'm out of n' bootstrap (200 replications).

Starting with the sample containing all observations (Figure 2), one observes a sharp dip in the first 7 to 8 weeks, followed by a moderate increase to roughly 4.70 pounds per hour at 22 weeks. However, while the pointwise confidence bands are relatively tight around the upwards sloping part of the curve, they are quite loose at the beginning casting some doubt on the significance of this initial dip. In fact, while the upwards movement is clearly visible also in Figures 3 and 4 (albeit much less pronounced in the latter case), the dip turns more into a bump when examining the curve from Sample III. A look at the underlying data reveals that a substantial fraction of observations (roughly 10%) exits unemployment after 8 to 9 weeks (with significantly less individuals leaving unemployment in the weeks before and after) suggesting that some of these spells might have been temporary from the outset, either because unemployment was seasonal or because these people already had an onward employment at the time their unemployment started. This, in turn, could be a possible explanation for the initial slump as one would typically expect individuals who know their unemployment is only of temporary nature to have higher reservation wages than those who do not expect their unemployment to end soon.

The null hypothesis to be tested in the following is $H_0 : \nabla^{(1)} m(t) \leq 0 \quad \text{for all} \quad t \in \mathcal{T}_1$ and $t \in \mathcal{T}_2$, where $\mathcal{T}_1 = [1, 7]$ and $\mathcal{T}_2 = [1, 22]$. That is, for all three samples, I test for a declining reservation wage function on these two intervals (in two separate tests). For the test, both intervals are discretized into 20 equi-spaced points. Estimator and kernel function are as described in Section 5, and the bandwidth for the test is set to be $h_{2n} = \text{sd}(\cdot) \cdot n^{-\frac{1}{4}}$. The latter choice has been made in absence of an "optimal" bandwidth theory for the test, but is at least partially supported by reasonable test results below. Moreover, to check the test's robustness against modifications of the latter, I evaluate the test's performance for $C \cdot h_{2n}$ with $C \in \{0.5, 0.75, 1\}$ when testing on $\mathcal{T}_1$, and $C \in \{0.75, 1, 1.25\}$ when testing on $\mathcal{T}_2$. Notice that the reason for choosing slightly different norming constants in the first case is motivated by the observation that critical values derived from the asymptotic expansion $F_n$ are quite distant from those of the asymptotic distribution $F_\infty$ when choosing norming constants that are too big, which suggests that the former might not be a very accurate approximation in cases where the bandwidth is large relative to the width of the interval.

The test results are displayed in Tables 3 and 4 of Appendix A2. As expected from the graphs, one observes no rejection of the null hypothesis of a declining reservation wage function on the sub-interval $\mathcal{T}_1$, but values of the test statistic from Sample III get somewhat close to the 10% critical values (at least when using critical values from the asymptotic expansion $F_n$). The conclusions are stable across all norming constants as well as across both types of critical values (i.e., derived from either the asymptotic expansion or the asymptotic distribution). Turning to the test results for $\mathcal{T}_2$ in Table 4, a very different picture is obtained: declining reservation wages are rejected for most specifications at 1% and 5% levels. A notable exception is the case where $C = 0.75$ for Sample III, where no rejection even at a 10% level is achieved.

In summary, this section has provided a first, tentative insight into the nonparametric behaviour of the reservation wage as a function of elapsed unemployment duration: even though these results have to be taken with care as many important factors or labour market reforms have been ignored (e.g., the introduction of a national minimum wage in 1999 or Working (Family) Tax Credit reforms), the preliminary evidence of this section suggests that the behaviour of reservation wages appears to be more complicated than typically assumed in applied studies using linear estimation techniques. For instance, a linear two-stage least squares regression using Sample I with the instrumental variables from above yields an estimated coefficient for elapsed unemployment duration of $-0.068$ with a t-statistic of $|t| = -4.95$. This estimate, obtained under the implicit assumption of monotonicity, does not convey any information about the fluctuating behaviour observed in Figure 2 and supported by the formal test results.

# 7 Conclusion

This paper proposes a test for monotonicity of a nonparametric regression model when the (continuous) regressor of interest is endogenous. I argue that this kind of test is relevant for various empirical setups. Technically, the test is an "endogeneity-bias" corrected version of the test statistic of Ghosal et al. (2000) that purges the endogeneity using suitable control functions that satisfy a conditional mean independence condition. The asymptotic distribution of this test statistic is shown to belong to a non-standard Gumbel distribution from which critical values can directly be derived. Moreover, using results from Piterbarg (1996), I show that a similar approximation (with a different norming constant) can be derived for the test at hand. A Monte Carlo experiment demonstrates that this approximation provides critical values with a superior finite sample performance. Finally, the test is extended to accomodate multidimensional exogenous regressors, which can be of interest if the researcher wants to control for additional factors.

As an illustration, the paper studies the behaviour of hourly reservation wages as a function of weekly 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 instruments suggested by the literature, test results indicate that reservation wage functions do in fact not seem to decline monotonically during the first 22 weeks of unemployment. This finding is robust across different sample specifications.

There are several topics that are left for future research. Most importantly, the paper has not considered an "optimal" choice of the bandwidths used in the test statistic as the asymptotic and the consistency results do not distinguish different bandwidths $h_{1n}$ and $h_{2n}$ provided they satisfy broad conditions on the rates of convergence outlined in Sections 3 and 4. As mentioned in the main text, over- as well as under-smoothing at both stages ($h_{1n}$ and $h_{2n}$) might lead to a loss of power and this problem is likely to be exacerbated in the multivariate case. Thus, despite a relatively good performance of the test for the selected bandwidths in the Monte Carlo experiment and the illustration, deriving tighter conditions on the bandwidth sequences and a more data-driven bandwidth choice are important topics for future research.

From an empirical perspective, it would be desirable to explore the preliminary results on the non-monotonic behaviour of the reservations wages further. For instance, many important controls that have been left unconsidered in this entirely nonparametric setting could be re-examined in a more parameterized setup. Likewise, to obtain a more robust picture, the tests of Section 6 would also need to be carried out for various subsamples in order to avoid any interference with effects from labour market reforms that potentially affected the pool of unemployed and their reservation wages.

# Appendix A1

**Notation.** For any real-valued function $f$, let $\nabla_j^{(i)} f$ denote the i-th order derivative w.r.t. the j-th argument of $f$. Moreover, let $\iota$ be the index set $(l_1, \ldots, l_k)$ such that $\sum_m l_m = p$. For a vector $v = (v_1, \ldots, v_k)$, define $v^\iota = v_1^{l_1} \ldots v_k^{l_k}$ and let the p-th order derivative be given by $\nabla_\iota^{(p)} f$. For the p-th order kernel function $\overline{K}(\cdot)$, I adopt the following notation: the product kernel for a vector valued random variable $Z_i$ with $k$ elements and bandwidth sequence $h_n$ is written as:

$$\overline{\mathbf{K}}_h^k(Z_i - Z_l) \equiv \overline{K}_h(Z_{i1} - Z_{l1}) \times \ldots \times \overline{K}_h(Z_{ik} - Z_{lk}),$$

while:

$$\overline{\mathbf{K}}_{h,j}^k(Z_{i1}, \ldots, Z_{ik}) \equiv \overline{K}_h(Z_{i1} - Z_{j1}) \times \ldots \times \overline{K}_h(Z_{ik} - Z_{jk})$$

stands for the product of $k$ (univariate) kernel functions evaluated at $Z_{i1}, \ldots, Z_{ik}$ with summation over $j$. Also, it is implicitly assumed that whenever $\overline{\mathbf{K}}_{h,j}^k(\cdot)$ is evaluated at $V_i$ instead of the estimated $\widehat{V}_i$, the summation is over $V_j$ rather than $\widehat{V}_j$. Finally, I will use $\sup_t$ and $\inf_t$ in place of $\sup_{t \in \mathcal{T}}$ and $\inf_{t \in \mathcal{T}}$.

**Lemma A1.** Let

$$U_n(t) = \frac{1}{n(n-1)} \sum_{1 \le i < j \le n} \Big(\Big(W_j - \lambda(V_j)\Big) - \Big(W_i - \lambda(V_i)\Big)\Big) \operatorname{sign}\Big(U_i - U_j\Big) \frac{1}{h_{2,n}^2} K_{h2}(U_i - t) K_{h2}(U_j - t),$$

where $\lambda(V_i) = \mathbb{E}[\epsilon_i | V_i]$ and let $\widehat{U}_n(t)$ be defined as in Equation (10) of the main text. Under conditions A1, A2, A3, A4, A6, and the bandwidth conditions of Theorem 1, it holds that:

$$\sup_t \Big| \widehat{U}_n(t) - U_n(t) \Big| = O_p(n^{-\frac{1}{2}}).$$

**Lemma A2.** Let $\nabla^{(1)} m(t) = 0$ for all $t \in \mathcal{T}$. Moreover, define:

$$U_n^{p_0}(t) = \frac{1}{n} \sum_{i=1}^n \Big(\epsilon_i - \lambda(V_i)\Big) \int \operatorname{sign}\Big(U_i - \omega\Big) \frac{1}{h_{2n}^2} K_{h2}(\omega - t) dF(\omega) K_{h2}(U_i - t).$$

and recall the definition of $U_n(t)$ from the previous lemma. Then, under assumptions A1 to A4, it holds that:

$$\frac{1}{2} \sup_t \Big| 2U_n(t) - 2U_n^{p_0}(t) \Big| = O_p(n^{-1} h_{2n}^{-2}).$$

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

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

where $\Psi_{n,t}(\epsilon, V, U) = \int \Big(\epsilon - \lambda(V)\Big) \operatorname{sign}\Big(U - \omega\Big) \frac{1}{h_{2n}^2} K_{h2}(\omega - t) dF(\omega) K_{h2}(U - t)$ such that:

$$\sup_t \Big| \sqrt{n} U_n^{p_0}(t) - G_n(t) \Big| = O(n^{-\frac{1}{6}} h_{2n}^{-1} \log(n)^{\frac{1}{2}}).$$

**Lemma A4.** Define $\widetilde{\sigma}_n^2(t)$ as $\widehat{\sigma}_n^2(t)$ in Equation (13), but with $(\epsilon_i - \lambda(V_i))$ replacing $(\widehat{\epsilon}_i - \widehat{\lambda}(\widehat{V}_i))$. Moreover, let $\sigma_n^2(t)$ be defined by Equation (14) in the main text, while $\sigma^2(t) = \int (\mathbb{E}[\epsilon^2 | V] - \lambda(V)^2) f(V) dV f(t)^3 \int q(z)^2 K(z)^2 dz$. Under assumptions A1 to A6 and the bandwidth conditions of Theorem 1, the following holds:

- **(i)** $\sup_t \Big| h_{2n} \sigma_n^2(t) - \sigma^2(t) \Big| = o(1)$
- **(ii)** $\lim \inf_{n \longrightarrow \infty} h_{2n} \inf_t \sigma_n^2(t) > 0$

**(iii)** Uniformly in $t \in \mathcal{T}$: $\sup_t \Big|\widehat{\sigma}_n^2(t) - \sigma_n^2(t)\Big| = \sup_t \Big|\widetilde{\sigma}_n^2(t) - \sigma_n^2(t)\Big| + o_p(n^{-\frac{1}{2}} h_{2n}^{-2}) = O_p(n^{-\frac{1}{2}} h_{2n}^{-2})$.

**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.:

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

with:

$$\rho_t(s) = \frac{\int q(r)q(r-s)K(r)K(r-s)dr}{\int q^2(r)K^2(r)dr},$$

where $q(\cdot)$ was defined in Theorem 1 and:

$$\sup_t \Big|\frac{G_n(t)}{\sigma_n(t)} - \xi_n(h_{2n}^{-1}(t-a))\Big| = O_p(h_{2n}\sqrt{\log(h_{2n}^{-1})})$$

with $\sigma_n(t)$ as in Lemma A4.

**Proof of Theorem 1.** The proof consists of two main steps:

Step 1: This step follows closely the proof of Theorem 3.1 in Ghosal et al. (2000) and Lemma 3.1-3.4 therein, which correspond to Lemma A2 to A5 above, and establishes an approximation of the test statistic $S_n$ in Equation (12) with a sequence of zero-mean Gaussian processes $\{\xi_n(s) : s \in \mathcal{T}_n\}$ with continuous sample paths that satisfy:

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

where:

$$\rho_t(s) = \frac{\int q(r)q(r-s)K(r)K(r-s)dr}{\int q^2(r)K^2(r)dr},$$

and $q(\cdot)$ is defined in Theorem 1. The main differenes w.r.t. Ghosal et al. (2000) occur in Lemma A1 and Lemma A4 (iii): Lemma A1 establishes that the error of replacing $\widehat{U}_n(t)$, which contains the estimated $\widehat{\lambda}(\widehat{V}_i)$, by $U_n(t)$, which contains the unknown $\lambda(V_i)$, is of order $O_p(n^{-\frac{1}{2}})$ uniformly in $t \in \mathcal{T}$. Similarily, Lemma A4 (iii) shows that replacing the estimated variance $\widehat{\sigma}_n^2(t)$ by $\widetilde{\sigma}_n^2(t)$, which involves the unknown $\mu(U_i, V_i)$, leads to an error of order $o_p(n^{-\frac{1}{2}} h_{2n}^{-2})$ uniformly in $t$. These additional results help to bound expressions that are analogous to the terms (3.10) and (3.16) in the proof of Theorem 3.1 (Ghosal et al., 2000, p. 1062f.). That is, by virtue of Lemma A1, Lemma A2, Lemma A4 (ii), and the bandwidth conditions of Theorem 1, it holds for the analogue of (3.10) that:

$$\begin{aligned}\frac{\sqrt{n}\sup_t\Big|\widehat{U}_n(t) - U_n^{po}(t)\Big|}{\inf_t \sigma_n(t)} \leq & \frac{\sqrt{n}\sup_t\Big|\widehat{U}_n(t) - U_n(t)\Big|}{\inf_t \sigma_n(t)} + \frac{\sqrt{n}\sup_t\Big|U_n(t) - U_n^{po}(t)\Big|}{\inf_t \sigma_n(t)} \\ =& O_p(1)O(h_{2n}^{\frac{1}{2}}) + O_p(n^{-\frac{1}{2}} h_{2n}^{-2})O(h_{2n}^{\frac{1}{2}}) \\ =& o_p(1).\end{aligned}$$

