SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

**CREATES Research Paper 2011-13**

# Nonparametric Detection and Estimation of Structural Change

**Dennis Kristensen**

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

# Nonparametric Detection and Estimation of Structural Change*

Dennis Kristensen†
Columbia University and CREATES

18th April 2011

**Abstract**

We propose a nonparametric approach to the estimation and testing of structural change in time series regression models. Under the null of a given set of the coefficients being constant, we develop estimators of both the nonparametric and parametric components. Given the estimators under null and alternative, generalized $F$ and Wald tests are developed. The asymptotic distributions of the estimators and test statistics are derived. A simulation study examines the finite-sample performance of the estimators and tests. The techniques are employed in the analysis of structural change in US productivity and the Eurodollar term structure.

JEL Classification: C12, C13, C14, C22.

Keywords: structural change, regression, nonparametric, estimation, testing, generalized likelihood ratio, time-varying, locally stationary.

*A previous version of this paper was circulated under the title "Semi-nonparametric Estimation and Testing of Time-varying Regression Models." I would like to thank Bruce Hansen for sharing data and Matlab code with me and for his valuable comments and suggestions. I would also like to thank Andrew Ang, Jianqin Fan, Ulrich Müller and Anders Rahbek for stimulating discussions, and participants at seminars at Carleton University, University of Montreal and Princeton University, and at the Society for Nonlinear Dynamics and Econometrics 18th Annual Symposium at the University of Piemonte Orientale for comments and suggestions. This research was supported by the Danish National Science Foundation (through a grant to CREATES) and the National Science Foundation (grant no. SES-0961596). Parts of this research was conducted while the author was visiting Princeton University and University of Copenhagen whose hospitality is gratefully acknowledged.

†E-mail: dk2313@columbia.edu. Department of Economics, Columbia University, 420 West 118th Street, New York, NY 10027.

# 1 Introduction

There is amble empirical evidence of structural changes in many economic and financial time series such as GDP (McConnell and Perez-Quiros, 2000), interest rates (Stock and Watson, 1996), labour productivity (Hansen, 2001), and stock returns (Ang and Kristensen, 2009). Neglecting these changes in the analysis of data can lead to spurious conclusions. This has lead to a large literature on detection and estimation of structural changes in time series regression models. Most studies assume a fully parametric structure of time variation in parameters. This has the advantage that the model maintains much of its parsimonious structure. The disadvantage is that the researcher runs the risk of choosing a misspecied model. This in turn may lead to misleading conclusions being drawn from the fitted time-varying specification.

This paper proposes a general methodology for nonparametric estimation and testing of time-varying coefficients in a linear regression model with heteroskedastic errors. We impose no parametric structure on neither the regression coefficients nor the conditional variance and instead estimate both components nonparametrically. This way, the risk of mispecification is smaller and so more robust inference can be conducted. We consider the null of a given (sub)set of the regression coefficients being constant, and develop estimators under null and alternative. The estimators take the form of simple kernel-weighted OLS estimators and so are very simple to implement in contrast to existing parametric estimators whose implementation can be computationally burdensome. We propose to test the null by comparing the two sets of estimators through a generalized likelihood-ratio test statistic. We also show how the proposed methods can be used as guidance in the search and estimation of a parsimonious parametric model of structural change.

We derive the asymptotic properties of the estimators and test statistics: All estimators follow normal distributions in large samples. In particular, under the null, the parametric (constant) components can be estimated with standard parametric rate, and can be made asymptotically efficient. The proposed test statistics are also shown to follow normal distributions under the null, and by suitable choice of weighting functions entering the tests they can be made nuisance parameter-free. These are attractive features when compared to standard parametric estimators and test statistics that tend to suffer from non-standard, non-pivotal distributions thereby further complicating inference in a parametric setting.

Our framework allows for both deterministically and randomly changing parameters, and as such allow for a rich class of data generating processes, incl. random walk type dynamics in the parameters. Moreover, while we here mainly focus on the case of smoothly changing parameters, our methodology also allows for discontinuous breaks in the parameter values over time. As such, our estimators and tests are very robust and should be able to detect structural change under may different scenarios. This is supported by a simulation study that reveals that our estimators and tests have good finite sample properties both when the parameters follow a random walk and a smooth transition. In particular, the performance of the estimators of the constant components are comparable to the oracle estimators assuming full knowledge of the structural changes. Moreover, the tests have precise size properties and exhibit strong power against both random walk and smooth transition alternatives.

The usefulness and simplicity of the proposed estimation and testing strategy are demonstrated through two empirical applications: In the first one, we investigate structural instabilities in US productivity within an autoregressive setting. We find strong evidence of structural changes in the AR coefficient while there is weaker support for variation in the intercept. We test two parametric structural breaks models against the nonparametric alternative, and accept the more general model with three breaks while there is mixed evidence for the simpler one with one break. In the second application, we analyze structural changes in an affine three-factor model for the Eurodollar term structure. We find substantial time variation in all factor loading over the period 1971-2004 and reject the null of constant loadings both individually and when tested in pairs. The variation in the loadings is found to be partially driven by underlying macro factors.

Our approach to the modelling of the changing regression coefficients, which we denote $\beta_t$, are based on the idea of Robinson (1989) who treats the estimation of the unstable parameter paths as a nonparametric curve fitting problem. This approach has also been pursued in more work such as Cai (2007) and Orbe et al. (2004). However, these studies focus solely on the nonparametric estimation of the changing parameters, and do not consider estimation and testing of constant parameters. Our approach also shares similarities with the literature on inference in nonparametric varying-coefficient models as developed in a cross-sectional setting; see Fan and Zhang (2008) for an overview. Our estimators are also related to rolling-window type estimators widely used in empirical finance as an informal way of estimating unstable regressions; see e.g. Ferson and Schadt (1996). Our theoretical analysis provide a formal asymptotic theory allowing for valid inference based on these estimators; see Ang and Kristensen (2009) for a further discussion.

There is a large literature on parametric testing of structural change in regression models. One popular way to parametrically specify time-variation in the regression coefficients is through deterministic breaks; see, among others, Andrews and Ploberger (1994) and Bai and Perron (1998). Other modelling approaches include the smooth transition models (Lin and Teräsvirta, 1994), hidden Markov models (Akharif and Hallin, 2003, Hamilton, 1992 and Hansen, 1992) and threshold models (Chan, 1990) and Hansen, 2000a). Elliott and Müller (2006) show that under regularity conditions, all parametric tests within a certain class of breaking processes are equivalent. They however restrict their analysis to a class of global alternatives and one single break, and as such do not allow for so-called local (or high-frequency) alternatives. This is an important point since parametric tests are in general less powerful at detecting such alternatives compared to Neymann-type tests; see e.g. Eubank and LaRiccia (1992) and Fan et al (2001). We conjecture that the optimality results of Fan et al (2001) can be extended to our setting in which case our tests are able to detect local alternatives at an optimal rate. Finally, even if a given parametric test is able to detect structural changes, it will only deliver a consistent estimator of the process $\beta_t$ if the alternative is correctly specified. Thus, parametric procedures may not be very informative about the type of variation in $\beta_t$ under the alternative, and so do not deliver any robust guidance in modelling potential time variation.

Other nonparametric tests can be found in the literature with the CUSUM test of Brown et al. (1975) being the most prominent. However, this procedure does not deliver any estimates of the breaking process under the alternative and has a non-standard asymptotic distribution. These tests furthermore involve integration/summation over the changing parameters and as such may suffer from the same problem as parametric tests, namely that they cannot detect high-frequency alternatives very well. These less appealing features are shared by many other nonparametric tests such as Chu et al. (1995).

Our testing approach is instead most closely related to the work by Chen and Hong (2009) and Juhl and Xiao (2005) who also develop kernel-based tests for stability in regression models using a strategy similar to ours. Juhl and Xiao (2005) focus on models where only the intercept is potentially time-varying, while Chen and Hong (2009) develop estimators and tests under the null of *all* regression coefficients being constant. We here extend their results in a number of directions: First, we develop estimators and tests under the hypothesis of time invariance for any given subset of the regression coefficients. This is an important extension since it is often of interest to identify which regressors have unstable coefficients (see e.g. Ang and Kristensen, 2009). Secondly, we allow for heteroskedastic errors and modify estimators and test statistics to handle this. Thirdly, we accomodate for non-stationary (but mixing) regressors; this is important since we thereby can handle autoregressive models which are excluded from the theory in Chen and Hong (2009).

The remains of the paper is organized as follows: In the next section, we introduce our model and develop the proposed estimator and test statistics. Section 3 contains theoretical results for these, while Section 4 gives some extensions. Bandwidth selection and bootstrapping is discussed in Section 5. The results of a simulation study is presented in Section 6, while the two empirical applications can be found in in Section 7. Section 8 concludes. All proofs and lemmas have been relegated to the Appendix.

## 2 Framework

Suppose we have observed $(y_t, X_t)$, $t = 1, ..., n$, from the following regression model:

$$y_t = \beta_t' X_t + \sigma_t z_t. \tag{2.1}$$

Here, $y_t \in \mathbb{R}$ is the dependent variable, $X_t \in \mathbb{R}^m$ is a set of regressors, and $z_t \in \mathbb{R}$ is the normalized error term which satisfies

$$E[z_t|X_t, \beta_t, \sigma_t] = 0, \quad E[z_t^2|X_t, \beta_t, \sigma_t] = 1. \tag{2.2}$$

As such, $\sigma_t^2 > 0$ represents the conditional variance of $y_t$ while the regression coefficients $\beta_t$ can be expressed as

$$\beta_t = E[X_t X_t'|\beta_t, \sigma_t]^{-1} E[X_t y_t|\beta_t, \sigma_t].$$

In a standard regression model, it is assumed that the regression coefficients and the variance are constant over time, $\beta_t = \beta$ and $\sigma_t^2 = \sigma^2$.

We are interested in testing the hypothesis that (part of) the regression coefficients are in fact constant over time, and also in obtaining estimates both under this null and its alternative. To be specific, let $X_{1,t} \in \mathbb{R}^{m_1}$ denote the set of regressors whose associated slope parameters we are interested in testing for time invariance. The remaining regressors are collected in $X_{2,t} \in \mathbb{R}^{m_2}$ whose regression coefficients may potentially be unstable. We can then write the complete set of regressors as $X_t = (X_{1,t}, X_{2,t})$ with $m_1 + m_2 = m$. With these definitions, the model can be written as:

$$y_t = \beta_{1,t}' X_{1,t} + \beta_{2,t}' X_{1,t} + \sigma_t z_t, \tag{2.3}$$

and we then interested in testing he following null hypothesis,

$$H_0 : \beta_{1,t} = \beta_1 \in \mathbb{R}^{m_1},$$

against the maintained (alternative) hypothesis

$$H_A : \beta_{1,t} \text{ and } \beta_{2,t} \text{ are time-varying.}$$

The above framework is quite standard in the literature on structural changes in regression models. However, in order to develop statistical estimators and tests, most studies now proceed to impose parametric assumptions on the parameter sequences $\beta_t$ and/or $\sigma_t^2$. One popular way of modelling the variation is through deterministic breaks, see e.g. Andrews and Ploberger (1994), Bai (1999) and Bai and Perron (1998). In the simplest case, with two breaks, the dynamics of the regression coefficients are modelled as $\beta_t = \bar{\beta}_1$ for $t = 1, ..., [\pi n]$, and $\beta_t = \bar{\beta}_2$ for $t = [\pi n] + 1, ..., n$ for some (unknown) $\pi \in (0, 1)$ and $\bar{\beta}_1, \bar{\beta}_2 \in \mathbb{R}^m$. This can be written more compactly as $\beta_t = \bar{\beta}_1 \mathbb{I}\{t \leq [\pi n]\} + \bar{\beta}_2 \mathbb{I}\{t > [\pi n]\}$, where $\mathbb{I}\{\cdot\}$ denotes the indicator function. Another widely used specification is the smooth transition model of Lin and Teräsvirta (1994) where the variation is specified as $\beta_t = \bar{\beta}_1 F(t/n; \gamma) + \bar{\beta}_2 [1 - F(t/n; \gamma)]$ for some parametric family of cdf's, $F(t; \gamma)$. While these two models impose a deterministic model on the dynamics of the coefficients, another approach is to model $\beta_t$ as a stochastic process; see e.g. Akharif and Hallin (2003), Hansen (1992), and Hamilton (1992). Finally, some models specify $\beta_t$ as a parametric function of observables, for example threshold models as treated in Chan (1990) and Hansen (2000a).

We here take an alternative approach and do not impose any such parametric restrictions on the nature of the time-variation, and instead build nonparametric estimators of the relevant parameters. However, some additional restrictions has to be imposed on the type of time-variatio in order to make any further progress. In particular, at the current level of generality, we are not able to nonparametrically identify the sequences $\beta_t$ and $\sigma_t^2$, $t = 1, ..., n$, as we have as many parameters as observations. We here resolve this problem by imposing the following rescaling

$$\beta_t = \beta\left(t/n\right), \quad \sigma_t^2 = \sigma^2\left(t/n\right), \tag{2.4}$$

for some functions $\beta : [0,1] \mapsto \mathbb{R}^m$ and $\sigma^2 : [0,1] \mapsto \mathbb{R}_+$ We here use $\beta$ and $\sigma^2$ to denote both functions and the corresponding sequences; this should hopefully not cause any confusion. This restriction on coefficients imply that as the sample size grows, a growing number of observations carry information regarding the variation in the coefficients in any given neighbourhood of the normalized time domain. This will allow us to identify the functions and thereby the parameter sequences.

The above assumption is a standard one in the literature on time-varying parameters, and is also imposed in, for example, the analysis of structural break estimators. We note that the class of models satisfying in eq. (2.4) is rich enough to include many of the parametric models discussed earlier. Clearly, the structural break and the smooth transtion models are contained. Moreover, hidden Markov models $\beta_t$ is modelled as a latent random process can be approximated by our model by choosing the function $\beta\left(\cdot\right)$ in eq. (2.4) as the corresponding continuous-time equivalent. For example, the random walk model can be approximated by letting $\beta\left(\tau\right)$, $\tau \in [0,1]$, be the realized trajectory of a Brownian motion.

The above rescaling was also used in Robinson (1989), who proposed to use kernel methods to nonparametrically estimate time-varying coefficients; see Cai (2007) and Orbe et al. (2004) for some extensions. In an autoregressive setting, the above scaling leads to so-called locally stationary models as analyzed in Dahlhaus (1997).