Likewise, the analogue of (3.16) can be bounded by:

$$\sqrt{n}\sup_t\Big|\frac{\widehat{U}_n(t)}{\widehat{\sigma}_n(t)} - \frac{\widehat{U}_n(t)}{\sigma_n(t)}\Big| \leq \sqrt{n}\sup_t\Big|\frac{\widehat{U}_n(t)}{\sigma_n(t)}\Big| \sup_t\Big|\frac{\sigma_n(t)}{\widehat{\sigma}_n(t)} - 1\Big|. \tag{A-1}$$

By the bandwidth conditions, Lemma A1, Lemma A4 (ii), and the fact that $\sqrt{n}\sup_t\Big|U_n(t)/\sigma_n(t)\Big| = O_p((\log(h_{2n}^{-1}))^{\frac{1}{2}})$, which follows by identical steps to the ones in the proof of Theorem 3.1 (Ghosal et al., 2000, p. 1062f.) replacing the results of Lemma 3.1 to 3.4 by those of Lemma A2 to A5, the first term on the right hand side (RHS) of

(A-1) yields (uniformly in $t \in \mathcal{T}$):

$$\begin{aligned}\sqrt{n}\sup_t\Big|\frac{\widehat{U}_n(t)}{\sigma_n(t)}\Big| \leq & \sqrt{n}\sup_t\Big|\frac{U_n(t)}{\sigma_n(t)}\Big| + \frac{\sqrt{n}\sup_t\Big|\widehat{U}_n(t) - U_n(t)\Big|}{\inf_t \sigma_n(t)}\\ = & \sqrt{n}\sup_t\Big|\frac{U_n(t)}{\sigma_n(t)}\Big| + o_p(1).\end{aligned}$$

Moreover, since $|x-1| \leq |x^2 - 1|$ for $x \geq 0$, note that:

$$\sup_t\Big|\frac{\widehat{\sigma}_n(t)}{\sigma_n(t)} - 1\Big| \leq \sup_t\Big|\frac{\widehat{\sigma}^2_n(t)}{\sigma^2_n(t)} - 1\Big| \leq \frac{\sup_t\Big|\widehat{\sigma}^2_n(t) - \sigma^2_n(t)\Big|}{\inf_t \sigma^2_n(t)},$$

where the last term, by application of Lemma A4 (ii) and (iii) and the bandwidth conditions, is of order:

$$\frac{\sup_t\Big|\widehat{\sigma}^2_n(t) - \sigma^2_n(t)\Big|}{\inf_t \sigma^2_n(t)} = \frac{\sup_t\Big|\widetilde{\sigma}^2_n(t) - \sigma^2_n(t)\Big|}{\inf_t \sigma^2_n(t)} + o_p(n^{-\frac{1}{2}}h_{2n}^{-2})O(h_{2n}) = O_p(n^{-\frac{1}{2}}h_{2n}^{-1})$$

uniformly in $t \in \mathcal{T}$. As shown in the proof of Theorem 3.1 (Ghosal et al., 2000, p. 1063), the last result can be used to prove that $\sup_t\Big|(\sigma_n(t)/\widehat{\sigma}_n(t)) - 1\Big|$ from the RHS of (A-1) is also of order $O_p(n^{-\frac{1}{2}}h_{2n}^{-1})$ uniformly in $t$, and hence the analogue of (3.16) in (A-1) is of order $O_p(n^{-\frac{1}{2}}h_{2n}^{-1}\log(n)^{\frac{1}{2}}) = o_p(1)$ uniformly in $t \in \mathcal{T}$ since $\log(h_{2n}^{-1}) = O(\log(n))$ and $n^{\frac{1}{2}}h_{2n}^{\frac{3}{2}}\log(n)^{-3} \longrightarrow \infty$. The remaining steps are identical to the ones in the proof of Theorem 3.1. Minor technical differences w.r.t. the proofs of Lemma 3.1 to 3.4 in Ghosal et al. (2000) are highlighted in the proofs of the corresponding Lemma A2 to Lemma A5. Thus, combining all results, invoking the bandwidth conditions of Theorem 1, and recalling the definition of the Gaussian process $\xi_n(\cdot)$ from above, the first step establishes that:

$$\sup_t\Big|\frac{\sqrt{n}\widehat{U}_n(t)}{\widehat{\sigma}_n(t)} - \xi_n(h_{2n}^{-1}(t-a))\Big| = O_p(h_{2n}^{\frac{1}{2}} + n^{-\frac{1}{6}}h_{2n}^{-\frac{1}{2}}\log(n)^{\frac{1}{2}} + h_{2n}\log(h_{2n}^{-1})^{\frac{1}{2}}) = o_p(1).$$

Step 2: The second step derives the asymptotic behaviour of the maximum of that Gaussian process $\xi_n(\cdot)$. As noted by Ghosal et al. (2000), since the interest lies in distributions only and the covariance function $\rho(\cdot)$ does not depend on $n$, I assume in the following that the Gaussian processes are all the same with:

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

To derive the limiting distribution of the maximum of $\xi(\cdot)$ on an increasing set $\mathcal{T}_n = [0, (b-a)/h_{2n}]$, I follow the technique in Piterbarg (1996) and Lee et al. (2009) by first establishing the excursion probability of the maximum of $\xi(\cdot)$ on a fixed set $H = [0, L]$ for some finite $L > 0$, before extending the result to the increasing set $\mathcal{T}_n$. That is, noting that by (4.9) in Ghosal et al. (2000) the covariance function $\rho_t(\cdot)$ satisfies:

$$\rho_t(s) = 1 - \frac{\theta_t}{2}s^2 + o(s^2)$$

as $s \longrightarrow 0$, one can apply Theorem D.2 in Piterbarg (1996, p.16). Since Theorem D.2 formally requires that the covariance function takes the form $1 - s^2 + o(s)$, rescale $s = \left(\frac{2}{\theta}\right)^{\frac{1}{2}} s^*$ and $L = \left(\frac{2}{\theta}\right)^{\frac{1}{2}} L^*$. Then, as $u \longrightarrow \infty$:

$$\mathbb{P}\Big(\max_{s \in I}\xi(s) > u\Big) = H_\alpha L^* u\Psi(u)\Bigg(1 + o(1)\Bigg) = \frac{L\theta^{\frac{1}{2}}}{2\pi}\exp\Bigg(-\frac{u^2}{2}\Bigg)\Bigg(1 + o(1)\Bigg)$$

where the constants and approximations of Theorem D.2 have been chosen as $\alpha = 2$ (Piterbarg, 1996, p. 13), $H_2 = 1/\sqrt{\pi}$ (Piterbarg, 1996, p. 31), and $u\Psi(u) \sim (2\pi)^{-\frac{1}{2}}\exp(-u^2/2)$ (Piterbarg, 1996, p. 15).

The extension to an increasing set is now straightforward: by arguments that are identical to the proofs of Theorem G.1 in Piterbarg (1996, p. 32) and of Theorem A.3 in the supplementary material of Lee et al. (2009)[p. 9f.], one can establish that for any $x$:

$$\mathbb{P}\Big(l_n(\sup_{s\in\mathcal{T}_n}\xi(s)-l_n)<x\Big)=\exp\bigg(-\exp\Big(-x-\frac{x^2}{2l_n^2}\Big)\bigg)\Big(1+o(1)\Big),$$

where $l_n$ was defined in Theorem 1. Finally, notice that the bandwidth sequences in Theorem 1 are chosen such that: $l_n[h_{2n}^{\frac{1}{2}}+n^{-\frac{1}{6}}h_{2,n}^{-\frac{1}{2}}\log(n)^{\frac{1}{2}}+h_{2n}\log(h_2^{-1})^{\frac{1}{2}}]=o(1)$, where $l_n=O(\log(h_{2n}^{-1})^{\frac{1}{2}})=O(\log(n)^{\frac{1}{2}})$. ■

**Proof of Theorem 2.** The proof follows similar steps to the ones of the proof of Theorem 5.1 in Ghosal et al. (2000): if the null is violated for a specific $t\in\mathcal{T}$, i.e. $\nabla_U m(t)<0$, one can show that $h_{2n}^{-1}U_n(t)$ converges to a positive limit in probability. Since $h_{2n}^{\frac{1}{2}}\widetilde{\sigma}_n(t)$ tends to a positive limit, too, the order of $S_n$ is $O_p(n^{\frac{1}{2}}h_{2n}^{\frac{3}{2}})$. The bandwidth restriction $n^{\frac{1}{2}}h_{2n}^{\frac{3}{2}}\log(h_{2n}^{-1})^{-3}\longrightarrow\infty$ ensures that this rate also exceeds the order of $l_n=O(\log(h_{2n}^{-1})^{\frac{1}{2}})$ (recall that the test is constructed as $l_n(S_n-l_n)$). ■

**Proof of Theorem 4.** The proof follows closely the one of Theorem 1. Hence, only the main differences will be sketched. As before, the proof consists of two main steps:

Step 1: The first step involves the approximation of the test statistic $\overline{S}_n$ by a sequence of Gaussian processes $\{\xi_n(t,r):(t,r)\in[0,h_{2n}^{-1}]\times[0,h_{3n}^{-1}]^{d_x}\}$ with continuous sample paths such that:

$$\mathbb{E}\Big[\xi_n(t,r)\Big]=0\qquad\mathbb{E}\Big[\xi_n(t_1,r_1)\xi_n(t_2,r_2)\Big]=\rho_{tr}(t_1-t_2,r_1-r_2)$$

for $t,t_1,t_2\in\mathcal{T}_n$ and $r,r_1,r_2\in\mathcal{R}_n$, where the covariance function is given by $\rho_{tr}(t,s)\equiv\rho_t(t)\prod_{i=1}^{d_x}\rho_r(s_i)$ with $\rho_t(t)=1-\frac{\theta_t}{2}t^2(1+o(1))$ and $\rho_r(s_i)=1-\frac{\theta_r}{2}s_i^2(1+o(1))$, and $\theta_t$ and $\theta_r$ defined in Theorem 4.

In analogy to Lemma A1, using the modified bandwidth conditions of Theorem 4 and repeating the steps of the proof, an according result yields:

$$\sup_{t,r}\Big|\widehat{\overline{U}}_n(t,r)-\overline{U}_n(t,r)\Big|=O_p(n^{-\frac{1}{2}}),$$

where $\overline{U}_n(t,r)$ is the U-process indexed by $t$ and $r$ containing the unknown $\lambda(V_i)$ and $\lambda(V_j)$. Moreover, letting:

$$\overline{U}_n^{p_0}(t,r)=\frac{1}{n}\sum_{i=1}^n\Big(\epsilon_i-\lambda(V_i)\Big)\int\int\operatorname{sign}\Big(U_i-\omega\Big)\frac{1}{h_{2n}^2h_{3n}^{2d_x}}K_{h2}(\omega_1-t)$$
$$\times\mathbf{K}_{h3}^{d_x}(\omega_2-r)dF(\omega_1)dF(\omega_2)K_{h2}(U_i-t)\,\mathbf{K}_{h3}^{d_x}(X_i-r),$$

be the linear approximation of $\overline{U}_n(t,r)$, one can show, by arguments paralleling the proof of Lemma A2, that:

$$\frac{1}{2}\sup_{t,r}\Big|2\overline{U}_n(t,r)-2\overline{U}_n^{p_0}(t,r)\Big|=O_p(n^{-1}h_{2n}^{-2}h_{3n}^{-2d_x}).$$

Similarily, using again the same steps as in Lemma A3 and arguments from the proof of Lemma 3.2 in Ghosal et al. (2000) to verify the conditions of Rio's (1994) Theorem 1.1 (see Section 4, p. 35f.), one can show that for the multivariate analogue of (3.9) in Ghosal et al. (2000, p. 1062) it holds that:

$$\sup_{t,r}\Big|\frac{\sqrt{n}\overline{U}_n^{p_0}(t,r)}{\overline{\sigma}_n(t,r)}-\frac{\overline{G}_n(t,r)}{\overline{\sigma}_n(t,r)}\Big|=O_p(n^{-\frac{1}{2(d_x+3)}}(h_{2n}h_{3n}^{d_x})^{-\frac{1}{2(d_x+1)}}\log(n)^{\frac{1}{2}}),$$

where $\overline{G}_n(\cdot,\cdot)$ is a sequence of Gaussian processes (indexed by $t$ and $r$) with continuous sample paths defined in analogy to Lemma A3 and $\overline{\sigma}_n(t,r)$ was defined in Section 4. Next, let $\overline{\sigma}(t,r)=\int\Big(\mathbb{E}\Big[\epsilon^2\Big|V\Big]-\lambda(V)^2\Big)f(V)dVf(t)^3f(r)^{3d_x}\int q(z)^2K(z)^2dz\int\mathbf{K}^{d_x}(z)^2dz$ and recall the definition of $\widehat{\overline{\sigma}}_n(t,r)$ from Section 4.

Then, Lemma A4 can be replaced by the following results:

**(i)** $\sup_{t,r}\Big|h_{2n}h_{3n}^{d_x}\overline{\sigma}_n^2(t,r)-\overline{\sigma}^2(t,r)\Big|=o(1)$

**(ii)** $\lim\inf_{n\longrightarrow\infty} h_{2n}h_{3n}^{d_x}\inf_{t,r}\overline{\sigma}_n^2(t,r)>0$

**(iii)** Uniformly in $(t,r)\in\mathcal{T}\times\mathcal{R}$: $\sup_{t,r}\Big|\widehat{\overline{\sigma}}_n^2(t,r)-\overline{\sigma}_n^2(t,r)\Big| = \sup_{t,r}\Big|\widetilde{\overline{\sigma}}_n^2(t,r)-\overline{\sigma}_n^2(t,r)\Big|+o_p(n^{-\frac{1}{2}}h_{2n}^{-2}h_{3n}^{-2d_x}) = O_p(n^{-\frac{1}{2}}h_{2n}^{-2}h_{3n}^{-2d_x})$.

Finally, since $h_{2n}^{-1}h_{3n}\longrightarrow 0$, one can show that the multivariate equivalent of Lemma A5 with $\xi_n(\cdot,\cdot)$ from above yields:

$$\sup_{(t,r)}\Big|\frac{G_n(t,r)}{\overline{\sigma}_n(t,r)}-\xi_n(h_{2n}^{-1}t,h_{3n}^{1}r)\Big|=O_p(h_{2n}\sqrt{\log(h_{2n}^{-1})})$$

Thus, collecting all the results, it can be established that:

$$\begin{aligned}&\sup_{t,r}\Big|\frac{\sqrt{n}\widehat{\overline{U}}_n(t,r)}{\widehat{\sigma}_n(t,r)}-\xi_n(h_{2n}^{-1}t,h_{3n}^{-1}r)\Big|\\ =&O_p(h_{2n}^{\frac{1}{2}}h_{3n}^{\frac{1}{2}d_x}+n^{-\frac{1}{2}}h_{2n}^{-\frac{3}{2}}h_{3n}^{-\frac{3}{2}d_x}+n^{-\frac{1}{2(d_x+3)}}(h_{2n}h_{3n}^{d_x})^{-\frac{1}{2(d_x+1)}}\log(n)^{\frac{1}{2}}+h_{2n}\log(h_{2n}^{-1})^{\frac{1}{2}})\\ =&o_p(1)\end{aligned}$$

by the same arguments as in the proof of Theorem 1.

Step 2: Start by defining the Gaussian process $\{\xi(t,r):(t,r)\in\mathbb{R}\times\mathbb{R}^{d_x}\}$ with continuous sample paths such that:

$$\mathbb{E}\Big[\xi(t,r)\Big]=0 \qquad \mathbb{E}\Big[\xi(t_1,r_1)\xi(t_2,r_2)\Big]=\rho_{tr}(t_1-t_2,r_1-r_2)$$

for $t,t_1,t_2\in\mathbb{R}$ and $r,r_1,r_2\in\mathbb{R}^{d_x}$ such that $\xi_n(\cdot,\cdot)$ is the restriction of $\xi(\cdot,\cdot)$ to $\mathcal{T}_n\times\mathcal{R}_n$. The remainder of this step follows the same logic as the proofs of Theorem A.2 and A.3 in the supplementary material of Lee et al. (2009). That is, even though all components of the multidimensional Gaussian process $\xi(t,r)$ are stationary (unlike in the case of Lee et al., 2009), the covariance functions $\rho_t(\cdot)$ and $\rho_r(\cdot)$ do have different norming constants and hence standard results from Piterbarg (1996) are not directly applicable. In addition, to render the proof more accessible, I will start by considering the case of $d_x=1$ and subsequently generalize this proof to $d_x>1$. Thus, unless stated otherwise, $r$ will denote a scalar in the following.

Let $\varepsilon>0$ be a fixed constant and define the additional stationary Gaussian processes $\psi_1^-(t)$ and $\psi_1^+(t)$ with mean zero and covariance functions:

$$\rho_1^-(t)=\frac{1}{2}[1-\theta_t(1-\varepsilon)t^2+o(t^2)]$$

and

$$\rho_1^+(t)=\frac{1}{2}[1-\theta_t(1+\varepsilon)t^2+o(t^2)].$$

Moreover, define the stationary Gaussian processes $\psi_2^-(r)$ and $\psi_2^+(r)$ such that they are independent of $\psi_1^-(t)$ and $\psi_1^+(t)$ and that they have mean zero and covariance functions:

$$\rho_2^-(r)=\frac{1}{2}[1-\theta_r(1-\varepsilon)r^2+o(r^2)]$$

and

$$\rho_2^+(r)=\frac{1}{2}[1-\theta_r(1+\varepsilon)r^2+o(r^2)].$$

Finally, define, as in the proof of Theorem 1, $I_1=[0,L_1]$ and $I_2=[0,L_2]$ to be intervals of fixed length, and let $\psi^-(t,r)=\psi_1^-(t)+\psi_2^-(r)$ and $\psi^+(t,r)=\psi_1^+(t)+\psi_2^+(r)$ be the convolutions of both Gaussian processes. Notice that $\psi^-(t,r)$ and $\psi^+(t,r)$ have the same mean and variance as $\xi(t,r)$. Thus, as noted in the proof of Lemma A.3 in Lee et al. (2009), by the fact that the distribution of the maximum is monotone with respect to the variance and the Slepian inequality (e.g., Piterbarg, 1996, p. 6), it holds that:

$$\mathbb{P}\Big(\max_{(t,r)\in I_1\times I_2} \psi^-(t,r) > u\Big) \le \mathbb{P}\Big(\max_{(t,r)\in I_1\times I_2} \xi(t,r) > u\Big) \le \mathbb{P}\Big(\max_{(t,r)\in I_1\times I_2} \psi^+(t,r) > u\Big). \tag{A-2}$$

Next, applying Theorem D.2 of Piterbarg (1996, p. 16) with the same choices for the parameters and approximations as in Theorem 1 yields as $u \longrightarrow \infty$:

$$\mathbb{P}\Big(\max_{t\in I_1} 2^{\frac{1}{2}}\psi_1^-(t) > u\Big) = \Big(\frac{\theta_t}{2}(1-\varepsilon)\Big)^{\frac{1}{2}}\frac{L_1}{\pi}\exp\Big(-\frac{u^2}{2}\Big)\Big(1+o(1)\Big),$$

$$\mathbb{P}\Big(\max_{t\in I_1} 2^{\frac{1}{2}}\psi_1^+(t) > u\Big) = \Big(\frac{\theta_t}{2}(1+\varepsilon)\Big)^{\frac{1}{2}}\frac{L_1}{\pi}\exp\Big(-\frac{u^2}{2}\Big)\Big(1+o(1)\Big),$$

as well as

$$\mathbb{P}\Big(\max_{r\in I_2} 2^{\frac{1}{2}}\psi_2^-(r) > u\Big) = \Big(\frac{\theta_r}{2}(1-\varepsilon)\Big)^{\frac{1}{2}}\frac{L_2}{\pi}\exp\Big(-\frac{u^2}{2}\Big)\Big(1+o(1)\Big),$$

$$\mathbb{P}\Big(\max_{r\in I_2} 2^{\frac{1}{2}}\psi_2^+(r) > u\Big) = \Big(\frac{\theta_r}{2}(1+\varepsilon)\Big)^{\frac{1}{2}}\frac{L_2}{\pi}\exp\Big(-\frac{u^2}{2}\Big)\Big(1+o(1)\Big),$$

for any $\varepsilon > 0$. By Lemma 8.6 of Piterbarg (1996, p. 128), this implies for the convolutions $\psi^-(\cdot,\cdot)$ and $\psi^+(\cdot,\cdot)$:

$$\mathbb{P}\Big(\max_{(t,r)\in I_1\times I_2} 2^{\frac{1}{2}}\psi^-(t,r) > u\Big) = L_1L_2(1-\varepsilon)\Big(\frac{\theta_t\theta_r}{2^2\pi^3}\Big)^{\frac{1}{2}} u\exp\Big(-\frac{u^2}{4}\Big)\Big(1+o(1)\Big) \tag{A-3}$$

and

$$\mathbb{P}\Big(\max_{(t,r)\in I_1\times I_2} 2^{\frac{1}{2}}\psi^+(t,r) > u\Big) = L_1L_2(1+\varepsilon)\Big(\frac{\theta_t\theta_r}{2^2\pi^3}\Big)^{\frac{1}{2}} u\exp\Big(-\frac{u^2}{4}\Big)\Big(1+o(1)\Big), \tag{A-4}$$

respectively. Since the choice of $\varepsilon > 0$ can be made arbitrarily small and the constants on the right hand side of Equations (A-3) and (A-4) are continuous at $\varepsilon = 0$, it follows immediately from Equation (A-2) that:

$$\mathbb{P}\Big(\max_{(t,r)\in I_1\times I_2} \xi(t,r) > u\Big) = L_1L_2\Big(\frac{\theta_t\theta_r}{2\pi^3}\Big)^{\frac{1}{2}} u\exp\Big(-\frac{u^2}{2}\Big)\Big(1+o(1)\Big).$$

Since $(b-a) = 1$ and $\mathcal{T}_n = [0, h_{2n}^{-1}]$, the remaining steps are almost identical to the proof of Theorem A.3 in the supplementary material of Lee et al. (2009, p. 9f.). In particular, for $i = 2, 3$, define the sequences of sets:

$$I_{k_i} = \Big[k_i(m_{in}h_{in})^{-1}, (k_i+1)(m_{in}h_{in})^{-1} - 2\Big)$$

for $k_i = 0, 1, \ldots, m_{in} - 1$, where $m_{in}$ are increasing sequences $m_{in} \longrightarrow \infty$ as $n \longrightarrow \infty$ satisfying $h_{in}m_{in} \longrightarrow 0$, and choose $u_n = \bar{l}_n + x/\bar{l}_n$, where, for the case of $d_x = 1$, $\bar{l}_n$ is defined as the largest solution to $h_{2n}^{-1}h_{3n}^{-1}(\theta_t\theta_r/2\pi^3)^{\frac{1}{2}}\bar{l}_n\exp(\bar{l}_n^2/2) = 1$. Then, one can show that:

$$\mathbb{P}\Big(\max_{(t,r)\in[\bigcup_{k_2} I_{k_2}]\times[\bigcup_{k_3} I_{k_3}]} \xi(t,r) < u_n\Big) = \exp\Bigg(-m_{2n}m_{3n}\Bigg\{\Big[(m_{2n}h_{2n})^{-1} - 2\Big] \times \Big[(m_{3n}h_{3n})^{-1} - 2\Big]\Bigg\}\Big(\frac{\theta_t\theta_r}{2\pi}\Big)^{\frac{1}{2}} u_n\exp\Big(-\frac{u_n^2}{2}\Big)\Bigg) + o(1)$$

so that:

$$\mathbb{P}\Big(\max_{(t,r)\in[\bigcup_{k_2} I_{k_2}]\times[\bigcup_{k_3} I_{k_3}]} \xi(t,r) < u_n\Big) = \exp\Bigg(-\exp\Bigg(-x - \frac{x^2}{2\bar{l}_n^2}\Bigg)\Bigg(1 + \frac{x}{\bar{l}_n^2}\Bigg)\Bigg) + o(1).$$

The remaining steps parallel the ones in the proof of Theorem A.3 in the supplementary material of Lee et al. (2009).

Next, consider the case for $d_x > 1$. Defining $d_x$ Gaussian processes of the kind of $\psi_2^-(r)$ and $\psi_2^+(r)$ and letting $I_2^{d_x} = [0, L_2]^{d_x}$ be a rectangle with a fixed volume $L_2^{d_x}$, one can show, by applying Lemma 8.6 of Piterbarg

(1996) repeatedly, that:

$$\mathbb{P}\Big(\max_{(t,r)\in I_1\times I_2^{d_x}} \xi(t,r) > u\Big) = L_1 L_2^{d_x} 2^{-d_x+1}\left(\frac{\theta_t}{2}\right)^{\frac{1}{2}} \frac{\theta_r^{\frac{d_x}{2}}}{\pi^{\frac{3}{2}+\frac{d_x}{2}}} u^{d_x}\exp\Big(-\frac{u^2}{2}\Big).$$

as $u \longrightarrow \infty$. This result then allows to mimic the steps from before to show that:

$$\mathbb{P}\Big(\max_{(t,r)\in[\bigcup_{k_2} I_{k_2}]\times[\bigcup_{k_3} I_{k_3}]^{d_x}} \xi(t,r) < u_n\Big) = \exp\left(-\exp\left(-x-\frac{x^2}{2\bar{l}_n^2}\right)\left(1+\frac{x}{\bar{l}_n^2}\right)^{d_x}\right)+o(1),$$

where $\bar{l}_n$ is now the largest solution to:

$$h_{2n}^{-1}h_{3n}^{-d_x}2^{-d_x+1}\left(\frac{\theta_t}{2}\right)^{\frac{1}{2}}\frac{\theta_r^{\frac{d_x}{2}}}{\pi^{\frac{3}{2}+\frac{d_x}{2}}}\bar{l}_n^{d_x}\exp\left(-\frac{\bar{l}_n^2}{2}\right)=1.$$

■

**Proof of Lemma A1.** In the following, define $R_{ij,n}(t) \equiv \text{sign}(U_i - U_j)K_{h2}(U_j - t)K_{h2}(U_i - t)$. Notice that:

$$\sup_t\Big|\widehat{U}_n(t)-U_n(t)\Big| = \sup_t\Big|\frac{1}{n(n-1)}\sum_{1\le i<j\le n}\Big((\widehat{\lambda}(\widehat{V}_i)-\lambda(V_i))+(\lambda(V_j)-\widehat{\lambda}(\widehat{V}_j))\Big)\frac{1}{h_{2n}^2}R_{ij,n}(t)\Big|.$$

This term can be bounded by:

$$\begin{aligned}
&\sup_t\Big|\frac{1}{n(n-1)}\sum_{1\le i<j\le n}\Big((\widehat{\lambda}(\widehat{V}_i)-\widehat{\lambda}(V_i))+(\widehat{\lambda}(V_j)-\widehat{\lambda}(\widehat{V}_j))\Big)\frac{1}{h_{2n}^2}R_{ij,n}(t)\Big| \\
&+\sup_t\Big|\frac{1}{n(n-1)}\sum_{1\le i<j\le n}\Big((\widehat{\lambda}(V_i)-\lambda(V_i))+(\lambda(V_j)-\widehat{\lambda}(V_j))\Big)\frac{1}{h_{2n}^2}R_{ij,n}(t)\Big| \qquad \text{(A-5)}\\
=&I_{1n}(t)+I_{2n}(t).
\end{aligned}$$