## 2.1 Estimation

We first develop estimators under the null and alternative. In order to motivate our nonparametric estimators under the alternative $H_A$, suppose that $\varepsilon_t | X_t \sim N\left(0, \sigma_t^2\right)$; we will however not impose this restriction when deriving theoretical properties. In this case, the *global* likelihood takes the form

$$L_n(\beta, \sigma^2) = -\frac{1}{2n} \sum_{t=1}^{n} \left\{ \log\left(\sigma_t^2\right) + \frac{\varepsilon_t^2(\beta_t)}{\sigma_t^2} \right\},$$

for any sequences $\left\{\beta_t, \sigma_t^2 : t \geq 1\right\}$, where $\varepsilon_t(\beta)$ is the residual,

$$\varepsilon_t(\beta) = y_t - \beta' X_t. \tag{2.5}$$

Let $\tau = t_0/n \in (0,1)$ denote a given (normalized) point in time. We define the *local* Gaussian log-likelihood at $\tau$ by

$$L_n^{\text{local}}\left(\beta, \sigma^2|\tau\right) = -\frac{1}{2}\sum_{t=1}^{n}\left\{\log\left(\sigma^2\right) + \frac{\varepsilon_t^2(\beta)}{\sigma^2}\right\} K_h\left(t/n - \tau\right),$$

for any constants $\beta \in \mathbb{R}^m$ and $\sigma^2 > 0$. Here, $K_h(z) = K(z/h)/h$ with $K(\cdot)$ being a kernel and $h > 0$ a window width. The kernel weights $K_h(t/n - \tau)$, $t = 1, ..., n$, determine how we use information around the time point $\tau$ to learn about $\beta(\tau)$ and $\sigma^2(\tau)$. As the time window shrinks to zero, $h \to 0$, only observations very close in time to $\tau$ are used while as $h \to \infty$, all observations are used.

We then propose to estimate $\left(\beta(\tau), \sigma^2(\tau)\right)$ by maximizing the local likelihood at $\tau$,

$$(\hat{\beta}(\tau), \hat{\sigma}^2(\tau)) = \arg\max_{(\beta,\sigma^2)} L_n^{\text{local}}\left(\beta, \sigma^2|\tau\right).$$

Solving $\partial L_n^{\text{local}}\left(\beta, \sigma^2|\tau\right)/\partial\beta = 0$ and $\partial L_n^{\text{local}}\left(\beta, \sigma^2|\tau\right)/\partial\sigma^2 = 0$, we find that they take the form of kernel-weighted least-squares estimators,

$$\hat{\beta}(\tau) = \left[\sum_{t=1}^{n} K_h(t/n - \tau) X_t X_t'\right]^{-1}\left[\sum_{t=1}^{n} K_h(t/n - \tau) X_t y_t\right], \tag{2.6}$$

$$\hat{\sigma}^2(\tau) = \frac{\sum_{t=1}^{n} K_h(t/n - \tau)\,\varepsilon_t^2(\hat{\beta}_t)}{\sum_{t=1}^{n} K_h(t/n - \tau)}. \tag{2.7}$$

The above estimator of $\beta(\tau)$ is identical to the one proposed by Robinson (1989), and is similar to nonparametric estimators of varying-coefficient models in a cross-sectional setting; see Fan and Zhang (2008) for an overview. The estimator of the volatility is akin to the volatility estimator considered in Fan and Yao (1998) except that we here employ normalized time $t/n$ as a regressor; see also Kristensen (2010) for a similar volatility estimator in a continuous-time framework. For notational convenience, we here use the same bandwidth for all regression coefficients and the volatility. There may be finite sample improvements from using different bandwidths for the individual coefficients, see e.g. Fan and Yao (1998) and Fan and Zhang (1999) for results on this in a i.i.d. setting.

Next, we consider estimation of the parametric ($\beta_1$) and nonparametric ($\beta_{2,t}$) components under $H_0$. We propose to estimate the time-varying and constant coefficients by profiled least-squares akin to the local linear profile estimator of Fan and Huang (2005) relying on a first-step kernel estimator of $\beta_{2,t}$. As is well-known in the literature on two-step semiparametric estimators, different bandwidth rules apply depending on whether the interest lies in the estimation of the nonparametric or parametric component. This will also be the case here, and so we here introduce an additional bandwidth $b > 0$ to avoid any confusion. We will then reserve $h$ for the use in the estimation and testing of nonparametric components, while $b$ is employed in the estimation of parametric components.

As a first step towards an estimator of $\beta_1$, treat the parameter as known and estimate $\beta_2(\tau)$ by

$$\tilde{\beta}_2(\tau) = \arg\max_{\beta_2} L_n^{\text{local}}\left(\beta_1, \beta_2, \sigma^2|\tau\right),$$

where $L_n^{\text{local}}\left(\beta_1, \beta_2, \sigma^2|\tau\right)$ is on the same form as before except that we have replaced the bandwidth $h$ by the new bandwith $b$; that is, the kernel weights in the definition of the local log-likelihood now takes the form $K_b(t/n - \tau)$. In order to give an explicit expression of $\tilde{\beta}_2(\tau)$, we introduce some additional notation: For any random sequence $A_t$, define $\hat{M}_b(\tau, A)$ by

$$\hat{M}_b(\tau, A) = \left[\sum_{s=1}^{n} K_b(s/n - \tau) X_{2,s} X_{2,s}'\right]^{-1} \left[\sum_{s=1}^{n} K_b(s/n - \tau) X_{2,s} A_s'\right],$$

which is a kernel estimator of $M(\tau, A) = M_{[\tau n]}(A)$, where

$$M_t(A) = E\left[X_{2,t} X_{2,t}'\right]^{-1} E\left[X_{2,t} A_t'\right]. \tag{2.8}$$

Then it is easily shown that $\tilde{\beta}_2(\tau)$ can be written as

$$\tilde{\beta}_2(\tau) = \hat{M}_b(\tau, y) - \hat{M}_b(\tau, X_1)' \beta_1.$$

Given the estimator of $\beta_{2,t}$, we propose to estimate $\beta_1$ by profile-likelihood: By plugging the conditional estimator $\tilde{\beta}_{2,t}$ into the global log-likelihood together with some preliminary estimator of $\sigma_t^2$ (for example, the unconstrained estimator, $\hat{\sigma}_t^2$), a natural estimator would be the maximize of this w.r.t. $\beta_1$. However, due to bias problems with our kernel estimators for $\tau$ close to the two boundaries, $\tau = 0$ and 1, we first redefine global likelihood function to include trimming,

$$L_n(\beta, \sigma^2) = -\frac{1}{2n} \sum_{t=1}^{n} \mathbb{I}_t(a) \left\{\log\left(\sigma_t^2\right) + \frac{\varepsilon_t^2(\beta_t)}{\sigma_t^2}\right\},$$

where $\mathbb{I}_t(a) = \mathbb{I}\{a \leq t/n \leq 1 - a\}$ for some trimming parameter $a > 0$. That is, we only include observations which are observed a time distance $a$ away from the two end points of the sample. We will let $a$ vanish as $n \to \infty$ such that the impact of trimming is asymptotically negiglible. As an alternative to trimming, a boundary kernel or local linear kernel estimator could be used in the nonparametric estimation since these do not carry any biases at the boundaries, see Kristensen (2010) for a further discussion.

We would like to emphasize that the purpose of the trimming employed here is fundamentally different from the type of trimming introduced in other semiparametric two-step estimators such as Robinson (1987,1988). Usually, trimming is used to handle denominator problems of a first-step nonparametric estimator. Our nonparametric estimator does not suffer from any denominator problems, but rather a boundary problem: That close to either $\tau = 0$ and $\tau = 1$, the estimator is asymptotically biased. The trimming is here used to control this bias component.

Given the redefined log-likelihood, we estimate $\beta_1$ by $\tilde{\beta}_1 = \arg\max_{\beta_1} L_n(\beta_1, \tilde{\beta}_2, \tilde{\sigma}^2)$. This estimator can be written on closed form: First note that

$$L_n(\beta_1, \tilde{\beta}_2, \tilde{\sigma}^2) \propto -\sum_{t=1}^{n} \mathbb{I}_t(a)\, \tilde{\sigma}_t^{-2} \left[\hat{y}_t - \beta_1' \hat{X}_{1,t}\right]^2,$$

where for any random sequence $A_t$ we have defined its corresponding residual as

$$\hat{A}_t = A_t - \hat{M}_{b,t}(A)' X_{2,t}, \quad \hat{M}_{b,t}(A) = \hat{M}_b(t/n, A).$$

As such, $\tilde{\beta}_1$ is the solution to a least-squares problem and is given by

$$\tilde{\beta}_1 = \left[\sum_{t=1}^{n} \mathbb{I}_t(a)\, \tilde{\sigma}_t^{-2} \hat{X}_{1,t} \hat{X}_{1,t}'\right]^{-1} \sum_{t=1}^{n} \mathbb{I}_t(a)\, \tilde{\sigma}_t^{-2} \hat{X}_{1,t} \hat{y}_t'. \tag{2.9}$$

We can substitute this back into the expression of $\tilde{\beta}_2(\tau)$ to obtain an estimator of $\beta_2(\tau)$ under the null:

$$\tilde{\beta}_2(\tau) = \hat{M}_h(\tau, y) - \hat{M}_h(\tau, X_1)\, \hat{\beta}_1, \tag{2.10}$$

where we here use the bandwidth $h$ instead of $b$, since now the interest lies in the estimation of a nonparametric component. One can potentially update the variance estimator by:

$$\tilde{\sigma}^2(\tau) = \frac{\sum_{t=1}^{n} K_h(t/n - \tau)\, \varepsilon_t^2(\tilde{\beta}_t)}{\sum_{t=1}^{n} K_h(t/n - \tau)}. \tag{2.11}$$

One can also iterate between the two estimators given in Eq. (2.9)-(2.10) and the variance estimator in Eq. (2.11), but as we shall see this will not lead to any first-order improvements.

While the above GLS estimator $\tilde{\beta}_1$ is asymptotically efficient (see Section 3), one may worry about its precision for small and moderate sample sizes. In particular, the estimator involves a preliminary estimator of the time-varying variance, $\tilde{\sigma}_t^2$, which in turn requires choosing an additional bandwidth. We therefore introduce a more general estimator depending on weights that can be chosen in a given application,

$$\tilde{\beta}_1^w = \left[\sum_{t=1}^{n} \mathbb{I}_t(a)\, \hat{w}_t \hat{X}_{1,t} \hat{X}_{1,t}'\right]^{-1} \sum_{t=1}^{n} \mathbb{I}_t(a)\, \hat{w}_t \hat{X}_{1,t} \hat{y}_t', \tag{2.12}$$

where $\hat{w}_t = \hat{w}(t/n)$ for some (potentially estimated) weighting function $\hat{w} : [0,1] \mapsto \mathbb{R}_+$. With $\hat{w}_t = \tilde{\sigma}_t^{-2}$, the efficient GLS estimator $\tilde{\beta}_1$ appears, while with $\hat{w}_t = 1$ the standard OLS estimator is obtained.

The resulting estimators share some similarities with the estimators in partially linear models and regression models with heteroskedasticity of unknown form as proposed by Robinson (1988) and Robinson (1987) respectively. The above estimator is a weighted time series version of the estimator proposed Fan and Huang (2005) who use uniform weights, $\hat{w}_t = 1$, and analyze its properties in a cross-sectional setting.

An alternative estimator of $\beta_1$ is obtained by simply averaging the unrestricted estimator $\hat{\beta}_1(\tau)$ over $\tau \in [0,1]$, $\check{\beta}_1 = \int_0^1 \omega(\tau)\hat{\beta}_1(\tau)\,d\tau$, for any weighting function $\omega$ satisfying $\int_0^1 \omega(\tau)\,d\tau = 1$. Ang and Kristensen (2009) show that $\check{\beta}_1$ is $\sqrt{n}$-asymptotically normally distributed but in general not as efficient as $\tilde{\beta}_1^w$. It should be possible to obtain full efficiency by suitable choice of $\omega$, but the optimal weighting function will however depend on unknown components and therefore has to be estimated. We will in the following focus on the likelihood-based estimator $\tilde{\beta}_1^w$.

## 2.2 Testing

Once the restricted estimators have been computed, we may then test $H_0$ by comparing the unrestricted and restricted fit of the model: We here propose two different tests: The first test is a Likelihood-Ratio type test that compares the sums of squared residuals (SSR's) associated with the unrestricted and restricted model, while the second directly compares the restricted and unrestricted estimator of $\beta_1(\tau)$.

To obtain our test statistics, we first define the residuals under the $H_0$ and its alternative respectively,

$$\tilde{\varepsilon}_t = y_t - \tilde{\beta}_1^{w\prime} X_{1,t} - \tilde{\beta}_{2,t}' X_{1,t}, \quad \hat{\varepsilon}_t = y_t - \hat{\beta}_{1,t}' X_{1,t} - \hat{\beta}_{2,t}' X_{1,t},$$

where, as before, $\tilde{\beta}_1^w$ is computed using the "semiparametric" bandwidth $b$ while $\hat{\beta}_t$ and $\tilde{\beta}_{2,t}$ relies on the "nonparametric" bandwidth $h$. The corresponding sums of (weighted) squared residuals under null and alternative are given by

$$SSR_0^w = \sum_{t=1}^n \mathbb{I}_t(a)\,\hat{w}_t \tilde{\varepsilon}_t^2, \quad SSR_A^w = \sum_{t=1}^n \mathbb{I}_t(a)\,\hat{w}_t \hat{\varepsilon}_t^2, \tag{2.13}$$

where $\hat{w}_t$ are some weights chosen by the econometrician (not necessarily the same used to compute $\tilde{\beta}_1^w$). We then propose to test $H_0$ using a generalized $F$ statistic given by

$$F_n = \frac{n}{2}\frac{SSR_0^w - SSR_A^w}{SSR_A^w}.$$

The statistic $F_n$ is similar to the generalized likelihood-ratio (GLR) test statistic proposed in Fan et al (2002) for varying-coefficient models. In particular, with $\hat{w}_t = \tilde{\sigma}_t^{-2}$, $F_n$ can be seen as a first-order approximation of the GLR based on $L_n(\beta, \sigma^2)$. For $\hat{w}_t = 1$, $F_n$ is the first-order approximation of the GLR proposed in Fan et al (2002) in a cross-sectional setting with homoskedastic errors.