Start with $I_{1n}(t)$ and notice that $(\widehat{\lambda}(\widehat{V}_i)-\widehat{\lambda}(V_i))+(\widehat{\lambda}(V_j)-\widehat{\lambda}(\widehat{V}_j)) = (\widehat{\mu}(U_i,\widehat{V}_i)-\widehat{\mu}(U_i,V_i))+(\widehat{\mu}(U_j,V_j)-\widehat{\mu}(U_j,\widehat{V}_j))$. I only consider the term involving $(\widehat{\mu}(U_i,\widehat{V}_i)-\widehat{\mu}(U_i,V_i))$, the other term will follow analogously. Define:

$$\widehat{s}(U_i,\widehat{V}_i)=\frac{1}{nh_{1n}^2}\sum_{k=1}^{n}\widehat{\tau}_{in}W_k\overline{\mathbf{K}}_{h1,k}^2(U_i,\widehat{V}_i)$$

and

$$\widehat{f}(U_i,\widehat{V}_i)=\frac{1}{nh_{1n}^2}\sum_{k=1}^{n}\overline{\mathbf{K}}_{h1,k}^2(U_i,\widehat{V}_i),$$

and let $\widetilde{s}(\cdot,\cdot)$ and $\widetilde{f}(\cdot,\cdot)$ be defined accordingly with $V_i$ and $V_k$ replacing $\widehat{V}_i$ and $\widehat{V}_k$ and $\tau_{in}$ replacing $\widehat{\tau}_{in}$. Then:

$$I_{1n}(t)=\sup_t\Big|\frac{1}{n(n-1)}\sum_{1\le i<j\le n}\left\{\frac{\widehat{s}(U_i,\widehat{V}_i)-\widetilde{s}(U_i,V_i)}{\widetilde{f}(U_i,V_i)}-\frac{\widetilde{f}(U_i,V_i)-\widehat{f}(U_i,V_i)}{\widetilde{f}(U_i,V_i)}\widehat{\mu}(U_i,\widehat{V}_i)\right\}\frac{1}{h_{2n}^2}R_{ij,n}(t)\Big|.$$

I concentrate on the first term, the second one follows by similar arguments. Expanding the first term yields:

$$\sup_t\Big|\frac{1}{n(n-1)}\sum_{1\le i<j\le n}\left\{\frac{\widehat{s}(U_i,\widehat{V}_i)-\widetilde{s}(U_i,V_i)}{f(U_i,V_i)}-\left(\frac{1}{\widetilde{f}(U_i,V_i)}-\frac{1}{f(U_i,V_i)}\right)\left(\widehat{s}(U_i,\widehat{V}_i)-\widetilde{s}(U_i,V_i)\right)\right\}\frac{1}{h_{2n}^2}R_{ij,n}(t)\Big|.$$

The second term is of smaller order (uniformly in t) since by Theorem 2 of Hansen (2008) for slowly expanding sets (set $\beta = \infty$ and $q = \infty$):

$$\sup_{U_i, Z_i \in \mathcal{W}_n} \left| \widetilde{f}(U_i, V(U_i, Z_i)) - f(U_i, V(U_i, Z_i)) \right| = O_p\left( \left( \frac{\log(n)}{n h_{1n}^2} \right)^{\frac{1}{2}} \right) = o_p(1),$$

where $\mathcal{W}_n = O(c_n) = O(\log(n)^{\frac{1}{d_z+1}})$ and the $o_p(1)$ term follows from the bandwidth conditions. The density of the first term is strictly bounded away from zero on $\mathcal{W}_n$ by A2 and the term can be bounded by:

$$\begin{aligned}
&\sup_t \Big| \frac{1}{n^2(n-1)} \sum_{1 \le i < j \le n} \sum_{k=1}^{n} \frac{\tau_{in}}{h_{1n}^2 h_{2n}^2 f(U_i, V_i)} W_k \Big\{ \overline{\mathbf{K}}_{h1,k}^2(U_i, \widehat{V}_i) - \overline{\mathbf{K}}_{h1,k}^2(U_i, V_i) \Big\} R_{ij,n}(t) \Big| \\
&+ \sup_t \Big| \frac{1}{n^2(n-1)} \sum_{1 \le i < j \le n} \sum_{k=1}^{n} \frac{(\widehat{\tau}_{in} - \tau_{in})}{h_{1n}^2 h_{2n}^2 f(U_i, V_i)} W_k \overline{\mathbf{K}}_{h1,k}^2(U_i, V_i) R_{ij,n}(t) \Big| \qquad \text{(A-6)} \\
&+ o_p(1),
\end{aligned}$$

where the $o_p(1)$ term contains the cross-products that are of smaller order uniformly in $t$. Let the first and second term be denoted as $I_{11n}(t)$ and $I_{12n}(t)$. Starting with $I_{11n}(t)$, I will show that these two expressions are or order $o_p(n^{-\frac{1}{2}})$ and $O_p(n^{-\frac{1}{2}})$ uniformly in $t$, respectively. Since $(\widehat{V}_i - \widehat{V}_k) = (\widehat{g}(Z_i) - \widehat{g}(Z_k))$, a mean value expansion of $I_{11n}(t)$ around $(V_i - V_k)$ yields:

$$\begin{aligned}
I_{11n}(t) = \sup_t \Big| &\frac{1}{n^2(n-1)} \sum_{1 \le i < j \le n} \sum_{k=1}^{n} \frac{\tau_{in}}{h_{1n}^3 h_{2n}^2 f(U_i, V_i)} W_k \overline{K}_{h1}(U_i - U_k) \\
&\times \nabla_1^{(1)} \overline{K}_{h1}(\overline{V}_i - \overline{V}_k) \{ (\widehat{g}(Z_i) - g(Z_i)) - (\widehat{g}(Z_k) - g(Z_k)) \} R_{ij,n}(t) \Big|,
\end{aligned}$$

where $(\overline{V}_i - \overline{V}_k)$ denote intermediate values. Since the term involving $(\widehat{g}(Z_i) - g(Z_i))$ can be treated identically to the one containing $(\widehat{g}(Z_k) - g(Z_k))$, I will focus on $(\widehat{g}(Z_i) - g(Z_i))$ in the following. Firstly, notice that by standard arguments $(\widehat{g}(Z_i) - g(Z_i))$ can be decomposed as:

$$\widehat{g}(Z_i) - g(Z_i) = \frac{(\widehat{g}(Z_i) - g(Z_i)) \widehat{f}(Z_i)}{f(Z_i)} + \left( \frac{f(Z_i) - \widehat{f}(Z_i)}{f(Z_i) \widehat{f}(Z_i)} \right) (\widehat{g}(Z_i) - g(Z_i)) \widehat{f}(Z_i),$$

where the second term is of smaller order uniformly in $t$ as $\sup_{Z \in \mathcal{Z}_n} |\widehat{f}(Z_i) - f(Z_i)| = o_p(1)$ by Theorem 2 of Hansen (2008) (set again $\beta = \infty$ and $q = \infty$) and the bandwidth conditions. Thus, neglecting the remainder term of smaller order and plugging the first part back into $I_{11n}(t)$ and writing $U_l$ as $g(Z_l) + V_l$ yields:

$$\begin{aligned}
&\sup_t \Big| \frac{1}{n^3(n-1)} \sum_{1 \le i < j \le n} \sum_{k=1}^{n} \sum_{l=1}^{n} \frac{\tau_{in}}{h_{1n}^{3+d_z} h_{2n}^2 f(U_i, V_i)} W_k \overline{K}_{h1}(U_i - U_k) \\
&\times \nabla_1^{(1)} \overline{K}_{h1}(\overline{V}_i - \overline{V}_k) \left\{ \frac{g(Z_l) - g(Z_i)}{f(Z_i)} \overline{\mathbf{K}}_{h1}^{d_z}(Z_i - Z_l) + \frac{V_l}{f(Z_i)} \overline{\mathbf{K}}_{h1}^{d_z}(Z_i - Z_l) \right\} R_{ij,n}(t) \Big| \\
&\le I_{111n}(t) + I_{112n}(t),
\end{aligned}$$

with the obvious definitions of $I_{111n}(t)$ and $I_{112n}(t)$.

In the following, I will examine $I_{111n}(t)$ and $I_{112n}(t)$ in turn and show that each can be approximated by a fourth order U-processes. Then, applying the Hoeffding decomposition and using results from Nolan and Pollard (1987), Pakes and Pollard (1989), and Sherman (1994) as well as the Rosenthal and Markov inequalities, I will establish uniform convergence rates for both expressions.

Start with $I_{111n}(t)$: first, in order to be able to approximate $I_{111n}(t)$ with a fourth order U-process, I will demonstrate that terms with $i = k$, $j = l$, $j = k$, and $k = l$ can be neglected asymptotically (notice that terms with $i = j$ and $i = l$ are automatically zero). Furthermore, since the number of pairs of observations with $i = k$, $j = l$, $j = k$, and $k = l$ dominates the number of triplets with $j = l = k$, it is clear that asymptotic negligibility of the former implies asymptotic negligibility of the latter.

Define $\Psi_{ijk,n}(t) \equiv \frac{\tau_{in}}{f(U_i,V_i)} W_k \overline{K}_{h1}(U_i - U_k)\nabla_1^{(1)}\overline{K}_{h1}(\overline{V}_i - \overline{V}_k)R_{ij,n}(t)$ and let the symmetrized 'kernel function' of $I_{111n}(t)$ (with 24 elements) be given by:

$$\begin{aligned} f_n(t) \equiv & \Psi_{ijk,n}(t)\frac{g(Z_l)-g(Z_i)}{f(Z_i)}\overline{\mathbf{K}}_{h1}^{d_z}(Z_i - Z_l)R_{ij,n}(t) + \ldots + \Psi_{kli,n}(t)\frac{g(Z_j)-g(Z_k)}{f(Z_k)}\overline{\mathbf{K}}_{h1}^{d_z}(Z_k - Z_j)R_{kl,n}(t) \\ \equiv & f_n^{(1)}(t) + \ldots + f_n^{(24)}(t), \end{aligned}$$

where $f_n^{(m)}(t)$ with $m = 1,\ldots,24$ are the single elements of $f_n(t)$. Notice also that $f_n(t)$ and $f_n^{(m)}(t)$ stand short for $f_n(\omega_i,\omega_j,\omega_k,\omega_l;t)$ and $f_n^{(m)}(\omega_i,\omega_j,\omega_k,\omega_l;t)$, respectively, with $\omega_i = \{W_i, U_i, Z_i\}$. Next, pick, for instance, terms with $i = k$, which can be bounded as in the proof of Theorem 3.1 of Ahn and Powell (1993, p. 25):

$$\begin{aligned} & \sup_t \Big| \frac{1}{n^3(n-1)h_{1n}^{3+d_z}h_{2n}^2} \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{l\neq i\\ l\neq j}} f_n(\omega_i,\omega_i,\omega_j,\omega_l;t)\Big| \\ \leq & \frac{\overline{K}(0)\overline{K}^{(1)}}{3nh_{1n}^3}\sup_t\Big|\frac{1}{h_{1n}^{d_z}h_{2n}^2}\binom{n}{3}\sum_{i=1}^{n-3}\sum_{j>i}\sum_{l>j}\frac{\tau_{in}}{f(U_i,V_i)}W_i\frac{g(Z_l)-g(Z_i)}{f(Z_i)}\overline{\mathbf{K}}_{h1}^{d_z}(Z_i - Z_l)R_{ij,n}(t) + \ldots \\ & + \frac{\tau_{ln}}{f(U_l,V_l)}W_l\frac{g(Z_i)-g(Z_l)}{f(Z_l)}\overline{\mathbf{K}}_{h1}^{d_z}(Z_l - Z_i)R_{lj,n}(t)\Big|, \end{aligned} \tag{A-7}$$

where $\overline{K}^{(1)}$ is a genuine positive constant that bounds $\nabla_1^{(1)}\overline{K}(0)$. Applying the Hoeffding decomposition to the U-process (indexed by $t$) in the last line of Equation (A-7) and using arguments that will be outlined in the next four paragraphs below, it can be shown that the different elements of the Hoeffding decomposition are of order $o_p(n^{-\frac{1}{2}})$ uniformly in $t$. Thus, since $n^{-1}h_{1n}^{-3} \longrightarrow 0$ by the bandwidth conditions of Theorem 1, the error of omitting observations with $i = k$ is of order $o_p(n^{-\frac{1}{2}})$ uniformly in $t$. Similar arguments can be applied to show that terms with $j = l$, $j = k$, and $k = l$ are asymptotically negligible and yield an error of order $o_p(n^{-\frac{1}{2}})$ uniformly in $t$.

Thus, $I_{111n}(t)$ can be approximated by:

$$I_{111n}(t) = \frac{1}{4h_{1n}^{3+d_z}h_{2n}^2}\binom{n}{4}\sum_{i=1}^{n-4}\sum_{j>i}\sum_{l>j}\sum_{k\geq l} f_n(\omega_i,\omega_j,\omega_l,\omega_k;t) + o_p(n^{-\frac{1}{2}})$$

uniformly in $t$. The Hoeffding decomposition of this fourth order U-process (indexed by $t \in \mathcal{T}$) yields:

$$\begin{aligned} \frac{1}{4h_{1n}^{3+d_z}h_{2n}^2}\binom{n}{4}\sum_{i=1}^{n-4}\sum_{j>i}\sum_{l>j}\sum_{k\geq l} f_n(\omega_i,\omega_j,\omega_l,\omega_k;t) = & \frac{1}{4h_{1n}^{3+d_z}h_{2n}^2}\Bigg\{\mathbb{E}\Big[f_n(t)\Big] + \frac{4}{n}\sum_{i=1}^{n}\Big\{\mathbb{E}\Big[f_n(t)\Big|\omega_i\Big] - \mathbb{E}\Big[\mathbb{E}\Big[f_n(t)\Big|\omega_i\Big]\Big]\Big\} \\ & + U_n^2 f_{2n}(t) + U_n^3 f_{3n}(t) + U_n^4 f_{4n}(t)\Bigg\} \\ = & J_{0n}(t) + J_{1n}(t) + J_{2n}(t) + J_{3n}(t) + J_{4n}(t), \end{aligned}$$

where $U_n^2 f_{2n}(t)$, $U_n^3 f_{3n}(t)$, and $U_n^4 f_{4n}(t)$ denote second, third, and fourth order degenerate U-processes (see, e.g., Sherman (1994, p. 449) for a definition of higher-order degenerate U-processes).