As an alternative to $F_n$, we also consider a generalized Wald statistic that measures the discrepancy between the restricted and unrestricted estimator of $\beta_1(\tau)$:

$$W_n = \sum_{t=1}^{n} \mathbb{I}_t(a) \left(\tilde{\beta}_1^w - \hat{\beta}_{1,t}\right)' \hat{\Omega}_t \left(\tilde{\beta}_1^w - \hat{\beta}_{1,t}\right),$$

for some sequence of (possibly estimated) weights, $\hat{\Omega}_t \geq 0$, $t = 1, ..., n$. One particular choice of $\hat{\Omega}_t$ is $\hat{\Omega}_t = X_{1,t}X_{1,t}'$, see Chen and Hong (2009), but others are possible too. In particular, when the errors are heteroskedastic, one can include a volatility weight in order for the test statistic to be asymptotically distribution free as we will discuss in the next section.

# 3 Asymptotics Properties

To derive the asymptotic properties of the above estimation testing procedure, we assume that data has arrived from the following sequence of models,

$$y_{n,t} = \beta_{n,t}' X_{n,t} + \sigma_{n,t} z_{n,t}, \quad t = 1, ..., n, \tag{3.1}$$

where $\beta_{n,t}'$ and $\sigma_{n,t}$ satisfy eq. (2.4). We allow the sequences $\{\beta_{n,t}\}$ and $\{\sigma_{n,t}\}$ to be random in which case all the following arguments and statements are implicitly made conditional on the realization of these two random sequences that generated data. Moreover, the set of regressors, $X_{n,t}$, and errors, $z_{n,t}$, may potentially depend on sample size $n$ such that structural change in their distributions are allowed for. As such, our model resembles the one considered in Hansen (2000b), except that we do not impose parametric assumptions on the changing parameters. We will however require that the regressors, while non-stationary, are mixing. One particular situation that our theoretical results cover is when $X_{n,t}$ includes lagged dependent variables in which case our regression model is an autoregressive model. The simplest example is

$$y_{n,t} = \mu(t/n) + \rho(t/n) y_{n,t-1} + \sigma(t/n) z_{n,t},$$

in which case $X_{n,t} = (1, y_{n,t-1})'$ is non-stationary when the functions $\mu(\cdot)$ and $\rho(\cdot)$ are non-constant. Under the restriction that $\sup_{\tau\in[0,1]} |\rho(\tau)| < 1$, $X_{n,t}$ is however still mixing, c.f. Orbes et al (2004), and our theoretical results apply.

To state our assumptions and results, we introduce some additional notation. Let $\Lambda_{n,t}$ denote the following moment matrix

$$\Lambda_{n,t} = \begin{bmatrix} \Lambda_{n,11,t} & \Lambda_{n,12,t} \\ \Lambda_{n,21,t} & \Lambda_{n,22,t} \end{bmatrix} \in \mathbb{R}^{m\times m},$$

where for $k, l \in \{1, 2\}$,

$$\Lambda_{n,kl,t} \equiv E\left[X_{n,k,t} X_{n,l,t}'\right] \in \mathbb{R}^{m_k \times m_l} \tag{3.2}$$

We will impose certain smoothness conditions on the parameters of interest, and for that purpose introduce the following function space of $r$ times continuously differentiable functions,

$$\mathcal{C}^r[0,1] = \{f : [0,1] \mapsto \mathbb{R} | f \text{ is } r \text{ times differentiable}\}.$$

We then impose the following assumptions conditional on $\beta(\cdot)$ and $\sigma^2(\cdot)$:

**A.1** For all $n \geq 1$: The joint sequence $\{Z_{n,t} = (X_{n,t}, z_{n,t}) : i = 1, ..., n\}$ satisfies

$$\sup_{n\geq 1} \sup_{t \leq n} E\left[\|Z_{n,t}\|^{4+\delta}\right] < \infty$$

for some $\delta > 0$; it is $\beta$-mixing where the mixing coefficients,

$$b_n(i) = \sup_{-n\leq k \leq n} \sup_{A\in\mathcal{F}^k_{n,-\infty}, B\in\mathcal{F}^\infty_{n,n+i}} |P(A\cap B) - P(A)P(B)|,$$

satisfy $b_n(i) \leq b(i)$, $n \geq 1$, and the dominating sequence $b(i)$ is geometrically decreasing.

**A.2** The errors $z_{n,t}$ is a MGD w.r.t. $\mathcal{F}_{n,t} = \mathcal{F}(X_{n,s}, z_{n,s-1} | s \leq t)$ with $E\left[z^2_{n,t} | X_{n,t}\right] = 1$ and $\lambda_{n,t} := E[(z^2_{n,t} - 1)^2] < \infty$.

**A.3** The sequences $\beta_{n,t}$, $\Lambda_{n,t}$ and $\sigma^2_{n,t}$ satisfy $\beta_{n,t} = \beta(t/n) + o(1)$, $\Lambda_{n,t} = \Lambda(t/n) + o(1)$, $\sigma^2_{n,t} = \sigma^2(t/n) + o(1)$ for some functions $\beta(\cdot)$, $\Lambda(\cdot)$ and $\sigma^2(\cdot)$. The elements of these functions are in $\mathcal{C}^r[0,1]$ for some $r \geq 1$. For all $\tau \in [0,1]$, $\Lambda(\tau)$ and $\sigma^2(\tau)$ are positive definite.

**A.4** The weighting functions $\hat{w}(\cdot)$ and $\hat{\Omega}(\cdot)$ satisfy:

$$\begin{aligned}
&\text{(i)} && \sup_{a\leq\tau\leq 1-a} |\hat{w}(\tau) - w(\tau)| &=& O_P\left(n^{1/4}\right), \\
&\text{(ii)} && \sup_{a\leq\tau\leq 1-a} |\hat{w}(\tau) - w(\tau)| &=& O_P\left(h^{1/2}\right), \\
&\text{(iii)} && \sup_{a\leq\tau\leq 1-a} \left|\hat{\Omega}(\tau) - \Omega(\tau)\right| &=& O_P\left(h^{1/2}\right),
\end{aligned}$$

where $w(\cdot)$ and $\Omega(\cdot)$ are continuous functions.

**A.5** The covariance matrices $\Phi_w$ and $\Sigma_w$ as defined below are non-singular:

$$\begin{aligned}
\Sigma_w &:= \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n w_t \Lambda_{11|2,t} = \int_0^1 w(\tau)\Lambda_{11|2}(\tau)\,d\tau, \\
\Phi_w &:= \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n w_t^2 \sigma_t^2 \Lambda_{11|2,t} = \int_0^1 w^2(\tau)\sigma^2(\tau)\Lambda_{11|2}(\tau)\,d\tau
\end{aligned}$$

where

$$\Lambda_{11|2,t} \equiv \Lambda_{11,t} - \Lambda_{12,t}\Lambda^{-1}_{22,t}\Lambda_{21,t},$$

and $\Lambda_{kl,t}$, $k, l = 1, 2$, are defined in eq. (3.2).

**A.6** The "semiparametric" bandwidth $b$ satisfies $nb^{2r} \to 0$, $\log^2(n)/(nb^2) \to 0$ and $n^{1-\epsilon}b^{7/4} \to \infty$ for some $\epsilon > 0$. The trimming parameter $a > 0$ satisfies $a/b \to 0$ and $\sqrt{n}a \to 0$.

The assumption of $\beta$-mixing in (A.1) is not required for all our results, but is imposed throughout for simplicity. Some results hold under the weaker assumption of $\alpha$-mixing, but for the semiparametric estimation and testing, we will rely on U-statistics results that only exist for $\beta$-mixing sequences. The assumption of geometrically decaying mixing coefficients is only imposed to make proofs and remaining conditions simpler, and could most likely be weakend. We do not assume stationarity in (A.1) and as such allow for situations where $X_{n,t}$ contains structural breaks; in particular, our framework includes unstable autoregressive models where $X_{n,t}$ contains lagged values of $y_{n,t}$. In a time-varying AR($q$)-model where $X_{n,t} = (y_{n,t-1}, ..., y_{n,t-q})'$, (A.1) is satisfied if the roots of the characteristic polynomail $\theta(\tau, z) = \beta_1(\tau) z + ... + \beta_q(\tau) z^q$ are inside of the unit circle for all $\tau \in [0,1]$ and the errors $z_{n,t}$ are i.i.d. with a continuous distribution. Sufficient conditions for (A.1) when $X_{n,t}$ solves a nonlinear model can be found in Kristensen (2011) and Subba Rao (2006).

Assumption (A.2) rules out correlated errors. We conjecture that our results can be extended to allow for this, but our asymptotic results and their proofs would become more complicated and burdensome, see e.g. Cai (2007) for some results in this direction.

The smoothness conditions imposed on the coefficients in (A.3) rule out discontinuities (jumps) in the coefficients. If jumps are present, then the pointwise kernel estimators of the time-varying coefficients will be inconsistent at any point in time where one such occurs. However, as discussed in the conclusion, by suitable adjustments of the estimators, jumps can be consistently estimated. Moreover, we expect that the asymptotic results for the semiparametric estimators and the test statistics remain valid when a finite jumps are present since these happen with measure zero. The assumption of twice differentiability is assumed for technical convenience, and could most likely be weakend to the assumption that the functions are Lipschitz by following the arguments of Kristensen (2010).

The two conditions on the estimated weighting functions imposed in (A.4) are made to ensure that their estimation errors do not affect the properties of the parametric estimators and test statistics. The first condition, (A.4.i), is used when deriving the asymptotics of $\tilde{\beta}$, while (A.4.ii)-(A.4.iii) are needed in the analysis of the test statistics. The two conditions are satisfied by standard kernel estimators such as $\tilde{\sigma}(\tau)$.

The rank condition in Assumption (A.5) is employed to ensure identification and asymptotic normality of $\beta_1$ under $H_0$. It is similar to the condition imposed for identification and estimation of partially linear models in Robinson (1988).

Restrictions on the bandwidth and trimming sequences used for the semiparametric estimators are imposed in (A.6). In general, undersmoothing is required (that is, $b$ should be chosen to converge faster than the optimal bandwidth minimizing MSE of the nonparametric estimators). The restrictions on the trimming parameter are on the other hand quite weak since this is only used to handle boundary issues.

Finally, we need to impose some regularity conditions on the kernel $K$:

**K**$(r)$ There exists $B, L < \infty$ such that either (i) $K(u) = 0$ for $\|u\| > L$ and $|K(u) - K(u')| \leq B\|u - u'\|$, or (ii) $K(u)$ is differentiable with $|\partial K(u)/\partial u| \leq B$ and, for some $\nu > 1$, $|\partial K(u)/\partial u| \leq B\|u\|^{-\nu}$ for $\|u\| \geq L$. Also, $|K(u)| \leq B\|u\|^{-\nu}$ for $\|u\| \geq L$. For some $r \geq 2$: $\int_{\mathbb{R}} K(z)\,dz = 1$, $\int_{\mathbb{R}} z^i K(z)\,dz = 0$, $i = 1, ..., r-1$, and $\int_{\mathbb{R}} |z|^r K(z)\,dz < \infty$.

The assumptions are satisfied by most kernels. In particular, for $r = 2$ the Gaussian kernel satisfies the condition. The order of the kernel, $r \geq 2$, is used in conjunction with the smoothness conditions imposed on the relevant functions in (A.3) to control the bias of the kernel estimators which will be of order $O(h^r)$. Some of our results will rely on higher-order kernels with $r > 2$ in order for the bias of the kernel estimators to vanish at a sufficiently fast rate. However, we believe higher-order kernels are only for needed for technical reasons in the theoretical proofs, and recommend the use of standard second order kernels in practice.

The first result states the pointwise asymptotic distribution of the unrestricted nonparametric estimators:

**Theorem 3.1** *Assume that (A.1)-(A.3) hold. Then, for any $\tau \in (0,1)$, as $h \to 0$, $nh \to \infty$ and $nh^{1+2r} \to 0$:*

$$\sqrt{nh}(\hat{\beta}(\tau) - \beta(\tau)) \to^d N\left(0, \|K\|^2 \Lambda^{-1}(\tau)\,\sigma^2(\tau)\right),$$

*where $\Lambda(\tau)$ is defined in (A.3) and $\|K\|^2 = \int K^2(z)\,dz$.*

Theorem 3.1 tells us how pointwise confidence bands of the regression coefficients and the volatility can be computed. These can be used as inputs in the initial analysis of whether there is any time variation in the individual elements of $\beta(\tau)$ and $\sigma^2(\tau)$. This can for example be done by plotting the individual estimators as functions of time together with confidence bands. This eyeballing test should of course be followed by the proposed formal statistical tests which are analyzed below.

A simple estimator of the asymptotic variance can be obtained by substituting in $\hat{\sigma}^2(\tau)$ together with:

$$\hat{\Lambda}(\tau) = \sum_{i=1}^{n} K_h(t/n - \tau)\,X_t X_t'. \tag{3.3}$$

The asymptotic distributional result in Theorem 3.1 is standard for nonparametric estimators. The result reveals an important advantage of our estimation strategy over most other nonparametric regression techniques, namely that there is no curse of dimensionality present here: The convergence rates of the estimators remain $\sqrt{nh}$ irrespectively of the number of regressors included since we only smooth over the time variable $t$.

Next, we consider the estimation under $H_0$. The next theorem states the asymptotic distribution of the semiparametric estimator of the constant component under $H_0$:

**Theorem 3.2** *Assume that (A.1)-(A.4,i) and (A.5)-(A.6) hold. Under $H_0$:*

$$\sqrt{n}(\tilde{\beta}_1^w - \beta_1) \to^d N\left(0, \Sigma_w^{-1}\Phi_w\Sigma_w^{-1}\right),$$

*where $\Phi_w$ and $\Sigma_w$ are given in (A.5).*

The above theorem is essentially a time-series version of the asymptotic result obtained for semi-varying coefficient models in Fan and Huang (2005) where in addition we allow for heteroskedastic errors. The theorem and its proof reveals that our estimator is first-order asymptotically equivalent to the weighted least-squares estimator of the regression $\bar{y}_t = \beta_1' V_t + \varepsilon_t$, where $\bar{y}_t = y_t - M_t(y)' X_{2,t}$, $V_t := X_{1,t} - M_t(X_1)' X_{2,t}$,and $M_t(A)$ is defined in eq. (2.8).