Starting with $J_{0n}(t)$, I first establish that the class of functions $\mathcal{F}_n \equiv \mathcal{F}_n^{(1)} + \ldots + \mathcal{F}_n^{(24)}$, with $\mathcal{F}_n^{(1)} \equiv \{\Psi_{ijk,n}(t)\frac{g(Z_i)}{f(Z_i)}\overline{\mathbf{K}}_{h1}^{d_z}(Z_i - Z_l) : t \in \mathcal{T}\}$ and $\mathcal{F}_n^{(2)},\ldots,\mathcal{F}_n^{(24)}$ defined analogously, to which $h_{1n}^{3+d_z}h_{2n}^2 J_{0n}(t)$ belongs, is Euclidean for a constant envelope (see Pakes and Pollard, 1989, p. 1032 f.): first, note that by invoking assumptions A1, A2, A3, A4, A6, and the Cauchy-Schwarz inequality together with (repeated) applications of Lemma 22 (ii) of Nolan and Pollard (1987) and Lemma 2.14 (ii) of Pakes and Pollard (1989), each class $\mathcal{F}_n^{(m)}$ with $m = 1,\ldots,24$ is Euclidean for a constant envelope. Then, multiple applications of Lemma 2.14 (i) of Pakes and Pollard (1989) allow to deduce that also $\mathcal{F}_n$ is Euclidean for a constant envelope. Next, using change of variables with $v_1 = ((U_i - U_k)/h_{1n})$, $v_2 = ((\overline{V}_i - \overline{V}_k)/h_{1n})$, $u_1 = ((U_i - t)/h_{2n})$, $u_2 = ((U_j - t)/h_{2n})$, and $v_3 = ((Z_i - Z_l)/h_{1n})$, integration by parts, iterated expectations, and a p-th order Taylor expansion around $h_{1n}v_3 = 0$ (where $v_3$ is a $(1 \times d_z)$ vector) jointly with A3 and A4, one obtains for the p-th order derivative of the first element of $J_{0n}(t)$ (the other 23 elements of the symmetric kernel function of $J_{0n}(t)$ follow accordingly):

$$\int \dots \int \tau_{k+h_1 v_1,n} \overline{K}(v_1)\overline{K}(v_2) \frac{\mathbb{E}[W_k|U_k,V_k]\nabla_2^{(1)} f(U_k + h_{1n}v_1, V_k + h_{1n}v_2) f(U_k,V_k)}{f(U_k + h_{1n}v_1, V_k + h_{1n}v_2)} \Big\{ \frac{h_{1n}^p}{p!} \sum_{\iota=1}^{d_z} v_3^{\iota} \overline{K}^{d_z}(v_3)$$
$$\times \nabla_{\iota}^{(p)} g(Z_l) f(Z_l) \Big\} \operatorname{sign}(u_1 - u_2) K(u_1) K(u_2) f(t + h_{2n}u_1) f(t + h_{2n}u_2) dv_1 dv_2 dv_3 du_1 du_2 dZ_l dU_k dV_k.$$

Since $h_{1n}^{3+d_z} h_{2n}^2 J_{0n}(t)$ belongs to a class of functions that is Euclidean for a constant envelope (which is independent of $t$) and $n^{\frac{1}{2}} h_{1n}^p c_n \longrightarrow \Delta_n$ by the bandwidth conditions, it is straightforward to deduce that this expression is of order $O(h_{1n}^p) = o(n^{-\frac{1}{2}})$ uniformly in $t \in \mathcal{T}$.

Next, I address $J_{1n}(t)$: using again change of variables and integration by parts, one can show that $\mathbb{E}\Big[f_n(t)\Big|\omega_i\Big]$, the leading term of $J_{1n}(t)$, is of order $O(h_{2n}^{-1})$ and hence $\mathbb{E}\Big[\mathbb{E}\Big[f_n(t)\Big|\omega_i\Big]^{\kappa}\Big] = O(h_{2n}^{-\kappa+1})$ for some positive integer $\kappa$. By Rosenthal's inequality (e.g., Petrov, 1995, p. 59):

$$\mathbb{E}\Big[\Big|J_{1n}(t)\Big|^{2\kappa}\Big] \leq n^{-2\kappa} C_\kappa (n h_{2n}^{-2\kappa+1} + n^{\kappa} h_{2n}^{-\kappa}), \tag{A-8}$$

where $C_\kappa$ is a positive constant constant that only depends on $\kappa$. Setting $\kappa = 1$ and using Markov's inequality, Equation (A-8) implies that $J_{1n}(t)$ is of order $O_p(n^{-1}h_{2n}^{-1})$, which is of order $o_p(n^{-\frac{1}{2}})$ by the bandwidth conditions. Moreover, since the bound is independent of $t$, convergence is uniform across $t \in \mathcal{T}$.

A similar line of argument can be used for $J_{2n}(t)$, $J_{3n}(t)$, and $J_{4n}(t)$. For instance, note that $J_{2n}(t)$ is defined as:

$$J_{2n}(t) = \frac{1}{4h_{1n}^{3+d_z} h_{2n}^2} \left\{ \frac{8}{n(n-1)} \sum_{i=1}^{n-2} \sum_{j>i} \mathbb{E}\Big[f_n(t)\Big|\omega_i,\omega_j\Big] - \mathbb{E}\Big[f_n(t)\Big|\omega_i\Big] - \mathbb{E}\Big[f_n(t)\Big|\omega_j\Big] + \mathbb{E}\Big[\mathbb{E}\Big[f_n(t)\Big|\omega_i\Big]\Big] \right\}.$$

Once again, it can be shown that the leading terms of $\mathbb{E}\Big[f_n(t)\Big|\omega_i,\omega_j\Big]$ are of order $O(h_{1n}^{-3})$, $O(h_{1n}^{-d_z})$ and $O(h_{2n}^{-2})$, depending on the position of the conditioning arguments in the different elements of the symmetrized kernel function $f_n(t)$ (i.e., integration by parts and change of variables cannot always be applied). Thus, $\mathbb{E}\Big[\mathbb{E}\Big[f_n(t)\Big|\omega_i,\omega_j\Big]^{\kappa}\Big] = O(h_{1n}^{-3\kappa+3} + h_{1n}^{-d_z\kappa+d_z} + h_{2n}^{-2\kappa+2})$, while $\mathbb{E}\Big[\Big|\mathbb{E}\Big[f_n(t)\Big|\omega_i,\omega_j\Big]\Big|^{\kappa}\Big] = O(h_{1n}^{-3\kappa+2} + h_{1n}^{-d_z\kappa+d_z} + h_{2n}^{-2\kappa+2})$. Applying again Rosenthal's inequality yields:

$$\mathbb{E}\Big[\Big|J_{2n}(t)\Big|^{2\kappa}\Big] \leq n^{-4\kappa} C_\kappa \Big( n^2 (h_{1n}^{-6\kappa+2} + h_{1n}^{-2\kappa d_z + d_z} + h_{2n}^{-4\kappa+2}) + n^{2\kappa} (h_{1n}^{-3\kappa} + h_{1n}^{-\kappa d_z} + h_{2n}^{-2\kappa}) \Big), \tag{A-9}$$

where $C_\kappa$ is a again a positive constant constant depending on $\kappa$ only. Setting $\kappa = 1$ and applying Markov's inequality, one can show that $J_{2n}(t)$ is of order $o_p(n^{-\frac{1}{2}})$ uniformly in $t \in \mathcal{T}$ since $n^{\frac{1}{2}} h_{2n}^{-1} h_{1n}^{3+d_z} c_n^{-2} \log(n)^{-1} \longrightarrow \infty$ with $3 + d_z \geq 4$ and $n^{\frac{1}{6}} h_{2n}^{\frac{1}{2}} \log(n)^{-1} \longrightarrow \infty$. Using identical arguments, $J_{3n}(t)$ and $J_{4n}(t)$ can also be shown to be of order $o_p(n^{-\frac{1}{2}})$ uniformly in $t \in \mathcal{T}$.

Turning to $I_{112n}(t)$ (p. 40), a similar line of argument can again be invoked to show that neglecting pairs of equal observations results in an error of order $o_p(n^{-\frac{1}{2}})$ uniformly in $t$, so that the U-process can be rewritten in symmetrized form as follows:

$$\frac{1}{4h_{1n}^{3+d_z} h_{2n}^2} \binom{n}{4} \sum_{i=1}^{n-4} \sum_{j>i} \sum_{l>j} \sum_{k \geq l} \Psi_{ijk,n}(t) \frac{V_l}{f(Z_i)} \overline{\mathbf{K}}_{h1}^{d_z}(Z_i - Z_l) + \ldots + \Psi_{lkj,n}(t) \frac{V_i}{f(Z_l)} \overline{\mathbf{K}}_{h1}^{d_z}(Z_l - Z_i).$$

Notice that, since $\mathbb{E}\Big[V_i\Big|Z_i\Big] = 0$, it is straightforward to show that the unconditional expectation of the symmetrized kernel function is zero. Moreover, apply Rosenthal's and Markov's inequalities as for $J_{1n}(t)$ and $J_{2n}(t)$ above to demonstrate that the terms of the Hoeffding expansion are of order $o_p(n^{-\frac{1}{2}})$ uniformly in $t$.

The next term to be examined is $I_{12n}(t)$ from Equation (A-6). A second order Taylor expansion around $f(U_i, Z_i)$ yields:

$$\sup_t \Big| \frac{1}{n^2(n-1)} \sum_{1\le i<j\le n} \sum_{k=1}^{n} \Big( \nabla_1^{(1)} \tau_{in} \Big( \widehat{f}(U_i, Z_i) - f(U_i, Z_i) \Big) + \frac{1}{2} \nabla_1^{(2)} \tau_{in}^* \Big( \widehat{f}(U_i, Z_i) - f(U_i, Z_i) \Big)^2 \Big) \times \frac{W_k \overline{\mathbf{K}}_{h1,k}^2(U_i, V_i)}{h_{1n}^2 h_{2n}^2 f(U_i, V_i)} R_{ij,n}(t) \Big|,$$

where $\tau_{in}^*$ stands for $\tau(f^*(U_i, Z_i))$ with $f^*(\cdot,\cdot)$ lying on the line segment between $\widehat{f}(U_i, Z_i)$ and $f(U_i, Z_i)$. Firstly, note that uniformly in $t$:

$$\sup_t \Big| \frac{1}{n^2(n-1)} \sum_{1\le i<j\le n} \sum_{k=1}^{n} \frac{W_k \overline{\mathbf{K}}_{h1,k}^2(U_i, V_i)}{h_{1n}^2 h_{2n}^2 f(U_i, V_i)} R_{ij,n}(t) \Big| = O_p(1)$$

by the same steps as for $I_{111n}(t)$ (arguments from Nolan and Pollard (1987), Pakes and Pollard (1989), and Sherman (1994) together with conditions A1, A2, A3, A4, and A6 can be used to show that the function belongs to a class of functions indexed by $t$ that is Euclidean for a constant envelope). Then, the uniform convergence rates of Theorem 2 in Hansen (2008) together with the bandwidth condition $n^{\frac{1}{2}} h_{1n}^{d_z+1} c_n^{-2} \log(n)^{-1} \longrightarrow \infty$ of Theorem 1 imply that:

$$\sup_t \Big| \frac{1}{n^2(n-1)} \sum_{1\le i<j\le n} \sum_{k=1}^{n} \frac{\nabla_1^{(1)} \tau_{in} \Big( \widehat{f}(U_i, Z_i) - f(U_i, Z_i) \Big)}{h_{1n}^2 h_{2n}^2 f(U_i, V_i)} W_k \overline{\mathbf{K}}_{h1,k}^2(U_i, V_i) R_{ij,n}(t) \Big| + o_p(n^{-\frac{1}{2}})$$

uniformly in $t$. The first term can be written out and bounded by:

$$\sup_{f(U_i,Z_i)} \Big| \nabla_1^{(1)} \tau_{in} \Big| \sup_t \Big| \frac{1}{n^2(n-1)} \sum_{1\le i<j\le n} \sum_{k=1}^{n} \sum_{l=1}^{n} \frac{\Big( \frac{1}{h_{1n}^2} \overline{\mathbf{K}}_{h1,l}^2(U_i, Z_i) - f(U_i, Z_i) \Big)}{h_{1n}^2 h_{2n}^2 f(U_i, V_i)} W_k \overline{\mathbf{K}}_{h1,k}^2(U_i, V_i) R_{ij,n}(t) \Big|.$$

As before, for the second expression, one can demonstrate that terms with $l = i$, $k = i$, $j = k$, $j = l$, and $l = k$ can be neglected asymptotically (uniformly in $t$). This allows to write down a symmetrized version of the kernel function and to apply the Hoeffding decomposition. That is, using the same arguments as for $I_{111n}(t)$, one can demonstrate that the symmetrized version of $(\overline{\mathbf{K}}_{h1,l}^2(U_i, Z_i) - f(U_i, Z_i)) f(U_i, V_i)^{-1} W_k \overline{\mathbf{K}}_{h1,k}^2(U_i, V_i) R_{ij,n}(t)$ is Euclidean for a constant envelope function, and hence that convergence holds uniformly across $t$. Then, after change of variables ($v_1 = ((U_i - U_l)/h_{1n})$, $v_2 = ((Z_i - Z_l)/h_{1n})$, $v_3 = ((Z_i - Z_k)/h_{1n})$, $v_4 = ((V_i - V_k)/h_{1n})$, $u_1 = ((U_i - t)/h_{2n})$, and $u_2 = ((U_j - t)/h_{2n})$), iterated expectations, a p-th order Talyor expansion of $f(\cdot,\cdot)$ around $h_{1n}v_1 = 0$ and $h_{1n}v_2 = 0$ ($v \equiv (v_1, v_2)$), A3 and A4, the p-th order derivative of the first element of the unconditional expectation is given by:

$$\int \dots \int \frac{f(U_i + h_{1n}v_3, V_i + h_{1n}v_4)}{f(U_i, V_i)} \Big\{ \frac{h_{1n}^p}{p!} \sum_{\iota=1}^{2} v^\iota \overline{\mathbf{K}}^2(v) \nabla_\iota^{(p)} f(U_i, Z_i) \Big\} \mathbb{E}\Big[ W_i \Big| U_i + h_{1n}v_3, V_i + h_{1n}v_4 \Big]$$
$$\times \overline{K}(v_3)\overline{K}(v_4) f(U_i, Z_i, V_i) \operatorname{sign}(u_1 - u_2) K(u_1) K(u_2) f(t + h_{2n}u_1) f(t + h_{2n}u_2) dv_1 dv_2 dv_3 dv_4 du_1 du_2 dU_i dZ_i dV_i,$$

which is of order $O(h^p) = o(n^{-\frac{1}{2}})$ uniformly in $t$. Hence, the leading terms converge at rate $O(n^{-\frac{1}{2}})$ since $nh_{1n}^{2p} c_n^2 \longrightarrow \Delta_n$. The higher order elements of the Hoeffding expansion can be addressed making again repeatedly use of Rosenthal's and Markov's inequality to show that they are of order $o_p(n^{-\frac{1}{2}})$ uniformly in $t$.