When the weighting function is chosen as $\hat{w}(\tau) = \hat{\sigma}^{-2}(\tau)$, we see that

$$\Phi_w = \Sigma_w = \int_0^1 \sigma^{-2}(\tau)\Lambda_{11|2}(\tau)\,d\tau,$$

in which case $\sqrt{n}(\tilde{\beta}_1^w - \beta_1) \to^d N\left(0, \Sigma_w^{-1}\right)$. We conjecture that for this choice of weighting function, our estimator is semiparametrically efficient. The theory on semiparametric efficiency in time series models is currently not fully developed, and so we are not able to verify this conjecture in the general case. Instead, we restrict ourselves to the case where $(X, z)$ are i.i.d.: Treating $\tau := t/n$ as i.i.d draws from a uniform distribution which is independent of $(X, z)$, our model then fits into the framework of Chamberlain (1992) with his moment condition here being on the form $\rho(y, X, \beta_1, \beta_2(\tau)) := y - \beta_1' X_1 - \beta_2(\tau) X_2$. We can now apply the results of Chamberlain (1992) stating that the efficiency bound is

$$\mathcal{I}_0 = E\left[E\left(D_0'\Sigma_0^{-1}D_0|\tau\right) - E\left(D_0'\Sigma_0^{-1}H_0|\tau\right)E\left(H_0'\Sigma_0^{-1}H_0|\tau\right)^{-1}E\left(H_0'\Sigma_0^{-1}D_0|\tau\right)\right],$$

where, in our case,

$$\begin{aligned}
D_0(X,\tau) &= E\left[\left.\frac{\partial\rho(y,X,\beta_1,\beta_2(\tau))}{\partial\beta_1}\right| X,\tau\right] = -X_1', \\
\Sigma_0(X,\tau) &= E\left[\left.\rho^2(y,x,\beta_1,\beta_2(\tau))\right| X,\tau\right] = \sigma^2(\tau), \\
H_0(X,\tau) &= E\left[\left.\frac{\partial\rho(y,x,\beta_1,\beta_2(\tau))}{\partial\beta_2}\right| X,\tau\right] = -X_2'.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathcal{I}_0 &= E\left[E\left(\sigma^{-2}(\tau)X_1X_1'|\tau\right) - E\left(\sigma^{-2}(\tau)X_1X_2'|\tau\right)E\left(\sigma^{-2}(\tau)X_2X_2'|\tau\right)^{-1}E\left(\sigma^{-2}(\tau)X_2X_1'|\tau\right)\right] \\
&= \int \sigma^{-2}(\tau)\Lambda_{11|2}(\tau)\,d\tau,
\end{aligned}$$

which matches up with $\Sigma_w$ when the weighting function satisfies $w = \sigma^{-2}$. As such our estimator extends the semiparametric estimator and results of Fan and Huang (2005) to allow for heteroskedastic errors and time series dependence: They show that the unweighted version ($w_t = 1$) of our estimator is semiparametric efficient in a cross-sectional setting when errors are homoskedastic.

The asymptotic variance terms can be consistently estimated by

$$\hat{\Phi}_w = \frac{1}{n}\sum_{t=1}^{n} \mathbb{I}_t(a)\,\hat{w}_t \hat{X}_{1,t}\hat{X}_{1,t}', \quad \hat{\Sigma}_w = \frac{1}{n}\sum_{t=1}^{n} \mathbb{I}_t(a)\,\hat{w}_t \hat{X}_{1,t}\hat{X}_{1,t}\tilde{\varepsilon}_t^2,$$

where $\hat{X}_{1,t} = X_{1,t} - \hat{M}_t(X_1)' X_{2,t}$.

Due to $\tilde{\beta}_1^w$ being $\sqrt{n}$-consistent, we can treat $\beta_1$ as known, when deriving the asymptotics of $\tilde{\beta}_2(\tau)$. Thus, we can apply the same arguments as in the proof of Theorem 3.1 to obtain that under $H_0$,

$$\sqrt{nh}(\tilde{\beta}_2(\tau) - \beta_2(\tau)) \to^d N\left(0, \|K\|^2 \Lambda_{22}^{-1}(\tau)\,\sigma_2^2(\tau)\right),$$

as $h \to 0$, $nh \to \infty$ and $nh^{1+2r} \to 0$. The asymptotic variance of $\tilde{\beta}_2(\tau)$ can be estimated using estimators similar to those given in eq. (2.7).

Next, we analyze the asymptotic properties of the two test statistics, $F_n$ and $W_{1,n}$. First, consider the test statistic $F_n$:

**Theorem 3.3** *Assume that (A.1)-(A.6) hold and:* $nh^{2r+1} \to 0$, $nh^{3/2}/\log(n)^2 \to \infty$, $a/h \to 0$ *and* $\sqrt{h}a \to 0$. *Then under* $H_0$,

$$\frac{F_n - \mu_n^F}{\sqrt{\nu_n^F}} \to^d N(0,1),$$

*where*

$$\mu_n^F = \frac{m_1\left[K(0) - \frac{1}{2}\kappa_2\right]}{h}, \quad \nu_n^F = \frac{2m_1}{h}\frac{\int_0^1 w^2(v)\,\sigma^4(v)\,dv}{\left(\int_0^1 w(\tau)\,\sigma^2(\tau)\,d\tau\right)^2} \times \left\|K - \tfrac{1}{2}(K*K)\right\|^2,$$

*and* $(K*K)(z) = \int K(v)\,K(z+v)\,dv$.

The asymptotic distribution of the normalized test statistic, $F_n$, follows a standard Normal distribution under the null hypothesis. The distribution is similar to the ones foound for the Generalized Likelihood Ratio (GLR) test statistics in Fan et al. (2001). In particular, using the notation that $r\lambda_n \overset{a}{\sim} \chi^2_{b_n}$ for a random sequence $\lambda_n$ that satisfies $(r\lambda_n - b_n)/\sqrt{2b_n} \to^d N(0,1)$, we observe that the above theoretical result also can be written as $r_K^F F_n \overset{a}{\sim} \chi^2_{r_K^F \mu_n^F}$, where

$$r_K^F := \frac{K(0) - \frac{1}{2}\kappa_2}{\int\left[K(z) - \frac{1}{2}(K*K)(z)\right]^2 dz}\,\frac{\left(\int_0^1 w(\tau)\,\sigma^2(\tau)\,d\tau\right)^2}{\int_0^1 w^2(\tau)\,\sigma^4(v)\,d\tau}.$$

However, it is important to note here that the distribution does depend on nuisance parameters in general, except in the case where the weighting function is chosen as $w(\tau) = \sigma^{-2}(\tau)$, in which case

$$r_K^F := \frac{K(0) - \frac{1}{2}\kappa_2}{\int \left[K(z) - \frac{1}{2}(K * K)(z)\right]^2 dz}.$$

So in general, one has to obtain a consistent estimator of the volatility process in order for the test statistic to enjoy the so-called Wilks phenomenon. This is not special to the time series setting, and is also the case in the cross-sectional setting where Fan et al. (2001) show that only in the case of homoskedastic errors (in which case one can choose $w(\tau) = 1$) will their GLR test be nuisance parameter free.

We conjecture that the GLR test statistic is asymptotically optimal in the sense that it can detect local alternatives with optimal rate, see Ingster (1993) for more details. In a cross-sectional setting, this is shown to hold in Fan et al (2001), and we expect that these results carry over to our time series model.

Next, we turn to the minimum-distance statistic. For this, the following asymptotic distributional result holds:

**Theorem 3.4** *Assume that (A.1)-(A.6) hold and:* $nh^{2r+1} \to 0$, $nh^{3/2}/\log(n)^2 \to \infty$, $a/h \to 0$ *and* $\sqrt{h}a \to 0$. *Then under* $H_0$,

$$\frac{W_n - \mu_n^W}{\sqrt{\nu_n^W}} \xrightarrow{d} N(0,1), \tag{3.4}$$

*where*

$$\mu_n^W = \frac{\kappa_2}{h}\int_0^1 \sigma^2(\tau)\, tr\left\{\Lambda_{11}^{-1}(\tau)\,\Omega(\tau)\right\} d\tau,$$

$$\nu_n^W = \frac{2}{h}\int \sigma^4(\tau)\, tr\left\{\Omega(\tau)\,\Lambda_{11}^{-1}(\tau)\,\Omega(\tau)\,\Lambda_{11}^{-1}(\tau)\right\} d\tau \times \|K * K\|^2$$

As with the GLR-statistic, we can express the above result on the form $r_K^W W_n \overset{a}{\sim} \chi^2_{r_K^W \mu_n^W}$, where

$$r_K^W := \frac{\kappa_2}{\|K * K\|^2}\frac{\left(\int_0^1 \sigma^2(\tau)\,\mathrm{tr}\left\{\Lambda_{11}^{-1}(\tau)\,\Omega(\tau)\right\} d\tau\right)^2}{\int \sigma^4(\tau)\,\mathrm{tr}\left\{\Omega(\tau)\,\Lambda_{11}^{-1}(\tau)\,\Omega(\tau)\,\Lambda_{11}^{-1}(\tau)\right\} d\tau}.$$

Again, the asymptotic distribution of the MD-statistic depends in general on nuisance parameters. However, by choosing the weighting matrix $\Omega_t$ as $\Omega_t = \Lambda_{11,t}\sigma_t^{-2}$, we obtain

$$r_K^W := \frac{m_1\kappa_2}{\|K * K\|^2},$$

and the limiting distribution is nuisance parameter free. In comparison to $F_n$, the location and scale sequences associated with $W_n$ are different. For general choices of the weighting functions $w_t$ and $\Omega_t$, it is not clear which of the two tests dominates. However, in the case where $w(\tau) = \sigma^{-2}(\tau)$ and $\Omega_t = \Lambda_{11,t}\sigma_t^{-2}$, it can be shown by following the arguments in Chen and Hong (2009) that $W_n$ is asymptotically more efficient in the sense of Pitmann; see also Hong and Lee (2008).

# 4 Some Extensions

We here extend the above results to two additional hypotheses which should be of general interest: First, we consider the situation where the researcher has tested and accepted the null that a subset of the coefficients are constant, and then wishes to test for time invariance of a set of the remaining (potentially) time-varying coefficients. This is for example of relevance in order to develop a recursive procedure testing the constancy of each coefficient one at a time. Second, we analyze the problem of testing a parametric specification of (some of the) time-varying parameters against a nonparametric alternative. This is of interest if one has rejected the null of constant parameters, and now wishes to find a parsimonious parametric specification of the time-varying parameters.

We start out by assuming that the following (maintained) model is correct:

$$y_t = \alpha' W_t + \gamma_{1,t}' Z_{1,t} + \gamma_{2,t}' Z_{2,t} + \sigma_t z_t. \tag{4.1}$$

We then wish to test the following null hypothesis against this model:

$$H_1 : \gamma_{1,t} = \gamma_1.$$

We proceed as in the testing of $H_0$: We first obtain estimators under null and alternative, and the compare the estimators through either an $F$ or a Wald-type test. The model under the alternative can be written on the form of the model under $H_0$ with $X_{1,t} = W_t$, $X_{2,t} = \left(Z_{1,t}', Z_{2,t}'\right)'$, $\beta_1 = \alpha$ and $\beta_{2,t} = \left(\gamma_{1,t}', \gamma_{2,t}'\right)'$. Thus, the estimators under the alternative are given by:

$$\hat{\alpha}^w = \left[\sum_{t=1}^n \mathbb{I}_t(a)\, \hat{w}_t \hat{W}_t \hat{W}_{Z,t}'\right]^{-1} \sum_{t=1}^n \mathbb{I}_t(a)\, \hat{w}_t \hat{W}_t \hat{y}_t', \quad \hat{\gamma}_t = \hat{M}_{h,t}(y) - \hat{M}_{h,t}(W)' \hat{\alpha}, \tag{4.2}$$

where $\hat{A}_t = A_t - \hat{M}_{b,t}(A)' Z_t$, and

$$\hat{M}_{b,t}(A) = \left[\sum_{s=1}^n K_b(s/n - t/n)\, Z_s Z_s'\right]^{-1} \left[\sum_{s=1}^n K_b(s/n - t/n)\, Z_s A_s'\right].$$

Similarly, under $H_1$, with $X_{1,t} := \left(W_t', Z_{1,t}'\right)$, $X_{2,t} := Z_{2,t}$, $\beta_1 := (\alpha, \gamma_1)$ and $\beta_{2,t} := \gamma_{2,t}$, we recognize the model as being on the same form as the one under $H_0$. Thus, the estimators, which we denote $\tilde{\alpha}^w$, $\tilde{\gamma}_1^w$ and $\tilde{\gamma}_{2,t}$, can again be written on the form of the estimators analyzed in the previous section. It now follows directly from Theorem 3.2 that both $\hat{\alpha}$, $\tilde{\alpha}$ and $\tilde{\gamma}_1$ are $\sqrt{n}$-asymptotically normally distributed under $H_1$.

To test $H_1$ against the maintained hypothesis, we proceed as before: Letting $\hat{\varepsilon}_t$ and $\tilde{\varepsilon}_t$ denoting the residuals under the alternative and under the null, we can compute (weighted) Sum of Squared Residuals, and use these to construct an $F$-test, which we denote $F_{1,n}$, while the Wald test is defined as $W_{1,n} = \sum_{t=1}^{n} \mathbb{I}_t(a) \left(\tilde{\gamma}_1^w - \hat{\gamma}_{1,t}\right)' \hat{\Omega}_t \left(\tilde{\gamma}_1^w - \hat{\gamma}_{1,t}\right)$. By using the exact same arguments as in the proofs of Theorems 3.3-3.4, we now obtain that under the same conditions as stated in these theorems:

$$\frac{F_{1,n} - \mu_{1,n}^F}{\sqrt{\nu_{1,n}^F}} \to^d N(0,1), \quad \frac{W_{1,n} - \mu_{1,n}^W}{\sqrt{\nu_{1,n}^W}} \to^d N(0,1) \tag{4.3}$$

where $\mu_{1,n}^F$, $\nu_n^F$, $\mu_{1,n}^W$ and $\nu_{1,n}^W$ are on the same form as the corresponding location and scale parameters in Theorems 3.3-3.4 except that $m_1$ should be replaced by the dimension of $\gamma_1$.

To develop parametric tests of the functional form of varying coefficients, we now wish to test the following hypothesis,

$$H_2 : \gamma_{1,t} = \gamma_1(t/n; \theta).$$

against the maintained hypothesis that eq. (4.1) holds and now. The parametric specification $\gamma_1(t/n;\theta)$ could for example be a structural break specification or a smooth transition model. Assuming that an estimator of $\theta$, $\tilde{\theta}$, is available under $H_2$,[1] such that $\sqrt{n}(\gamma_1(\tau;\tilde{\theta}) - \gamma_1(\tau)) = O_P(1)$ uniformly in $\tau \in [0,1]$, one can easily show that the corresponding $F$- and Wald test statistics of $H_2$ still satisfy eq. (4.3).

# 5 Implementation

In the previous sections, we analyzed the asymptotic properties of the proposed estimators and tests. In particular, for these results to hold the bandwidths have to converge at suitable rates as sample size grows. The stated conditions and results are however silent about the appropriate choice of the bandwidths in finite samples, and, as is well-known in the literature, kernel-based estimators and tests tend to be quite sensitive to the chosen bandwidths.

There are two bandwidth selection issues involved in the estimation and testing. We have to choose one bandwidth, $h$, for the point estimates of the nonparametric components, and a different bandwidth, $b$, for the parametric component. The use of two different bandwidths are necessary because in our theoretical framework the bandwidth selection rules differ depending on whether the interest lies in the estimation of the non- or fully parametric component. In particular, the asymptotic results suggest that for parametric estimators undersmoothing is necessary; that is, $b$ should in general be chosen smaller than $h$.

[1] It is outside of the scope of this paper to analyze this more general semiparametric estimation problem, but conjecture that the natural two-step estimator, obtained in the same fashion as the semiparametric estimator of the constant specification, $\tilde{\beta}^w$, could be shown to be $\sqrt{n}$-consistent by following the proof strategy used for Theorem 3.2.

While there is a large literature on bandwidth selection for fully nonparametric kernel estimators, there has been done little on how to choose bandwidths in semiparametric estimation problems since the impact of the bandwidth in the latter case is a lot more difficult to analyze. Similarly, very little work has been done on bandwidth selection for non- and semiparametric testing. Our proposed solution to this problem is to first determine the bandwidth $h$ that minimize the (an estimated version of) mean-square error of the nonparametric estimators, and then choose $b$ by adjusting $h$ in a suitable manner. We have no theoretical finite-sample justification for the proposed selection rule for $b$, but our simulation study reveals that it does an acceptable job.

To estimate $h$, we employ a plug-in method. We here focus on the estimation of $\beta_t$ under the alternative; the following arguments are easily adapted to the case of estimation of $\beta_{2,t}$ under the null. First, we note that from the proof of Theorem 3.1, we find that the bias and variance when a second order kernel ($r = 2$) is employed are given by

$$\text{Bias}\left(\hat{\beta}(\tau)\right) = h^2 b(\tau) + o\left(h^2\right) \quad \text{with} \quad b(\tau) := \mu_2 \beta^{(2)}(\tau), \tag{5.1}$$

and

$$\text{Var}\left(\hat{\beta}(\tau)\right) = \frac{1}{nh} v(\tau) + o\left(1/(Th)\right) \quad \text{with} \quad v(\tau) = \|K\|^2 \Lambda^{-1}(\tau) \sigma^2(\tau), \tag{5.2}$$

where $\mu_2 := \int K(z) z^2 dz$. Thus, the optimal bandwidth that minimizes the integrated MSE is

$$h_j^* = \left[\frac{||V||}{||B||^2}\right]^{1/5} n^{-1/5}, \tag{5.3}$$

where $V = \int v(\tau) d\tau$ and $B = \int b(\tau) d\tau$ are the integrated time-varying variance and bias components. In order to make the above bandwidth selection rule operational, we propose to obtain preliminary estimates of these through the following two-step method:[2]

1. Assume that $\Lambda_t = \Lambda$ and $\sigma_t = \sigma$ are constant, and $\beta_t = a_0 + a_1 t + ... + a_p t^p$ is a polynomial. We then obtain parametric least-squares estimates $\hat{\Lambda}$, $\hat{\sigma}^2$ and $\bar{\beta}_t = \bar{a}_0 + \bar{a}_1 t + ... + \bar{a}_p t^p$. Compute

$$\hat{V}_1 = \|K\|^2 \hat{\Lambda}^{-1} \otimes \hat{\sigma}^2 \quad \text{and} \quad \hat{B}_1 = \mu_2 \frac{1}{n} \sum_{t=1}^{n} \bar{\beta}_t^{(2)},$$

   where $\bar{\beta}_t^{(2)} = 2\bar{a}_2 + 6\bar{a}_3 t + ... + p(p-1)\bar{a}_p t^{p-2}$. Then, using these estimates we compute the first-pass bandwidth

$$\hat{h}_1 = \left[\frac{||\hat{v}_1||}{||\hat{b}_2||^2}\right]^{1/5} \times n^{-1/5}.$$

[2] Ruppert, Sheather and Wand (1995) discuss in detail how this can done in a standard kernel regression framework.

2. Given $h_1$, compute the kernel estimators $\hat{\beta}_t = \hat{\Lambda}_t^{-1} n^{-1} \sum_{s=1}^n K_{\hat{h}_1}((s-t)/n) X_s y_s$, where $\hat{\Lambda}_t$ and $\hat{\sigma}_t$ are computed as in equation (**??**) with $h = \hat{h}_1$. We use these to obtain

$$\hat{V}_2 = \|K\|^2 \frac{1}{n} \sum_{t=1}^n \hat{\Lambda}_t^{-1} \otimes \hat{\sigma}_t^2 \quad \text{and} \quad \hat{B}_2 = \mu_2 \frac{1}{n} \sum_{t=1}^n \hat{\beta}_t^{(2)},$$

where $\hat{\beta}_t^{(2)}$ is the second derivative of the kernel estimator with respect to $t$. These are in turn used to obtain a second-pass bandwidth:

$$\hat{h}_2 = \left[ \frac{||\hat{v}_2||}{||\hat{b}_2||^2} \right]^{1/5} \times n^{-1/5}. \tag{5.4}$$

One could alternatively use (generalized) cross-validation (CV) procedures to choose the bandwidth. These procedures are completely data driven and, in general, yield consistent estimates of the optimal bandwidth. However, it is well-known that cross-validated bandwidths may exhibit very inferior asymptotic and practical performance even in a cross-sectional setting (see, for example, Härdle, Hall, and Marron, 1988). This problem is further enhanced when CV procedures are used on time-series data as found in various studies (Hart, 1991; Opsomer, Wang and Yang, 2001).

The "semiparametric" bandwidth $b$ should ideally be chosen to minimize the mean-squared error $E\left[||\tilde{\beta}_1^w - \beta_1||^2\right]$. Unfortunately, this would require a higher-order expansion of the MSE since the leading variance term does not depend on $b$. This is a general issue with semiparametric estimators and outside of the scope of this paper. We instead simply propose to scale down the nonparametric bandwidth $h$ appropriately, $b = \hat{h}_2 \times n^{-1/(1+2r)}$ with $r = 2$ corresponding to a standard kernel being the leading choice.

In small and moderate sample sizes, the asymptotic distributions of estimators and test statistics derived in the previous section may deliver a poor finite-sample approximation. To improve on the finite-sample inference, we therefore propose to use a Wild bootstrap procedure that we expect will yield better confidence bands for the time-varying coefficients and critical values for the test statistic. Let $\breve{\beta}_t$ and $\breve{\sigma}_t$ be (either nonparametric or semiparametric) estimators of the regression coefficients and volatility (under the relevant hypothesis). We then proceed as in Franke, Kreiss and Mammen (2002) and propose the following bootstrap procedure: (i) Compute residuals $\breve{\varepsilon}_t = y_t - \breve{\beta}_t' X_t$, $t = 1, ..., n$; (ii) resample the dependent variable by $y_t^* = \breve{\beta}_t' X_t + \varepsilon_t^*$, $\quad t = 1, ..., n$, where $\varepsilon_t^* = \breve{\varepsilon}_t \eta_t^*$ and $\eta_t^*$ are i.i.d. $(0, 1)$ satisfying $E^*[|\eta_t^*|^4] < \infty$; (iii) compute estimators and/or test statistic given the bootstrap sample $(y_t^*, X_t)$, $t = 1, ..., n$; (iv) repeat Steps (ii)-(iii), $B \geq 1$ times, and use the empirical distributions to obtain confidence intervals and/or critical values.

While it is outside of the scope to establish formally the validity of this bootstrap procedure, we expect that consistency can be shown along the lines of Franke et al (2002) and Li (2005) for the estimators and test statistics respectively.

# 6 A Simulation Study

In this section, we examine the finite-sample performances of our estimators and test statistics. We consider a bivariate model,

$$y_t = \beta_{1,t} X_{1,t-1} + \beta_{2,t} X_{2,t-2} + \sigma_t z_t,$$

where $X_t$ solves a VAR(2), $X_t = AX_{t-1} + \eta_t$ with $A$ chosen to be in the stationary range. We are interested in testing the hypothesis $H_0 : \beta_{1,t} = \beta_1$, and will investigate both size and power of our tests. Throughout (whether we work under null or alternative), we impose either of the two following DGP's on $\beta_{2,t}$:

$$\begin{aligned} \text{RW} &: \beta_{2,t} = \beta_{2,t-1} + \eta_{\beta,t}, \quad \eta_{\beta,t} \sim \text{i.i.d.} N\left(0, v_\beta^2\right); \\ \text{ST} &: \beta_{2,t} = \beta_{2,0} + \alpha \Phi\left(\frac{t/n - \mu}{\sigma}\right), \quad \Phi(\cdot) = \text{cdf of } N(0,1). \end{aligned}$$

This allows us to investigate how smoothness of the parameter trajectories affect the estimators and tests. The volatility DGP is specified as a stochatic volatility model,

$$\log \sigma_t^2 = \log \sigma_{t-1}^2 + \eta_{\sigma,t}, \quad \eta_{\sigma,t} \sim \text{i.i.d.} N\left(0, v_\sigma^2\right).$$

Finally, throughout we let the rescaled errors of the regression model be i.i.d. normally distributed, $z_t \sim N(0,1)$.

For the implementation of estimators and tests, we choose $K$ as a Gaussian kernel and the bandwidths $h$ and $b$ according to the plug-in rule described in Section 4. The semiparametric estimators and test statistics are computed with both $\hat{w}_t = 1$ and $\hat{w}_t = \hat{\sigma}_t^{-2}$, and their critical values are evaluated using the Wild bootstrap outlined in Section 5. We consider sample sizes of $n = 250$, 500 and 1000. In order to compare the performance across different sample sizes and simulations, we compute one (random) trajectory of $\beta_{2,t}$ and $\sigma_t^2$ and keep those fixed throughout. This mimicks the theoretical results in the paper which are developed conditional on the particular trajectories of the varying coefficients.

We first investigate the performance when the null is true such that $\beta_{1,t}$ is constant. Figures 1 and 2 report the performance of the fully nonparametric estimators of $\beta_{2,t}$ under the two different DGPS, RW or ST. From Figure 1 we see that while the estimator, by its nature, cannot completely track the discontinuous random walk time series, it still captures the overall structural change in the parameter quite precisely. It is also worth noting that the estimator works well even for small sample sizes ($n = 250$) and most of the improvement as the sample size grows is in terms of variance. Similar findings are reported in Figure 2 where $\beta_{2,t}$ follows a smooth transition. The overall bias is significantly smaller compared to Figure 1 though since the trajectory now is a smooth function of time.

Table 1 reports biases, standard deviations and root-MSE's (RMSE's) of the unweighted and weighted semiparametric estimators of $\beta_1$. For comparison, we also report results for the

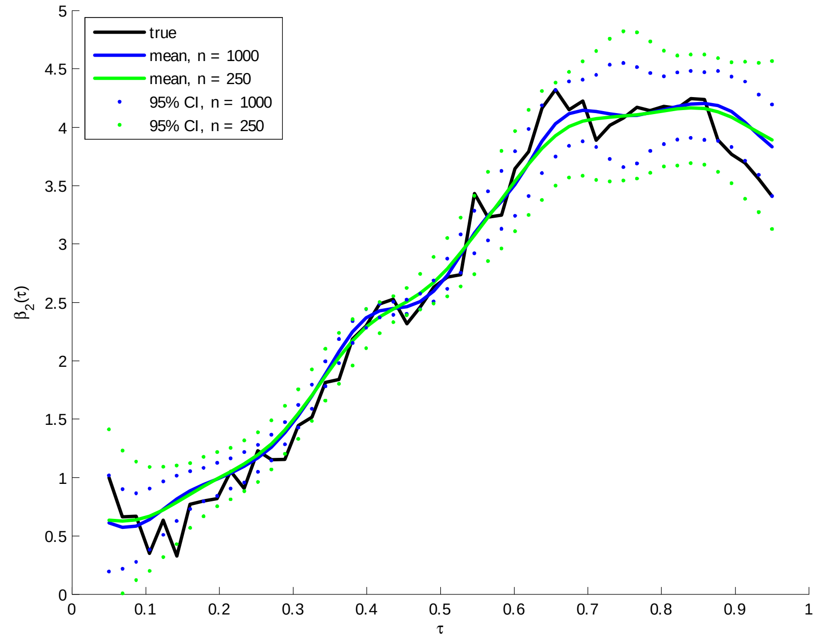


Figure 1: Simulation study, performance of estimator when $\beta_{2,t}$ follows a random walk, $n = 250$ and 1000.