Finally, it remains to show that $I_{2n}(t)$ from (A-5) is $o_p(n^{-\frac{1}{2}})$ uniformly in $t$. First of all, notice that $\widehat{\lambda}(V_i) - \lambda(V_i) = \Big( \frac{1}{n} \sum_{j=1}^{n} \widehat{\mu}(U_i, V_j) - \int \mathbb{E}[W_i | U_i, V_i = v] f(v) dv \Big) + \Big( \mathbb{E}[W_i | U_i, V_i] - \widehat{\mu}(U_i, V_i) \Big)$. Thus, applying the triangle inequality, $I_{2n}(t)$ can be bounded by the sum of two $\sup_t$ expressions, each involving one of the terms in round brackets for $i$ and $j$, respectively. I will only sketch the steps for the expression involving $\Big(\mathbb{E}[W_i|U_i,V_i] - \widehat{\mu}(U_i,V_i)\Big)$, the ones for the first term follow by similar arguments. That is, notice that the expression containing $\Big(\Big(\mathbb{E}[W_i|U_i,V_i] - \widehat{\mu}(U_i,V_i)\Big) + \Big(\widehat{\mu}(U_j,V_j) - \mathbb{E}[W_j|U_j,V_j]\Big)\Big)$ can be rewritten as:

$$\sup_t\Bigg|\frac{1}{n^2(n-1)}\sum_{1\le i<j\le n}\sum_{k=1}^{n}\Bigg(\frac{\tau_{in}(\mathbb{E}[W_i|U_i,V_i]-W_k)\overline{\mathbf{K}}^2_{h1,k}(U_i,V_i)}{\widehat{f}(U_i,V_i)} + \frac{\tau_{jn}(W_k-\mathbb{E}[W_j|U_j,V_j])\overline{\mathbf{K}}^2_{h1,k}(U_j,V_j)}{\widehat{f}(U_j,V_j)}\Bigg)\frac{1}{h_{2n}^2}R_{ij,n}(t)\Bigg|.$$

The remaining steps are now as before: firstly, invoking Theorem 2 of Hansen (2008), one can replace the estimated densities in the denominator by the true densities $f(U_i,V_i)$ and $f(U_j,V_j)$. Then, showing that terms with $i=k$ and $j=k$ are asymptotically negligible uniformly in $t$ allows to construct a symmetric kernel function and to apply the Hoeffding decomposition. Notice that the U-process (indexed by $t$) is degenerate in $i$ and $j$ since the conditional expecation of $(\mathbb{E}[W_i|U_i,V_i]-W_k)$ and $(W_k-\mathbb{E}[W_j|U_j,V_j])$ given $(U_i,V_i)$ and $(U_j,V_j)$ is zero. By iterated expecations, this implies that the first term of the Hoeffding expansion is zero. The remaining terms of the expansion can be addressed invoking the same arguments as before (i.e. applying Rosenthal's and Markov's inequality with $\kappa=1$) and hence the lead terms are of order $o_p(n^{-\frac{1}{2}})$ uniformly in $t$. This completes the proof of Lemma A1. ■

**Proof of Lemma A2.** The proof is almost identical to the one of Lemma 3.1 in Ghosal et al. (2000). The main difference w.r.t. that proof consists in the fact that $\epsilon_i$ and $\lambda(V_i)$ are not contained inside an indicator function. This requires invoking assumption A1 jointly with the application of Lemma 2.6.15 and 2.6.18 (multiple times) of Van der Vaart and Wellner (1996) to show that the class of functions $\mathcal{P} = \{p(t) : t\in\mathcal{T}\}$ with $p(t) \equiv p((U_1,V_1,\epsilon_1),(U_2,V_2,\epsilon_2);t) = h_{2n}^{-2}(\epsilon_1-\lambda(V_1)-\epsilon_2+\lambda(V_2))\,\mathrm{sign}(U_2-U_1)K_{h2}(U_2-t)K_{h2}(U_1-t)$ is a VC (Vapnik-Cervonenkis) class with the envelope function $Ch_{2n}^{-2}$ and C some positive, finite constant. The remaining steps are identical to the proof of Lemma 3.1. ■

**Proof of Lemma A3.** The result follows by the same arguments as in the proof of Lemma 3.2 of Ghosal et al. (2000), which verifies the conditions of Theorem 1.1 of Rio (1994): thus, define $\widetilde{V}\equiv\lambda(V)$ and let

$$\Phi_{n,t}(\epsilon,\widetilde{V},U) = 2\Big(\epsilon-\widetilde{V}\Big)\int \mathrm{sign}\Big(U-\omega\Big)\frac{1}{h_{2n}^2}K_{h2}(\omega-t)dF(\omega)K_{h2}(U-t).$$

Since Rio's (1994) Theorem 1.1 formally requires that the above function belongs to a VC class of bounded functions that is supported on $[0,1]^3$, invoke assumption A2 to carry out multivariate quantile transformations and conditions A1, A2, and A4 together with Lemma 2.6.15, 2.6.16, and 2.6.18 of Van der Vaart and Wellner (1996) to show that the class is VC. The remaining steps parallel the proof of Lemma 3.2. Also note that the difference in the rate of convergence w.r.t. to the original Lemma 3.2 arises because three ($U_i$, $V_i$, and $\epsilon_i$) instead of two variables are considered. ■

**Proof of Lemma A4.** Assertions (i) and (ii) are identical to the proof of Lemma 3.3 in Ghosal et al. (2000).

To prove (iii), I will start by showing that uniformly in $t\in\mathcal{T}$:

$$\sup_t\Big|\widehat{\sigma}_n^2(t)-\sigma_n^2(t)\Big| = \sup_t\Big|\widetilde{\sigma}_n^2(t)-\sigma_n^2(t)\Big| + o_p(n^{-\frac{1}{2}}h_{2n}^{-2}). \tag{A-10}$$

The second equality of (iii), namely $\sup_t\Big|\widetilde{\sigma}_n^2(t)-\sigma_n^2(t)\Big| = O_p(n^{-\frac{1}{2}}h_{2n}^{-2})$, then follows by steps that are identical to the ones in the proof of Lemma A2 above and Lemma 3.3 of Ghosal et al. (2000). That is, applying these arguments, one can show that:

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

which proves the second equality of (iii).

Since the proof of Equation (A-10) involves almost identical arguments to the ones employed in the proof of Lemma A1, I will only sketch the main differences: firstly, in analogy to the proof of Lemma A1, let $R_{ijk,n}(t) \equiv \text{sign}(U_i - U_j)\,\text{sign}(U_i - U_k) K_{h2}(U_i - t)^2 K_{h2}(U_j - t) K_{h2}(U_k - t)$ and recall that $(\widehat{\epsilon}_i - \widehat{\lambda}(\widehat{V}_i)) = (W_i - \widehat{\mu}(U_i, \widehat{V}_i))$. Then, notice that:

$$\begin{aligned}\sup_t \left|\widehat{\sigma}_n^2(t) - \sigma_n^2(t)\right| =& \sup_t \left| \frac{1}{n(n-1)(n-2)h_{2n}^4} \sum_{\substack{1\leq i,j,k\leq n\\ i\neq j\neq k}} \Big(W_i - \mu(U_i, V_i) + \mu(U_i, V_i) - \widehat{\mu}(U_i, \widehat{V}_i)\Big)^2 R_{ijk,n}(t) - \sigma_n^2(t)\right| \\ \leq& A_{1n}(t) + A_{2n}(t) + A_{3n}(t),\end{aligned} \tag{A-11}$$

where $A_{1n}(t)$ contains $\left(\widehat{\mu}(U_i, \widehat{V}_i) - \mu(U_i, V_i)\right)^2$, $A_{2n}(t)$ contains the cross-product, and $A_{3n}(t)$ is:

$$A_{3n}(t) \equiv \sup_t \left| \frac{1}{n(n-1)(n-2)h_{2n}^4} \sum_{\substack{1\leq i,j,k\leq n\\ i\neq j\neq k}} \Big(W_i - \mu(U_i, V_i)\Big)^2 R_{ijk,n}(t) - \sigma_n^2(t)\right| = \sup_t \left|\widetilde{\sigma}_n^2(t) - \sigma_n^2(t)\right|.$$

I start with $A_{1n}(t)$, which can be decomposed as in Lemma A1 and then bounded by:

$$\sup_t \left| \frac{1}{n(n-1)(n-2)h_{2n}^4} \sum_{\substack{1\leq i,j,k\leq n\\ i\neq j\neq k}} \left\{ \left( \frac{\widehat{s}(U_i, \widehat{V}_i) - \widetilde{s}(U_i, V_i)}{\widetilde{f}(U_i, V_i)} \right)^2 - \left( \frac{\widetilde{f}(U_i, V_i) - \widehat{f}(U_i, V_i)}{\widetilde{f}(U_i, V_i)} \widehat{\mu}(U_i, \widehat{V}_i) \right)^2 \right\} R_{ijk,n}(t) \right| + o_p(1),$$

where $\widehat{s}(\cdot)$, $\widetilde{s}(\cdot)$, $\widehat{f}(\cdot)$, and $\widetilde{f}(\cdot)$ are defined in the proof of Lemma A1 and the term of smaller order holds uniformly in $t$ and contains cross-products. As in the proof of Lemma A1, I focus on the first term: after replacing $\widetilde{f}(U_i, V_i)$ by $f(U_i, V_i)$, which results in an error of smaller order (uniformly in $t$) by Theorem 2 of Hansen (2008) and the bandwidth conditions, one can decompose and expand the first expression as follows:

$$\begin{aligned}\sup_t \Bigg| \frac{1}{n(n-1)(n-2)h_{2n}^4} \sum_{\substack{1\leq i,j,k\leq n\\ i\neq j\neq k}} \Bigg\{ \frac{1}{h_{1n}^2 n} \sum_{l=1}^n \Bigg\{ & \frac{(\widehat{\tau}_{in} - \tau_{in})}{f(U_i, V_i)} W_l \mathbf{K}_{h1,l}^2(U_i, V_i) \\ &+ \frac{\tau_{in} W_l}{f(U_i, V_i) h_{1n}} K_{h1}(U_i - U_l) K_{h1}^{(1)}(\overline{V}_i - \overline{V}_l)\{\widehat{V}_i - V_i + V_k - \widehat{V}_k\} \\ &+ \frac{(\widehat{\tau}_{in} - \tau_{in})}{f(U_i, V_i) h_{1n}} W_l K_{h1}(U_i - U_l) K_{h1}^{(1)}(\overline{V}_i - \overline{V}_l)\{\widehat{V}_i - V_i + V_k - \widehat{V}_k\} \Bigg\}\Bigg\}^2 R_{ijk,n}(t)\Bigg|.\end{aligned} \tag{A-12}$$

In the following, I will focus exclusively on the $n$ square terms since the $n(n-1)$ cross-products can be treated in the same manner. Moreover, I will neglect all cross-products involving the last term of Equation (A-12) with $(\widehat{\tau}_{in} - \tau_{in})$ and $(\widehat{V}_i - V_i + V_k - \widehat{V}_k)$ as these expressions are clearly of smaller order uniformly in $t$. That is, the remaining terms to be examined are:

$$\begin{aligned}&\sup_t \Bigg| \frac{1}{n^2(n-1)(n-2)h_{1n}^2 h_{2n}^4} \sum_{\substack{1\leq i,j,k\leq n\\ i\neq j\neq k}} \sum_{l=1}^n \Bigg\{ \frac{(\widehat{\tau}_{in} - \tau_{in})^2}{f(U_i, V_i)^2} W_l^2 \mathbf{K}_{h1,l}^2(U_i, V_i)^2 R_{ijk,n}(t) \\ &+ \frac{\tau_{in}^2 W_l^2}{f(U_i, V_i)^2 h_{1n}^2} K_{h1}(U_i - U_l)^2 K_{h1}^{(1)}(\overline{V}_i - \overline{V}_l)^2 \{\widehat{V}_i - V_i + V_k - \widehat{V}_k\}^2 R_{ijk,n}(t) \\ &+ \frac{2(\widehat{\tau}_{in} - \tau_{in})}{f(U_i, V_i) h_{1n}} W_l K_{h1}(U_i - U_l) K_{h1}^{(1)}(\overline{V}_i - \overline{V}_l)\{\widehat{V}_i - V_i + V_k - \widehat{V}_k\} R_{ijk,n}(t) \Bigg\}\Bigg| \\ \leq& A_{11n}(t) + A_{12n}(t) + A_{13n}(t).\end{aligned}$$

$A_{11n}(t)$ can be bounded by:

$$
\begin{aligned}
&\sup_t \Big| \frac{1}{n^2(n-1)(n-2)h_{2n}^4 h_{1n}^4} \sum_{\substack{1\le i,j,k\le n \\ i\neq j\neq k}} \sum_{l=1}^{n} \frac{(\widehat{\tau}_{in}-\tau_{in})^2}{f(U_i,V_i)^2} W_l^2 \mathbf{K}_{h1,l}^2(U_i,V_i)^2 R_{ijk,n}(t) \Big| \\
\le & \sup_{U_i,Z_i\in\mathcal{W}_n} \Big|\widehat{\tau}_{in}-\tau_{in}\Big|^2 \sup_t \Big| \frac{1}{n^2(n-1)(n-2)h_{2n}^4 h_{1n}^4} \sum_{\substack{1\le i,j,k\le n \\ i\neq j\neq k}} \sum_{l=1}^{n} \frac{W_l^2 \mathbf{K}_{h1,l}^2(U_i,V_i)^2 R_{ijk,n}(t)}{f(U_i,V_i)^2} \Big| \\
= & A_{111n}(t) \times A_{112n}(t).
\end{aligned}
$$

Using a mean value expansion for $A_{111n}(t)$ and invoking A6 and Theorem 2 of Hansen (2008), one obtains that $A_{111n}(t)$ is of order $O_p(\log(n)c_n^2/nh_{1n}^{1+d_z})$. The second term $A_{112n}(t)$ can, as in the proof of Lemma A1, be written as a fourth order U-process (indexed by $t$) with the symmetric version of $W_l^2\mathbf{K}_{h1,l}^2(U_i,V_i)^2 f(U_i,V_i)^{-2} R_{ijk,n}(t)$ as kernel function $q_n(t)$. Then, invoking A1, A2, A3, A4 together with the properties of the sign and indicator function, one can show, by the same arguments as in the proof of Lemma A1, that the kernel function $q_n(t)$ is Euclidean for a constant envelope. Applying again the Hoeffding decomposition, the leading term of the latter, $h_{1n}^{-4}h_{2n}^{-4}\,\mathbb{E}\Big[q_n(t)\Big]$, is of order $O(h_{2n}^{-1}h_{1n}^{-2})$ uniformly in $t$. Moreover, an application of Lemma 6 of Sherman (1994) yields that also the linear and the higher order terms of the decomposition are Euclidean for (different) constant envelope functions. Hence, letting $\omega_i \equiv \{W_i, U_i, V_i\}$, by the uniform convergence results of Corollary 7 of Sherman (1994), the linear term of the Hoeffding decomposition is of order:

$$
\frac{4}{nh_{1n}^4 h_{2n}^4} \sum_{i=1}^{n} \Big\{ \mathbb{E}\Big[q_n(t)\Big|\omega_i\Big] - \mathbb{E}\Big[\mathbb{E}\Big[q_n(t)\Big|\omega_i\Big]\Big] \Big\} = O_p(n^{-\frac{1}{2}} h_{1n}^{-2} h_{2n}^{-2})
$$

uniformly in $t$, which is of order $o_p(h_{2n}^{-1}h_{2n}^{-2})$ uniformly in $t$ by the bandwidth conditions. Similarily, the other higher order terms of the decomposition of the fourth order U-process are also of order $o_p(h_{2n}^{-1}h_{1n}^{-2})$ uniformly in $t$.