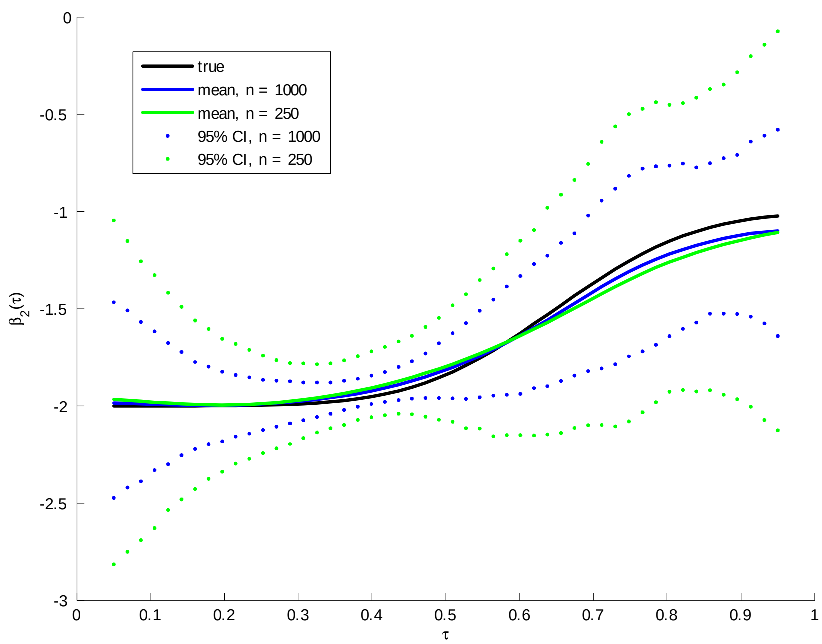


Figure 2: Simulation study, performance of estimator when $\beta_{2,t}$ follows a smooth transition, $n = 250$ and 1000.

infeasible OLS estimator which assumes knowledge of $\beta_{2,t}$ and $\sigma_t$. As expected the infeasible estimator clearly dominates the two semiparametric estimators in finite samples. But, as the theory predicts, as sample size grows these differences vanish. The semiparametric estimators are doing very well for all sample sizes with small biases and variances. Moreover, as also predicted by theory, the weighted version does better in terms of variance compared to the unweighted one in all cases but one.

| | $\beta_{2,t}$ random walk | | | $\beta_{2,t}$ smooth transition | | |
|---|---|---|---|---|---|---|
| $n$ | Infeasible | Unweighted | Weighted | Infeasible | Unweighted. | Weighted |
| | -0.93 | -0.64 | -1.96 | -0.87 | -1.65 | -1.50 |
| 250 | 28.06 | 46.61 | 41.04 | 29.50 | 42.60 | 31.48 |
| | 28.07 | 46.62 | 41.08 | 29.51 | 42.65 | 31.53 |
| | 0.12 | 0.57 | -0.57 | -0.21 | -0.88 | -0.91 |
| 500 | 19.67 | 31.46 | 27.57 | 20.52 | 30.10 | 24.38 |
| | 19.67 | 31.47 | 27.58 | 20.52 | 30.12 | 24.39 |
| | 0.06 | -0.70 | -0.52 | 0.05 | -0.31 | -0.65 |
| 1000 | 14.40 | 22.32 | 18.97 | 14.89 | 20.30 | 21.48 |
| | 14.40 | 22.33 | 18.98 | 14.89 | 20.30 | 21.48 |

Table 1: Bias, standard devision and RMSE of semiparametric estimators.
Note: In each cell, bias, standard devision and RMSE are reported. All numbers have been scaled up by a factor $10^3$.

Finally, we consider how the tests perform. In Table 2, we report sizes for the bootstrap tests based on weighted and unweighted statistics respectively. As we see, in terms of size, none of the two tests dominate the other with both having good size properties. As expected, the size in general improves as sample size grows. It is also notworthy, that size is better for the random walk model compared to the smooth transition one; we have no explanation for this.

To examine the power of the test, we now let both $\beta_{1,t}$ and $\beta_{2,t}$ vary over time. We examine the power under two different specifications of their dynamics - either they are random walks or smooth transitions. In Table 3, the powers of the two tests are reported for the two different specification. In general, the power increases with sample size as expected with good power for moderate and large samples. As expected, the weighted test has signficantly better power compared to the unweighted test, and so is better at detecting deviations from the null. The power depends on the underlying data-generating mechanism and the tests do better with the random walk specification despite the fact that the non- and semiparametric estimators are better at tracking the parameters under the smooth transition DGP. This is probably due to the fact that the realized variation of $\beta_{1,t}$ in the random walk specification is larger (between -1 and 2) compared to the one of the smooth transition specification (between 0 and 0.5).

| | $\beta_{2,t}$ random walk | | | $\beta_{2,t}$ smooth transition | | |
|---|---|---|---|---|---|---|
| $n$ | $p = 1\%$ | $p = 5\%$ | $p = 10\%$ | $p = 1\%$ | $p = 5\%$ | $p = 10\%$ |
| 250 | 1.4 | 5.9 | 8.4 | 2.2 | 7.7 | 13.2 |
| | 1.9 | 4.8 | 6.9 | 2.0 | 7.4 | 15.1 |
| 500 | 1.3 | 4.8 | 8.6 | 2.3 | 5.5 | 10.7 |
| | 1.2 | 4.1 | 8.1 | 1.8 | 5.8 | 10.9 |
| 1000 | 1.1 | 4.7 | 9.0 | 1.4 | 5.5 | 11.3 |
| | 1.0 | 4.1 | 9.7 | 1.3 | 5.4 | 11.2 |

Table 2: Size of tests using weighted and unweighted statistics.
Note: In each cell, the top and bottom number is size of weighted and unweighted test respectively.

| | $\beta_{1,t}$ and $\beta_{2,t}$ random walk | | | $\beta_{1,t}$ and $\beta_{2,t}$ smooth transition | | |
|---|---|---|---|---|---|---|
| $n$ | $p = 1\%$ | $p = 5\%$ | $p = 10\%$ | $p = 1\%$ | $p = 5\%$ | $p = 10\%$ |
| 250 | 35.8 | 48.0 | 59.4 | 26.6 | 34.1 | 43.5 |
| | 30.2 | 44.8 | 56.4 | 20.8 | 27.4 | 35.7 |
| 500 | 51.0 | 58.7 | 66.1 | 32.9 | 42.6 | 55.1 |
| | 44.9 | 50.3 | 59.8 | 25.6 | 33.2 | 48.4 |
| 1000 | 69.0 | 77.5 | 83.2 | 49.1 | 66.7 | 75.3 |
| | 57.6 | 58.7 | 70.7 | 38.3 | 54.3 | 65.8 |

Table 3: Power of tests using weighted and unweighted statistics.
Note: In each cell, the top and bottom number is power of weighted and unweighted test respectively.

To conclude, the non- and semiparametric estimators perform well for both small, moderate and large sample sizes with small biases and variances. The tests also show good performance with precise size and good power properties. In general, the weighted versions outperforms the unweighted ones which is in accordance with theory.

## 7 Empirical Applications

We employ the nonparametric techniques developed in the previous sections to investigate whether structural changes occurred in US productivity and the Eurodollar term structure. For an application to Fama-French type factor models for stock returns, we refer to Ang and Kristensen (2009).

### 7.1 US Productivity

Hansen (2001) examined whether US productivity exhbited structural changes this issue within the framework of parametric structural break models. He found that data supported one significant break in 1992 with the possibility of two more breaks in 1963 and 1982 respectively. The aim here is to see whether this is supported by the nonparametric estimators and tests. For comparison, we use the same data set for US productivity as in Hansen (2001) and refer to this paper for a more detailed description of data. Here, it suffices to say that the data is monhtly over the period of 1947 to 2001 giving us a total of $n = 651$ observations.

As in Hansen (2001), we model US productivity, $y_t$, by a time-varying AR($k$) model,

$$y_t = \mu_t + \sum_{i=1}^{k} \rho_{i,t} y_{t-i} + \sigma_t z_t.$$

We start out with $k = 3$ lags, and test for whether the 2nd and 3rd lags are significant using the bootstrapped version of the GLR test; we accept the null of $H_0 : \rho_{2,t} = \rho_{3,t}$ at a 5 and 10% level with a $p$-value of 18.3%. In the following, we therefore maintain an AR(1) model. For the AR(1) model, we examine how the fully nonparametric estimators of $\mu_t$ and $\rho_t = \rho_{1,t}$ perform in comparison to the one- and three-breaks AR(1) models estimated in Hansen (2001). In Figures 3-4, the nonparametric trajectories as obtained using our nonparametric estimators are plotted together with corresponding ones obtained from the two fitted structural break model. As an informal test of whether the two parametric models are consistent with our nonparametric estimates, we have also included pointwise 95% confidence intervals for the nonparametric estimators. Figure 3 shows the trajectory of the intecept, $\mu_t$, and we see that the nonparametric estimator supports the three break model with the red trajectory staying within the 95% confidence interval for the whole sample period. The one-break model on the other hand lies outside during the period 1985-1995. The same picture appears when examining the variation in $\rho_t$ as plotted in Figure 4: Again, the 3-break parametric model appears to be consistent with the nonparametric estimates while there is some evidence that the one-break model is not fully adequate in describing the parameter variation. It is also worth noting that while the nonparametric estimator is not able to capture the potential time-variation of the intercept very precisely, as the wide confidence intervals in Figure 1 indicate, it performs well for the AR coefficient with much tighther confidence intervals.

Since the confidence intervals are only pointwise, the above two figures cannot be used as formal statistical evidence for validity of the two parametric models. Instead, we implement the proposed bootstrap tests with the two nulls being that $\beta_t = (\mu_t, \rho_t)'$ either follows the one- or three-break model. For the one-break model, we obtain $F_1 = 16.25$ with the 5% critical value being 17.33; thus we only just accept the null with a $p$-value of 7.0%. This is in accordance with the plots that showed that the one-break model lead to parameter trajectories that were not fully supported by the nonparametric estimators. In contrast, for

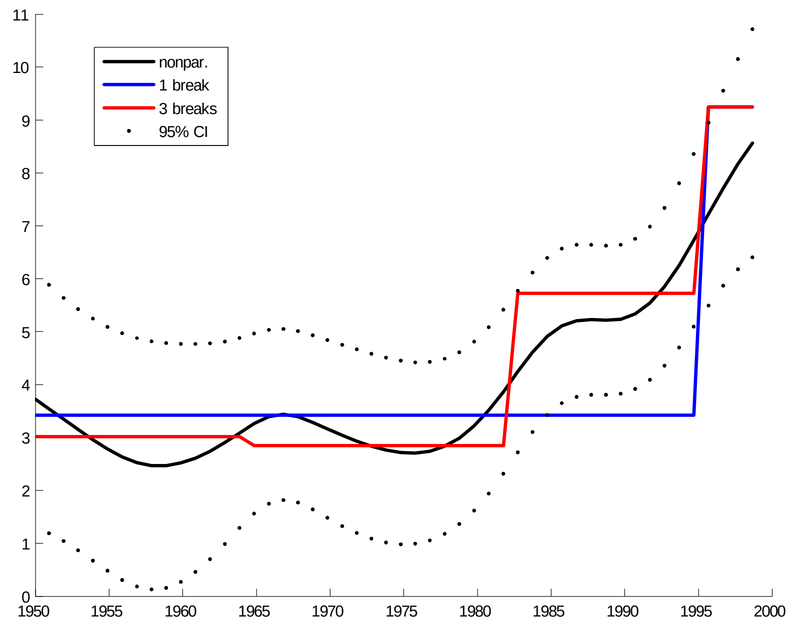


Figure 3: Estimates of structural change in $\mu$, 1950-2000.

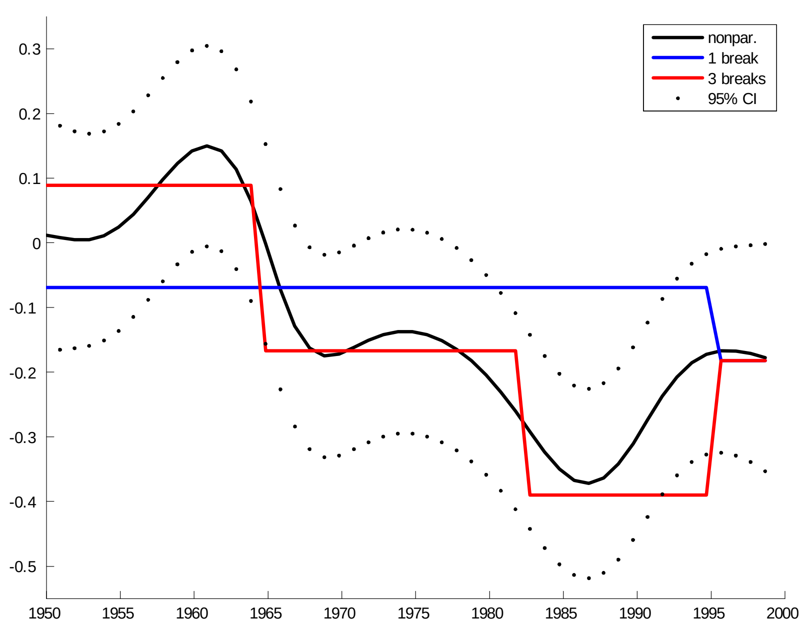


Figure 4: Estimates of time variation in $\rho$, 1950-2000.

the three-breaks model, the test yields $F_1 = 1.86$ with a 5% critical value of 16.48, and so we overwhelmingly accept the null of a three break model; the corresponding $p$-value is 96.5%. Our findings complement the analysis of Hansen (2001) who finds "a structural break in 1994, and possibly breaks in Dec. 1963 and Jan. 1982," and reports that the two latter break time points are very imprecisely estimated. Our nonparametric analysis shows that both models are supported by data, but that the one-break model is close to being rejected.

As can also be seen from the nonparametric estimates and their confidence intervals, there is not very strong support for breaks in the intercept while there appears to be strong evidence for breaks in the AR coefficients. We therefore now test the hypothesis that $\mu_t = \mu$ is constant against the nonparametric alternative using our semiparametric estimators. Under the null we obtain $\hat{\mu} = 4.37$ with 95% confidence interval being $(3.10, 5.14)$. Comparing the nonparametric and semiparametric model, we obtain $F_1 = 11.33$ with a 5% critical value of 8.93 and a $p$-value of 1.25%. Thus, we reject at a 5% level but not at a 1% level. In comparison, we strongly reject the hypothesis that $\rho_t = \rho$ is constant with $F_1 = 47.61$ and a 1% critical value of 14.85.

In conclusion, our nonparametric approach supports the findings of Hansen (2001) finding that a 3-breaks model adequate captures the time-variation in the regression coefficients, while a 1-break model may be too simple. Moreover, our techniques also shows that most of the time-variation is found in the AR coefficient while there is not as strong support for time-variation in the intercept.

### 7.2 Eurodollar Term Structure

Affine factor models are widely used in empirical finance to describe the dynamics of the yield curve. Affine term structure models assume that the short-term interest rate, $y_t$, is driven by a linear combination of factors with the loadings normally assumed to be constant. However, there is amble empirical evidence that the factor loadings are varying over time. This can have major implications for forecasting the yield curve and for bond pricing. Most studies examining time-variation in the factor loading take a parametric approach using, for example, Markov switching or random walk models to describe the possible time variation (see Ang and Bekaert, 2002; Bhansal and Zhou, 2002). Due to the numerical complications with estimating dynamic models with latent variables, most studies have confined themselves to single-factor models despite the fact that the consensus is that multiple factors are needed to adequately describe the yield curve dynamics. In contrast, the nonparametric techniques developed in this paper are straight-forward to implement even in a multi-factor setting.

The data set that we will use consists of $n = 8669$ daily observations of the 1, 3, and 6 months Eurodollar yield in the period 1971-2004. We use the 1-month interest rate as a proxy for the short-term rate, and choose our factors as the lagged level, slope and curvature of the yield curve as constructed from the three observed yields. These are plotted in Figure

5. All three series appear to be somewhat unstable over time with particularly the so-called Fed Experiment of the early 1980's having a large impact on the their dynamics.

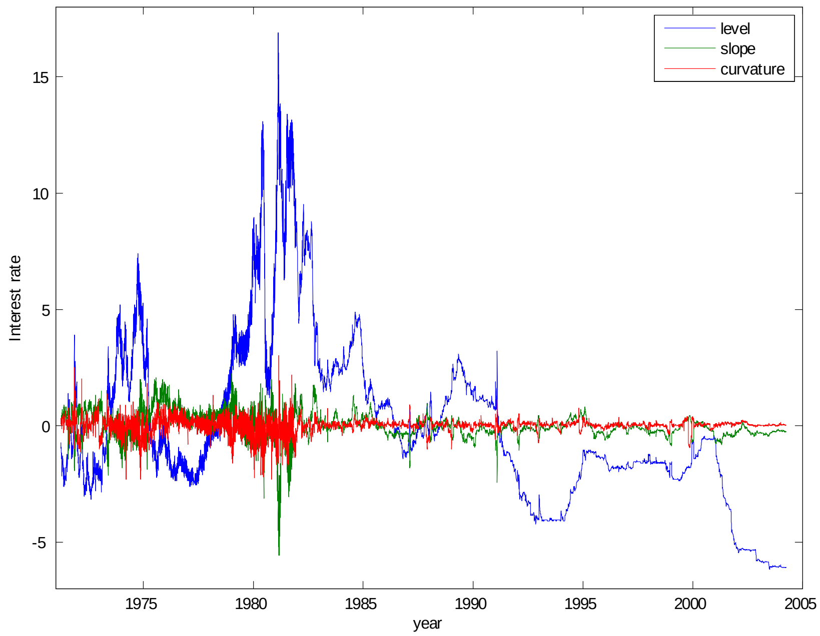


Figure 5: Time series plot of the level, slope and curvature factor, 1971-2004.

Next, we estimate the three-factor model allowing all loadings and the intercept to be time-varying. The resulting estimates of the intercept and the loadings over time are shown in Figures 6-8. These estimates deliver an informal rejection of the null of constant factor loadings with all three exhibiting substantial variation over time. In particular, the Fed Experiment changed the yield curve dynamics quite dramatically with all factor loadings showing pronounced changes. In the same period, the volatility of the short-term interest rate increased substantially which explains the wider pointwise confidence bands for the estimates in this period. Another interesting feature is that from 1995 and onwards, the loadings for the level and slope factors have stabilized and (based on the pointwise confidence intervals) we cannot reject that these two are constant for this period. Moreover, we cannot reject that the level factor is insignficant from 1994 and onwards. On the other hand, the curvature loading exhibits a significant change in the early 1990's and shows pronounced variation through the latter period.

For comparison, we also report the OLS estimates for each of the factor loadings in the three figures. We see that all the OLS estimators at some point lie outside of the pointwise confidence intervals. Treating this as an informal test for constant factor loadings, we therefore reject that they are constant over the sample period. More formally, we proceed

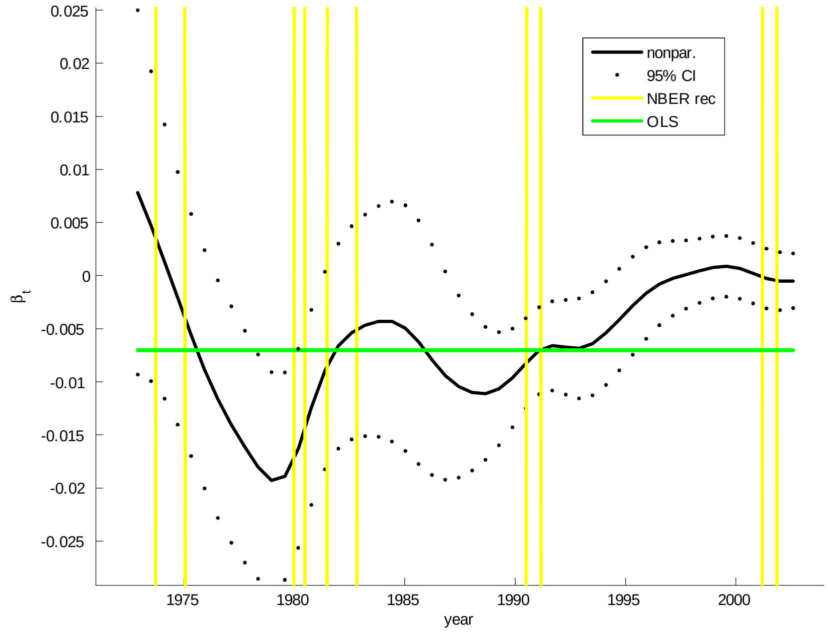


Figure 6: Estimated level factor loading with pointwise 95% confidence bands, 1971-2004.

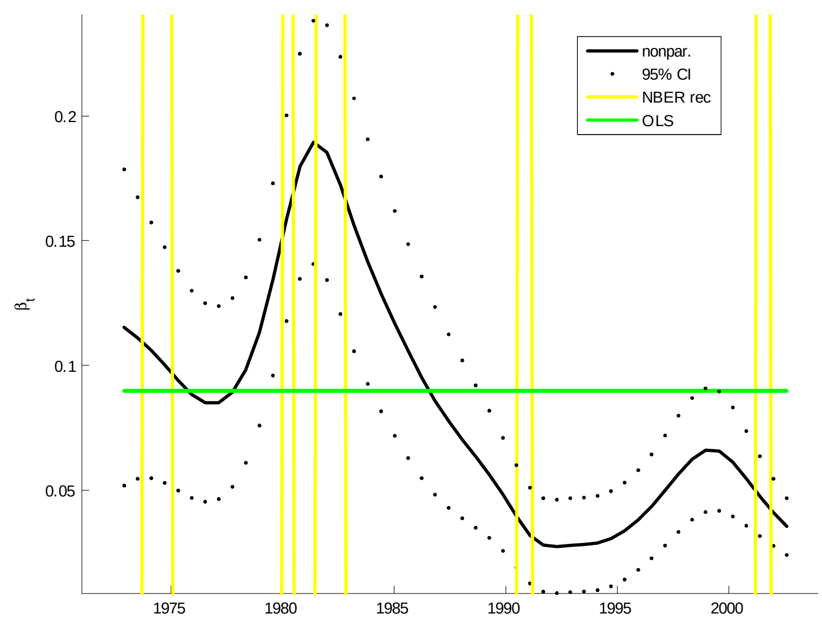


Figure 7: Estimated slope factor loading with pointwise 95% confidence bands, 1971-2004.

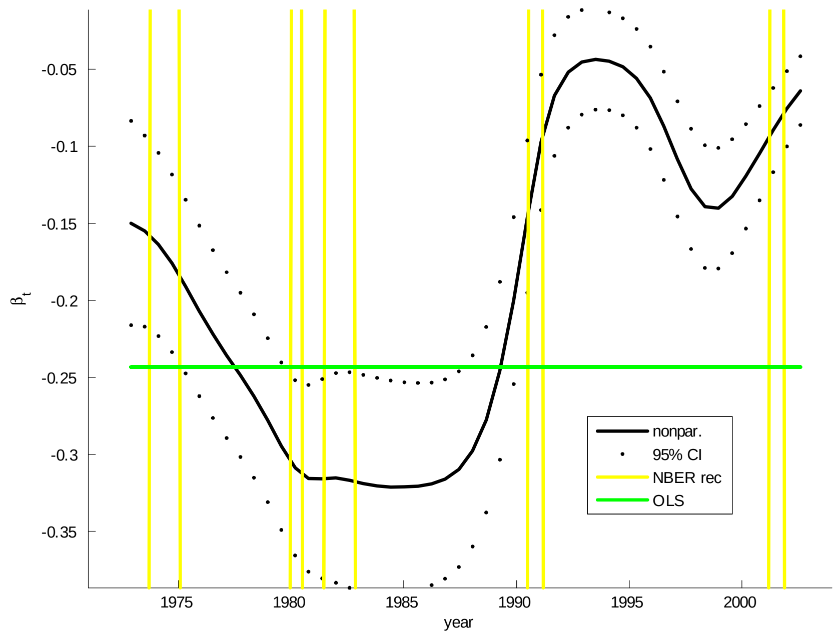


Figure 8: Estimated curvature factor loading with pointwise 95% confidence bands, 1971-2004.

to carry out a battery of tests regarding the time-variation in the loadings based on the test statistics proposed in the paper. First, we test the three hypotheses that any pair of the factor loadings are constant; we strongly reject all three nulls with $p$-values well below 1%. Next, we test whether any of the loadings individually is constant; again, we heavily reject the three nulls. Finally, the null of all loadings being constant is strongly rejected. In conclusion, the tests support the informal eyeballing test and we conclude that there is strong evidence of time-variation in all factor loadings. If on the other hand, we conduct the test for the subperiod of 1994-2004, we accept the null of no time-variation in the loadings for the level and slope factor. The corresponding test for the curvature factor is rejected on the other hand. Thus, a reasonable model for the recent Eurodollar term structure has constant level and slope loadings, while the curvature loading remains time-varying.

The above analysis is silent about the underlying causes for the time-variation in the loadings. In Figures 6-8, we also report the NBER recessions. There appears to be some correlation between whether the economy is in a recession and changes in the factor loadings, but the sign is not clear. In addition, other macro factors may also influence the variation. We therefore carry out an informal regression analysis where we treat the estimated factor loading paths as observed dependent variables and regress them onto the NBER recession indicator, US productivity and US inflation; a similar two-step procedure in a continuous-time setting was proposed and analyzed in Kanaya and Kristensen (2010). The three chosen macro regressors are only observed at a monthly frequency, but this causes no problems since we can estimate the factor loadings at any given frequency. The results of those second stage regressions are reported in Table 4. In general, the NBER recession and inflation are good predictors of the variation in the coefficients while US productivity is less informative. It should be noted though that the reported standard errors do not take into account the estimation error in the factor loading, and so the results probably over estimate the significance of the macro variables. The over all $R^2$ is ranges between 36%-51% and so substantial parts of the estimated variation in the loadings are explained by underlying macro factors. Looking at the individual regresssion coefficients, we see that recessions tend to increase the loadings for all three factors, while inflation and productivity have negative impacts on the level and curvature loadings of the yield curve, but positive impact on the slope coefficient.

| | Factor loadings | | |
|---|---|---|---|
| | level | slope | curvature |
| NBER recession | 0.5056 | 0.4037 | 3.4020 |
| | (0.0701) | (0.5989) | (1.0456) |
| US inflation | -0.1564 | 1.1966 | -2.3476 |
| | (0.0110) | (0.0603) | (0.1144) |
| US productivity | -0.0036 | 0.0518 | -0.0747 |
| | (0.0028) | (0.0216) | (0.0490) |
| $R^2$ | 0.466 | 0.510 | 0.362 |

Table 4: Second-stage regression of factor loadings onto macro variables.
Note: All regression coefficients and SE's have been scaled up by a factor $10^2$.

## 8 Conclusion

A general theory has been developed for the semi-nonparametric estimation and testing of partially time-varying regression models. A number of extensions would be of interest:

So far we have assumed that $\beta(\tau)$ is a smoothly varying function. It would be interesting to extend our estimators and results to allow for finite number of break points/discontinuities in the evolution of $\beta(\tau)$. We conjecture that the current semiparametric estimators and tests are robust to this situation: First, Theorem 3.1 remains valid at any continuity point. Thus, it seems plausible that Theorems 3.2-3.4 hold true even if $\beta(\tau)$ exhibits breaks as long as there is only a finite number of those in the sample: The nonparametric estimator will be inconsistent in shrinking neighbourhoods of the break points, and since there is only a finite number of those, we expect the biases to vanish asymptotically in the estimator and test statistics.

However, in finite samples, one would expect that the performance of estimators and tests could be improved by trying to adjust for jumps. Moreover, in situations where we reject $H_1$ or $H_2$, it is often of interest to identify abrupt changes (breaks) in the time-varying coefficients. In the nonparametric regression literature, procedures for the estimation and testing of jumps have been developed which should be adaptable to our setting. We refer to Gijbels (2003) for a review of existing methods, and Gijbels et al (2007) for a recent proposal; see also Casas and Gijbels (2009) for an application to nonparametric volatility estimators allowing for jumps.

The theory for the semiparametric estimators and test statistics require the bandwidths to vanish at non-standard rates. Data-driven bandwidth selection procedures for these are currently not available, and it would be highly useful to develop and analyze such.

Finally, the proposed estimators and test statistics are straightforward to extend to non-linear models whose time-invarying parameters can be characterized as minimizers of an objective function taking the form of a (time-varying) population moment. In this class of models, estimators of the time-varying parameters can be defined as minimizers of a kernel-weighted version of the corresponding sample moment. The theoretical analysis of this broader class of estimators is left for future research.

## References


Aït-Sahalia, Y., J. Fan and H. Peng (2009) "Nonparametric Transition-Based Tests for Jump-Diffusions," *Journal of the American Statistical Association* 104, 1102-1116.

Akharif, A. and M. Hallin (2003) "Efficient Detection of Random Coefficients in Autoregressive Models," *Annals of Statistics* 31, 675-704.

Andrews, D.W.K. (1993) "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61, 821–856.

Andrews, D.W.K. and W. Ploberger (1994) "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative," *Econometrica* 62, 1383-1414.

Ang, A. and G. Bekaert (2002) "Short Rate Nonlinearities and Regime Switches," *Journal of Economic Dynamics & Control* 26, 1243–1274.

Ang, A. and D. Kristensen (2009) "Testing Conditional Factor Models," CREATES Research Papers 2009-09.

Bansal, R. and H. Zhou (2002) "Term Structure of Interest Rates with Regime Shifts," *Journal of Finance* 57, 1997-2043.

Bai. J. (1999) "Likelihood Ratio Tests for Multiple Structural Changes," *Journal of Econometrics* 91, 299-323.

Bai, J. and P. Perron (1998) "Testing for and Estimation of Multiple Structural Changes," *Econometrica*, 66, 47-79.