The second expression $A_{12n}(t)$ can be bounded by:

$$
\sup_{Z_i\in\mathcal{Z}_n} \Big|\widehat{g}(Z_i)-g(Z_i)\Big|^2 \sup_t \Big| \frac{1}{n^2(n-1)(n-2)h_{2n}^4 h_{1n}^6} \sum_{\substack{1\le i,j,k\le n \\ i\neq j\neq k}} \sum_{l=1}^{n} \frac{\tau_{in} W_l^2 (K_{h1}(U_i-U_l)K_{h1}^{(1)}(\overline{V}_i-\overline{V}_l))^2 R_{ijk,n}(t)}{f(U_i,V_i)^2} \Big|.
$$

The same arguments to before yield that $\sup_{Z_i\in\mathcal{Z}_n} \Big|\widehat{g}(Z_i)-g(Z_i)\Big|^2 = O_p(\log(n)/nh_{1n}^{d_z})$, while the leading expression of the second term is $O(h_{2n}^{-1}h_{1n}^{-3})$ uniformly in $t$ by the same arguments as for $A_{112n}(t)$ and an application of integration by parts. Given the bandwidth conditions, this is $o_p(n^{-\frac{1}{2}}h_{2n}^{-2})$ uniformly in $t$. Finally, $A_{13n}(t)$, the last term of relevance, can be bounded by:

$$
\begin{aligned}
&\sup_{U_i,Z_i\in\mathcal{W}_n} \Big|(\widehat{\tau}_{in}-\tau_{in})(\widehat{g}(Z_i)-g(Z_i))\Big| \sup_t \Big| \frac{2}{n^2(n-1)(n-2)h_{2n}^4 h_{1n}^5} \sum_{\substack{1\le i,j,k\le n \\ i\neq j\neq k}} \\
&\qquad \times \sum_{l=1}^{n} \frac{\tau_{in} W_l^2 K_{h1}(U_i-U_l)^2 K_{h1}(V_i-V_l) K_{h1}^{(1)}(\overline{V}_i-\overline{V}_l) R_{ijk,n}(t)}{f(U_i,V_i)^2} \Big|,
\end{aligned}
$$

where the first part can be shown to be of order $O_p(\log(n)c_n/nh_{1n}^{\frac{1}{2}+d_z})$, while the leading term of the second expression is of order $O(h_{2n}^{-1}h_{1n}^{-2})$ uniformly in $t$. Hence, by the bandwidth conditions, the product is of order $o_p(n^{-\frac{1}{2}}h_{2n}^{-2})$ uniformly in $t$.

Finally, $A_{2n}(t)$ from Equation (A-11) can be shown to be of order $O_p(n^{-\frac{1}{2}}) = o_p(n^{-\frac{1}{2}}h_{2n}^{-2})$ uniformly in $t$ using arguments that are identical to the ones used in the proof of Lemma A1. ■

**Proof of Lemma A5.** The proof is identical to the one of Lemma 3.4 in Ghosal et al. (2000), making also use of Remark 8.3 in their paper, so only the starting point will be highlighted: let $\mathcal{G}_n$ denote the class of functions

$\{g_{n,t}(\epsilon, V, U) : t \in \mathcal{T}\}$, where $g_{n,t}(\epsilon, V, U) = \frac{G_n(t)}{\sigma_n(t)}$ with $G_n(t)$ defined in Lemma A3 and $\sigma_n(t)$ defined in the main text. Furthermore, let $\overline{\mathcal{G}}_n$ stand for the class of functions $\{\overline{g}_{n,t} : t \in \mathcal{T}\}$ with $\overline{g}_{n,t}(\epsilon, V, U) = \frac{\overline{\Psi}_{n,t}(\epsilon,V,U)}{\overline{\sigma}_n(t)}$ where:

$$\overline{\Psi}_{n,t}(\epsilon, V, U) = \int \Big(\epsilon - \lambda(V)\Big) \operatorname{sign}\Big(U - \omega\Big) \frac{1}{h_{2,n}^2} K_{h2}(\omega - t) d\omega K_{h2}(U - t)$$

and

$$\overline{\sigma}_n(t) = \bigg(\int \bigg(\int \Big(\overline{\epsilon} - \lambda(\overline{V})\Big) \operatorname{sign}\Big(\overline{U} - \omega\Big) \frac{1}{h_{2,n}} K_{h2}(\omega - t) d\omega\bigg)^2 \frac{1}{h_{2,n}^2} K_{h2}^2(\overline{U} - t) d\overline{\epsilon} d\overline{V} d\overline{U}\bigg)^{\frac{1}{2}} f(\epsilon, V, U)^{\frac{1}{2}}.$$

Now apply steps that are analogous to the proof of Lemma 3.4. Notice that in order to obtain the covariance $\rho_t(\cdot)$, whose level does not depend on $t \in \mathcal{T}$, the large support condition A5 needs to be invoked to ensure that:

$$\frac{\int (\epsilon_1 - \lambda(V_1))(\epsilon_2 - \lambda(V_2)) f(V_1, \epsilon_1 | t_1)^{\frac{1}{2}} f(V_2, \epsilon_2 | t_2)^{\frac{1}{2}} dV_1 dV_2 d\epsilon_1 d\epsilon_2}{\{\int (\epsilon_1 - \lambda(V_1))^2 f(V_1, \epsilon_1 | t_1) dV_1 d\epsilon_1\}^{\frac{1}{2}} \{\int (\epsilon_2 - \lambda(V_2))^2 f(V_2, \epsilon_2 | t_2) dV_2 d\epsilon_2\}^{\frac{1}{2}}} = 1$$

for any $t_1, t_2 \in \mathcal{T}$. ■

# Appendix A2

Table 1: Monte Carlo Results

| Asymptotic Expansion $F_n$, $\alpha = .05$ | | | | | |
|---|---|---|---|---|---|
| | | M1 | M2 | M3 | M4 |
| $h_{1n} = 0.8n^{-\frac{1}{4}}$ $h_{2n} = 0.8n^{-\frac{1}{4}}$ | $n = 200$ | 0.02667 | 0.61467 | 0.75467 | 0.97600 |
| | $n = 300$ | 0.02933 | 0.68467 | 0.84867 | 1.00000 |
| | $n = 400$ | 0.02800 | 0.72667 | 0.92800 | 1.00000 |
| $h_{1n} = 0.9n^{-\frac{1}{4}}$ $h_{2n} = 0.9n^{-\frac{7}{24}}$ | $n = 200$ | 0.04667 | 0.58467 | 0.78067 | 0.94067 |
| | $n = 300$ | 0.04733 | 0.65933 | 0.87000 | 1.00000 |
| | $n = 400$ | 0.04700 | 0.71267 | 0.90800 | 1.00000 |
| $h_{1n} = 2.34\,\mathrm{sd}(\cdot)n^{-\frac{1}{5}}$ $h_{2n} = 0.8n^{-\frac{1}{4}}$ | $n = 200$ | 0.03467 | 0.59133 | 0.75800 | 0.88400 |
| | $n = 300$ | 0.02333 | 0.67200 | 0.87133 | 0.99867 |
| | $n = 400$ | 0.02367 | 0.72600 | 0.94800 | 1.00000 |
| Asymptotic Distribution $F_\infty$, $\alpha = .05$ | | | | | |
| | | M1 | M2 | M3 | M4 |
| $h_{1n} = 0.8n^{-\frac{1}{4}}$ $h_{2n} = 0.8n^{-\frac{1}{4}}$ | $n = 200$ | 0.00467 | 0.30333 | 0.49267 | 0.87267 |
| | $n = 300$ | 0.00800 | 0.40800 | 0.63667 | 0.99867 |
| | $n = 400$ | 0.00600 | 0.47467 | 0.76600 | 1.00000 |
| $h_{1n} = 0.9n^{-\frac{1}{4}}$ $h_{2n} = 0.9n^{-\frac{7}{24}}$ | $n = 200$ | 0.01467 | 0.29333 | 0.54267 | 0.81600 |
| | $n = 300$ | 0.01133 | 0.39667 | 0.70200 | 0.99800 |
| | $n = 400$ | 0.01200 | 0.45067 | 0.77333 | 1.00000 |
| $h_{1n} = 2.34\,\mathrm{sd}(\cdot)n^{-\frac{1}{5}}$ $h_{2n} = 0.8n^{-\frac{1}{4}}$ | $n = 100$ | 0.00600 | 0.27600 | 0.49000 | 0.67933 |
| | $n = 200$ | 0.00533 | 0.39467 | 0.66733 | 0.98867 |
| | $n = 500$ | 0.00600 | 0.46867 | 0.79733 | 1.00000 |

Figure 1: Null Hypothesis False

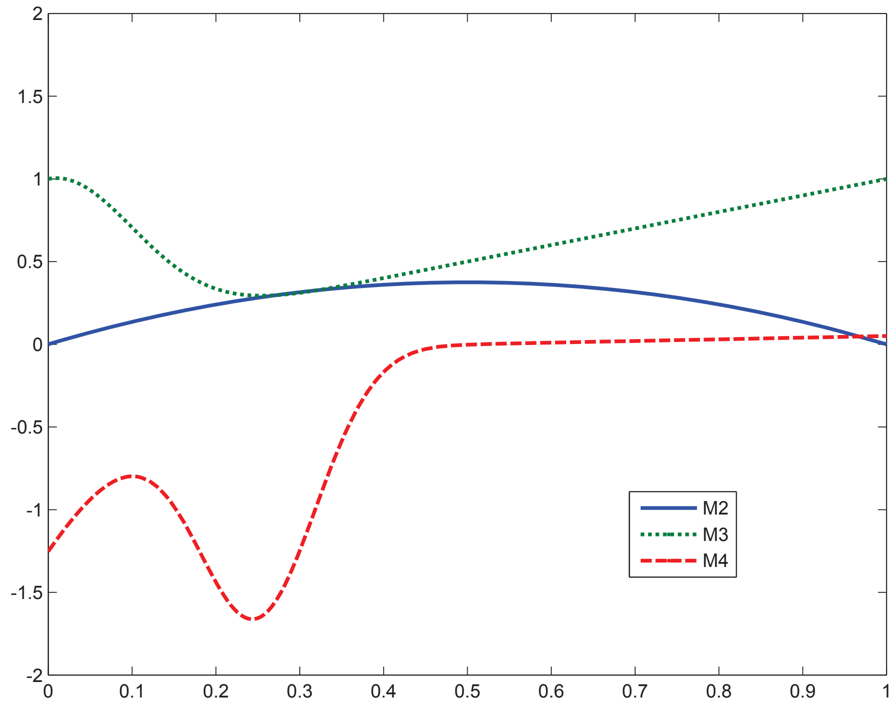


Table 2: Summary Statistics

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Hrly. Res. Wage in Pounds | 4.98 | 2.55 | 1 | 30 |
| Elapsed Duration in Weeks | 19.15 | 15.76 | 0 | 80 |
| Completed Duration in Weeks | 59.06 | 76.09 | 4 | 624 |
| Gender (0: Female 1: Male) | 0.54 | 0.50 | 0 | 1 |
| Unemployed (0: Inactive 1: Unemployed) | 0.73 | 0.44 | 0 | 1 |
| Weekly Unempl. Benefits (Individual-level)[1] | 37.58 | 36.28 | 0 | 138.73 |
| Weekly Other Benefits (Household-level) | 142.99 | 144.20 | 0 | 1042.20 |

[1] Weekly rates computed according to www.ifs.org.uk/fiscalFacts/benefitTables.

Figure 2: Sample I - Estimated Reservation Wage Function

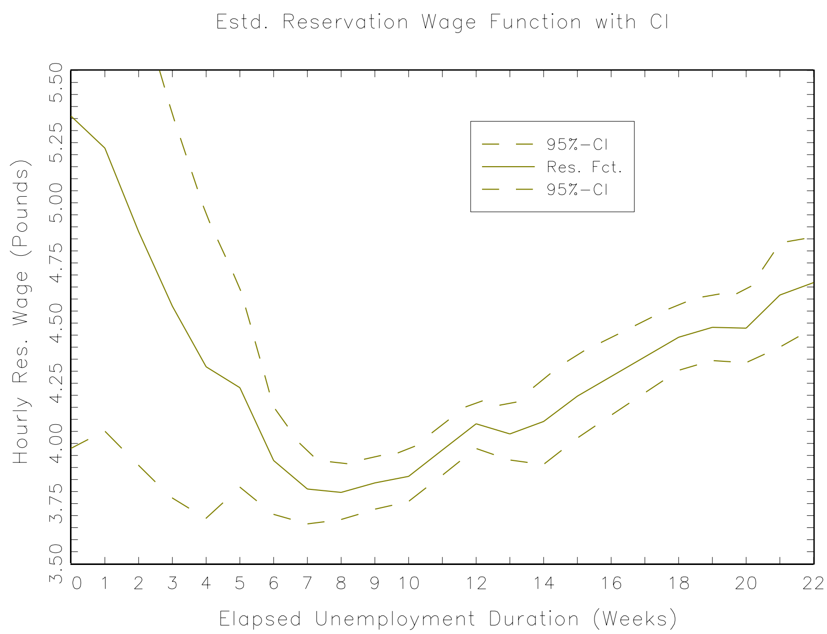


Figure 3: Sample II - Estimated Reservation Wage Function

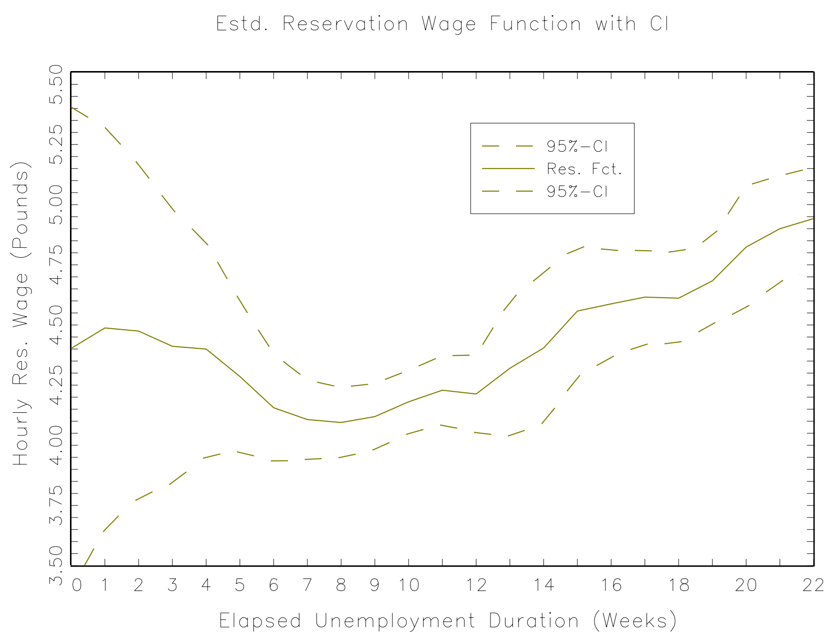

Figure 4: Sample III - Estimated Reservation Wage Function

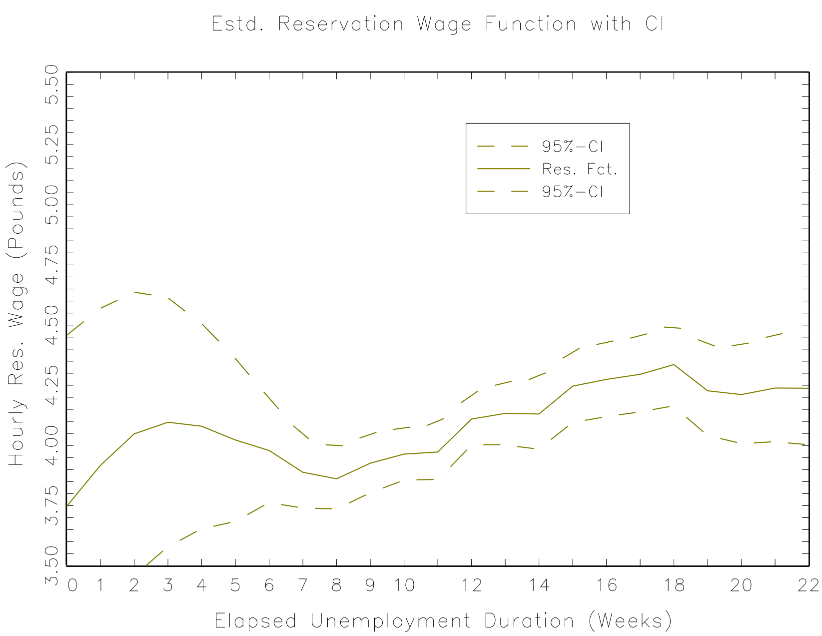

Table 3: Test Results - $\mathcal{T}_1 = [1, 7]$

| Asymptotic Expansion $F_n$, $\alpha = .05$ | | | | | |
|---|---|---|---|---|---|
| | $C \cdot h_{2n}$ | CV 1% | CV 5% | CV 10% | Test Statistic |
| Sample I | $C = 0.5$ | 3.3254 | 2.7929 | 2.5216 | 0.6890 |
| | $C = 0.75$ | 3.2009 | 2.6432 | 2.3561 | 0.4450 |
| | $C = 1$ | 3.1103 | 2.5330 | 2.2305 | $-0.4047$ |
| Sample II | $C = 0.5$ | 3.2802 | 2.7390 | 2.4620 | 2.0909 |
| | $C = 0.75$ | 3.1542 | 2.5870 | 2.2912 | 1.8403 |
| | $C = 1$ | 3.0623 | 2.4723 | 2.1629 | 0.7577 |
| Sample III | $C = 0.5$ | 3.3091 | 2.7734 | 2.5007 | 2.0799 |
| | $C = 0.75$ | 3.2079 | 2.6518 | 2.3647 | 2.3321 |
| | $C = 1$ | 3.1167 | 2.5404 | 2.2398 | 1.6828 |
| Asymptotic Distribution $F_\infty$, $\alpha = .05$ | | | | | |
| | $C \cdot h_{2n}$ | CV 1% | CV 5% | CV 10% | Test Statistic |
| Sample I | $C = 0.5$ | 4.7372 | 3.5418 | 3.0139 | 0.6890 |
| | $C = 0.75$ | 5.5168 | 3.9248 | 3.2218 | 0.4450 |
| | $C = 1$ | 7.3770 | 5.0068 | 3.9601 | $-0.4047$ |
| Sample II | $C = 0.5$ | 4.9319 | 3.6270 | 3.0507 | 2.0909 |
| | $C = 0.75$ | 6.1801 | 4.2970 | 3.4654 | 1.8403 |
| | $C = 1$ | 11.4495 | 7.5404 | 5.8140 | 0.7577 |
| Sample III | $C = 0.5$ | 4.799 | 3.5677 | 3.0238 | 2.0799 |
| | $C = 0.75$ | 5.4478 | 3.8877 | 3.1987 | 2.3321 |
| | $C = 1$ | 7.1213 | 4.8526 | 3.8507 | 1.6828 |

Table 4: Test Results - $\mathcal{T}_2 = [1, 22]$

| Asymptotic Expansion $F_n$, $\alpha = .05$ | | | | | |
|---|---|---|---|---|---|
| | $C \cdot h_{2n}$ | CV 1% | CV 5% | CV 10% | Test Statistic |
| Sample I | $C = 0.75$ | 3.5715 | 3.0813 | 2.9461 | 3.3712 |
| | $C = 1$ | 3.4896 | 2.9867 | 2.8847 | 5.1052 |
| | $C = 1.25$ | 3.4250 | 2.9110 | 2.8481 | 5.5290 |
| Sample II | $C = 0.75$ | 3.5291 | 3.0324 | 2.9127 | 4.7396 |
| | $C = 1$ | 3.4464 | 2.9357 | 2.8587 | 4.5652 |
| | $C = 1.25$ | 3.3815 | 2.8588 | 2.8327 | 4.5092 |
| Sample III | $C = 0.75$ | 3.5775 | 3.0882 | 2.9511 | 2.6987 |
| | $C = 1$ | 3.4958 | 2.9936 | 2.8889 | 3.7924 |
| | $C = 1.25$ | 3.4318 | 2.9184 | 2.8511 | 4.1737 |
| Asymptotic Distribution $F_\infty$, $\alpha = .05$ | | | | | |
| | $C \cdot h_{2n}$ | CV 1% | CV 5% | CV 10% | Test Statistic |
| Sample I | $C = 0.75$ | 4.3254 | 3.4607 | 3.0789 | 3.3712 |
| | $C = 1$ | 4.3913 | 3.4469 | 3.0298 | 5.1052 |
| | $C = 1.25$ | 4.4821 | 3.4578 | 3.0055 | 5.5290 |
| Sample II | $C = 0.75$ | 4.3539 | 3.4505 | 3.0515 | 4.7396 |
| | $C = 1$ | 4.4472 | 3.4514 | 3.0117 | 4.5652 |
| | $C = 1.25$ | 4.5727 | 3.4820 | 3.0003 | 4.5092 |
| Sample III | $C = 0.75$ | 4.3221 | 3.4627 | 3.0831 | 2.6987 |
| | $C = 1$ | 4.3846 | 3.4470 | 3.0329 | 3.7924 |
| | $C = 1.25$ | 4.4713 | 3.4556 | 3.0071 | 4.1737 |

**Acknowledgements:** Insightful comments from three anonymous referees have led to a greatly improved version of this paper. I would also like to thank Valentina Corradi and Wiji Arulampalam for various helpful discussions. Comments received by Rosa Matzkin, Steve Pudney, and Marcelo Fernandes as well as seminar participants at the ASSA meeting 2013 in San Diego, the University of Freiburg and Oxford are gratefully acknowledged, too. Finally, I would like to thank Jennifer Smith for her advice on the construction of the data set. All remaining errors are my sole responsibility.

# References

Addison, J.T., Centeno, M., Portugal, P., 2004. Reservation wages, search duration, and accepted wages in europe. Discussion Paper 1252. IZA Bonn.

Ahn, H., Powell, J.L., 1993. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. Journal of Econometrics 58, 3–29. URL: `http://www.sciencedirect.com/science/article/B6VC0-4582D09-2C/2/91e44e99e7e4ea61fee4284e0b299550`.

Van den Berg, G., 1990. Nonstationarity in job search theory. Review of Economic Studies 57, 255–277.

Birke, M., Neumeyer, N., 2013. Testing monotonicity of regression functions - an empirical process approach. Scandinavian Journal of Statisitics 40, 438–454.

Blundell, R., Matzkin, R.L., 2010. Conditions for the existence of control functions in nonseparable simultaneous equations models. Note. UCLA.

Blundell, R., Powell, J.L., 2003. Endogeneity in nonparametric and semiparametric regression models, in: Dewatripont, M., Hansen, L., Turnovsky, S. (Eds.), Endogeneity in nonparametric and semiparametric regression models. Cambridge University Press, Cambridge, UK. volume 2, pp. 312–357. Eigth World Congress.

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

Brown, S., Taylor, K., 2009. Reservation wages, expected wages and the duration of unemployment: Evidence from british panel data. Discussion Paper 3981. IZA Bonn.

Chetverikov, D., 2012. Testing regression monotonicity in econometric models. Job Market Paper. Massachusetts Institute of Technology, Cambridge, MA.

Delgado, M.A., Escanciano, J.C., 2012. Distribution-free tests of stochastic monotonicity. Journal of Econometrics 170, 68–75.

Durot, C., 2003. A kolmogorov-type test for monotonicity of regression. Statistics and Probability Letters 63, 425–433.

Einmahl, U., Mason, D., 2005. Uniform in bandwidth consistency of kernel-type function estimators. Annals of Statistics 33, 1380–1403.

Ellison, G., Ellison, S.F., 2011. Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. American Economic Journal: Microeconomics 3, 1–36.

Fermanian, J.D., Salanie, B., 2004. A nonparametric simulated maximum likelihood estimation method. Econometric Theory 20, 701–734.

Garen, J., 1984. The returns to schooling: A selectivity bias approach with a continuous choice variable. Econometrica 52, 1199–1218.

Ghosal, S., Sen, A., van der Vaart, A.W., 2000. Testing monotonicity of regression. Annals of Statistics 28, 1054–1092.

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

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

Hansen, B.E., 2008. Uniform convergence rate for kernel estimation with dependent data. Econometric Theory 24, 726–748.

Hansen, B.E., 2012. Econometrics. First-Year PhD Textbook, University of Wisconsin.

Hoderlein, S., Su, L., White, H., 2011. Specification testing for nonparametric structural models with monotonicity in unobservables. Research Collection. University of California, San Diego.

Imbens, G.W., Newey, W.K., 2009. Identification and estimation of triangular simultaneous equations models without additivity. Econometrica 77, 1481–1512.

Jones, M., Signorini, D., 1997. A comparison of higher-order bias kernel density estimators. Journal of the American Statistical Association 92, 1063–1073.

Kiefer, N., Neumann, G., 1979. A empirical job-search model, with a test of the constant reservation wage hypothesis. The Journal of Political Economy 87, 89–107.

Lancaster, T., 1985. Simultaneous equations models in applied search theory. Journal of Econometrics 28, 113–126. URL: http://www.sciencedirect.com/science/article/B6VC0-45828X7-5R/2/0d4f6ac9638061c1af01a1406db2e64a.

Lee, S., Linton, O., Whang, Y.J., 2009. Testing for stochastic monotonicity. Econometrica 77, 585–602.

Mortensen, D., 1986. Job search and labor market analysis, in: Ashenfelter, O., Layard, R. (Eds.), Job search and labor market analysis. North-Holland, Amsterdam.

Newey, W.K., Powell, J.L., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equations models. Econometrica 67, 565–603.

Nolan, D., Pollard, D., 1987. U-processes: Rates of convergence. The Annals of Statistics 15, 780–799. URL: http://www.jstor.org/stable/2241339.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57, 1027–1057. URL: http://www.jstor.org/stable/1913622.

Paull, G., 2002. Biases in the reporting of labour market dynamics. Working Paper 02/10. Institute for Fiscal Studies, London.

Petrov, V.V., 1995. Limit theorems of probability theory. Number 4 in Oxford Studies in Probability, Oxford University Press.

Piterbarg, V., 1996. Asymptotic methods in the theory of gaussian processes and fields. volume 148 of *Translation of Mathematical Monographs*. American Mathematical Society, Providence, Rhode Island.

Rio, E., 1994. Local invariance principles and their application to density estimation. Probability Theory and Related Fields 98, 21–45.

Sherman, R.P., 1994. Maximal inequalities for degenerate u-processes with applications to optimization estimators. The Annals of Statistics 22, 439–459. URL: http://www.jstor.org/stable/2242462.

Smith, J., 2011. The ins and outs of uk unemployment. The Economic Journal 121, 402–444.

Van der Vaart, A.W., Wellner, J., 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.

Wang, J., Meyer, M.C., 2011. Testing the monotonicity or convexity of a function using regression splines. The Canadian Journal of Statistics 39, 89–107.