Brown, R. L., J. Durbin and J.M. Evans (1975) "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society*, Series B, 37, 149-192.

Cai, Z. (2007) "Trending Time-varying Coefficient Time Series Models with Serially Correlated Errors," *Journal of Econometrics*, 136, 163–188.

Casas, I. and I. Gijbels. (2009) "Unstable Volatility Functions: the Break Preserving Local Linear Estimator," CREATES Research Paper 2009-48.

Chan, K.S. (1990) "Testing for Threshold Autoregression," *Annals of Statistics* 18, 1886-1894.

Chen, B. and Y. Hong (2009) "Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression," manuscript, Department of Economics, University of Rochester.

Chu, C.-S. J., K. Hornik and C.M. Kuan (1995) "The Moving-Estimates Test for Parameter Stability," *Econometric Theory* 11, 699-720.

McConnell, M.M. and G. Perez-Quiros (2000) "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?" *American Economic Review* 90, 1464-1476.

Dahlhaus, R. (1997) "Fitting time series models to nonstationary processes," *Annals of Statististics* 16, 1-37.

Denker, M. and G. Keller (1983) "On $U$-Statistics and von Mises' Statistics for Weakly Dependent Processes," *Zeitschrift für Wahrkeinlichkeitstheorie*, 64, 505-522.

Elliott, G. and U. Müller (2006): "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economic Studies* 73, 907-940.

Fan, J. and T. Huang (2005) "Profile Likelihood Inferences on Semiparametric Varying-coefficient Partially Linear Models," *Bernoulli* 11, 1031-1057.

Fan, Y. and Q. Li (1999) "Central Limit Theorem for Degenerate U-Statistics of Absolutely Regular Processes with Applications to Model Specification Testing," *Journal of Nonparametric Statistics* 10, 245-271.

Fan, J., and Q. Yao (1998) Efficient Estimation of Conditional Variance Functions in Stochastic Regression. *Biometrika* 85, 645-660.

Fan, J. and W. Zhang (1999) "Statistical Estimation in Varying Coefficient Models," *Annals of Statistics* 27, 1491-1518.

Fan, J. and W. Zhang (2008) "Statistical Methods with Varying Coefficient Models," *Statistics and Its Interface* 1, 179–195.

Fan, J., C. Zhang and J. Zhang (2001) "Generalized Likelihood Ratio Statistics and Wilks Phenomenon," *Annals of Statistics* 29, 153–193.

Ferson,W. and R. Schadt (1996) "Measuring Fund Strategy and Performance in Changing Economic Conditions," *Journal of Finance* 51, 425-462.

Franke, J., J-P. Kreiss and E. Mammen (2002) "Bootstrap of Kernel Smoothing in Nonlinear Time Series," *Bernouilli* 8, 1-37.

Gao, J., and M. King (2004) "Model Specification Testing in Nonparametric and Semiparametric Time Series Econometrics," technical report, University of Western Australia.

Gijbels, I. (2003) "Inference for Nonsmooth Regression Curves and Surfaces Using Kernel-based Methods," in *Recent Advances and Trends in Nonparametric Statistics* (eds. M.G. Akritas and D.N. Politis), p. 183-202, Elsevier (North Holland).

Gijbels, I., A. Lambert and P. Qiu (2007) "Jump-preserving Regression and Smoothing Using Local Linear Fitting: A Compromise," *Annals of the Institute of Statistical Mathematics* 59, 235–27.

Hamilton, J.D. (1989) "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica* 57, 357-384.

Hansen, B.E. (1992) "Testing for Parameter Instability in Linear Models," *Journal of Policy Modeling* 14, 517-533.

Hansen, B.E. (2000a) "Sample Splitting and Threshold Estimation," *Econometrica* 68, 575-603.

Hansen, B.E. (2000b) "Testing for Structural Change in Conditional Models," *Journal of Econometrics* 97, 93-115.

Hansen, B. E. (2001) "The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity," *Journal of Economic Perspectives* 15, 117-128.

Härdle, W., P. Hall, and J. S. Marron (1988) "How Far Are Automatically Chosen Regression Smoothing Parameters from Their Optimum?" *Journal of the American Statistical Association*, 83, 86-101.

Hart, J. D. (1991) "Kernel Regression Estimation with Time Series Errors," *Journal of the Royal Statistical Society, Series B (Methodological)* 53, 173-187.

Hong, Y. and Y.-J. Lee (2008) "A Loss Function Approach to Model Specification Testing and Its Relative Efficiency to the GLR Test," manuscript, Dep. of Economics, Indiana University.

Ingster, Y.I. (1993) "Asymptotic Minimax Hypothesis Testing for Nonparametric Alternatives I-III," *Mathematical Methods of Statistics* 2, 85-114; 3, 171-189; 4, 249-268.

Juhl, T. and Z. Xiao (2005) "A Nonparametric Test for Changing Trends," *Journal of Econometrics* 127, 179-199.

Kanaya, S. and D. Kristensen (2010) "Estimation of Stochastic Volatility Models By Nonparametric Filtering," CREATES Research Papers 2010-67, Aarhus University.

Kristensen, D. (2009) "Uniform Convergence Rates of Kernel Estimators with Heterogeneous, Dependent Data," *Econometric Theory* 25, 1433-1445.

Kristensen, D. (2010) "Nonparametric Filtering of the Realized Spot Volatility: A Kernel-based Approach," *Econometric Theory* 26, 60-93.

Kristensen, D. (2011) "Stationary Approximations of Time-inhomogenous Markov Chains with Applications," manuscript, Dep. of Economics, Columbia University.

Li, D. (2005) "A Simple Bootstrap Test for Time Series Regression Models," *Journal of Nonparametric Statistics* 17, 513-520.

Lin, C.-F. J., and T. Teräsvirta (1994) "Testing the Constancy of Regression Parameters Against Continuous Structural Change," *Journal of Econometrics* 62, 211–228.

McLeish, D. L. (1974) "Dependent Central Limit Theorems and Invariance Principles," *Annals of Probability* 2, 620-628.

Nyblom, J. (1989) "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association* 84, 223-230.

Orbe, S., E. Ferreira and J. Rodriguez-Poo (2004) "Nonparametric Estimation of Time Varying Parameters Under Shape Restrictions," *Journal of Econometrics*, 126, 53-77.

Opsomer, J. Y. Wang and Y. Yang (2001) "Nonparametric Regression with Correlated Errors," *Statistical Science*, 16, 134-153

Pollard, D. (1990) *Empirical Processes and Their Applications.* Institute of Mathematical Statistics, Hayward, California, USA.

Robinson, P.M. (1987) "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica* 55, 875-91.

Robinson, P.M. (1988) "Root-N-Consistent Semiparametric Regression," *Econometrica* 56, 931-954.

Robinson, P.M. (1989) "Nonparametric Estimation of Time-Varying Parameters," in *Statistical Analysis and Forecasting of Economic Structural Changes* (ed. P. Hackl), p. 253-264.

Robinson, P.M. (1991) "Time-Varying Nonlinear Regression," in *Economic Structural Change in Analysis and Forecasting* (eds. P. Hackl and A. Westlund), Springer-Verlag (1991), 179-190.

Stock, J., and M. Watson (1996) "Evidence of Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics* 14, 11-30.

Subba Rao, S. (2006) "On some nonstationary, nonlinear random processes and their stationary approximations," *Advances in Applied Probability* 38, 1155-1172.

Wooldridge, J. M., and H. White (1988) "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory* 4, 210-230.

# A Proofs

In the following we will for notational convenience often suppress the dependence of the variables on $n$ and for example write $X_t$ for $X_{n,t}$.

**Proof of Theorem 3.1.** Write $K_{t,\tau} = K_h(t/n - \tau)$, $X_t = X_{n,t}$ and similar for other variables. We have

$$\begin{aligned}\hat{\beta}(\tau) - \beta(\tau) &= \left(\sum_{t=1}^{n} K_{t,\tau} X_t X_t'\right)^{-1} \sum_{t=1}^{n} K_{t,\tau} X_t X_t' \{\beta_t - \beta(\tau)\} \\ &\quad + \left(\sum_{t=1}^{n} K_{t,\tau} X_t X_t'\right)^{-1} \sum_{t=1}^{n} K_{t,\tau} X_t \varepsilon_t.\end{aligned}$$

By Lemma C.1, we obtain $n^{-1}\sum_{t=1}^{n} K_{t,\tau} X_i X_i' = \Lambda(\tau) + o_P(1)$, while

$$E\left[\left\|\frac{1}{n}\sum\nolimits_{t=1}^{n} K_{t,\tau} X_t X_t' \{\beta_t - \beta(\tau)\}\right\|\right] \leq C \sup_{t:|t/n-\tau|<Bh} \|\beta(t/n) - \beta(\tau)\| = O(h^r),$$

where we have used the smoothness assumption imposed on $\beta(t)$. Thus, the following representation holds uniformly over $\tau \in (a, 1-a)$:

$$\hat{\beta}(\tau) - \beta(\tau) = \Lambda^{-1}(\tau) \frac{1}{n} \sum_{t=1}^{n} K_{t,\tau} X_t \varepsilon_t + O_P(h^r) \tag{A.1}$$

To complete the proof, we show that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_{n,t} \to^d N\left(0, \left\|K^2\right\| \Lambda(\tau) \sigma^2(\tau)\right), \quad nh \to \infty, \tag{A.2}$$

where $u_{n,t} = \sqrt{h} K_{t,\tau} X_t \varepsilon_t$. This is done by verifying the conditions of Lemma C.4: First, we note that $u_{n,t}$ is a MGD w.r.t. $\mathcal{F}_t = \mathcal{F}(X_t, z_t, X_{t-1}, z_{t-1}, \ldots)$. Furthermore, it satisfies

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} E\left[u_{n,t} u_{n,t}'\right] &= \frac{h}{n}\sum_{t=1}^{n} K_{t,\tau}^2 \sigma_t^2 \Lambda_t + o(1) \\ &= \frac{1}{h}\int K^2\left(\frac{s-\tau}{h}\right) \sigma^2(s) \Lambda(s)\, ds + o(1) \\ &= \left\|K^2\right\| \sigma^2(\tau) \Lambda(\tau) + o(1)\end{aligned}$$

and, as $nh \to \infty$,

$$\begin{aligned}\frac{1}{n^{1+\delta/2}}\sum_{t=1}^{n}E\left[\|u_{n,t}\|^{2+\delta}\right] &= \frac{h^{1+\delta/2}}{n^{1+\delta/2}}\sum_{t=1}^{n}K_{t,\tau}^{2+\delta}\sigma_t^{2+\delta}E\left[\|X_t\|^{2+\delta}|z_t|^{2+\delta}\right]\\ &= C\frac{1}{(nh)^{\delta/2}}\sigma^{2+\delta}(\tau)\int K^{2+\delta}(z)\,dz\\ &= o(1).\end{aligned}$$

■

**Proof of Theorem 3.2.** Define for any two sequences $A_t$ and $B_t$ and any weighting function $w$,

$$S_{A,B}^{w}=\frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t(a)\,w_tA_tB_t',$$

where $\mathbb{I}_t(a)=\mathbb{I}\{a\le t/n\le 1-a\}$, and let $S_A^w=S_{A,A}^w$. We may then write $\tilde{\beta}_1^w$ and the corresponding estimator based on known weights, say $\bar{\beta}_1^w$, as

$$\tilde{\beta}_1^w=\left(S_{X_1-\hat{X}_1}^{\hat{w}}\right)^{-1}S_{X_1-\hat{X}_1,y-\hat{y}}^{\hat{w}},\quad \bar{\beta}_1^w=\left(S_{X_1-\hat{X}_1}^{w}\right)^{-1}S_{X_1-\hat{X}_1,y-\hat{y}}^{w}.$$

We write $\tilde{\beta}_1^w-\beta_1=\left\{\bar{\beta}_1^w-\beta_1\right\}+\left\{\tilde{\beta}_1^w-\bar{\beta}_1^w\right\}$, and now show that:

$$\sqrt{n}\left\{\bar{\beta}_1^w-\beta_1\right\} \to^{d} N\left(0,\Sigma_w^{-1}\Phi_w\Sigma_w^{-1}\right), \tag{A.3}$$

$$\sqrt{n}\left\{\tilde{\beta}_1^w-\bar{\beta}_1^w\right\} = o_P(1). \tag{A.4}$$

**Proof of eq. (A.3):** Define

$$V_t:=X_{1,t}-\xi_t, \tag{A.5}$$

where

$$\xi_t:=M_t(X_1)'X_{2,t}, \tag{A.6}$$

and $M_t(X_1)$ is defined in eq. (2.8). We then have

$$E\left[X_{2,t}y_t'\right]=E\left[X_{2,t}X_{1,t}'\right]\beta_1+E\left[X_{2,t}X_{1,t}'\right]\beta_{2,t},$$

such that

$$M_t(y)=M_t(X_1)\beta_1+\beta_{2,t},$$

and

$$y_t-M_t(y)'X_{2,t}=\beta_1'\left[X_{1,t}-M_t(X_1)'X_{2,t}\right]+\varepsilon_t=\beta_1'V_t+\varepsilon_t$$

Furthermore,

$$\hat{X}_{1,t}=\hat{\xi}_t+\hat{V}_t,\quad \hat{y}_t=\beta_1'\hat{X}_{1,t}+\bar{\beta}_{2,t}'X_{2,t}+\hat{\varepsilon}_t,$$

where

$$\bar{\beta}_{2,t} = \left[\sum_{s=1}^{n} K_{s,t} X_{2,s} X_{2,s}'\right]^{-1} \sum_{s=1}^{n} K_{s,t} X_{2,s} X_{2,s}' \beta_{2,t}, \tag{A.7}$$

such that

$$X_{1,t} - \hat{X}_{1,t} = \left\{\xi_t - \hat{\xi}_t\right\} + V_t - \hat{V}_t$$

and

$$y_t - \hat{y}_t = \beta_1' \left[X_{1,t} - \hat{X}_{1,t}\right] + \left[\beta_{2,t} - \bar{\beta}_{2,t}\right]' X_{2,t} + \varepsilon_t - \hat{\varepsilon}_t$$

In total,

$$S^w_{X_1 - \hat{X}_1} = S^w_V + S^w_{\hat{V}} + S^w_{\xi - \hat{\xi}} + 2S^w_{V,\xi-\hat{\xi}} + 2S^w_{\hat{V},\xi-\hat{\xi}} + 2S^w_{V,\hat{V}}$$

and

$$\begin{aligned} S^w_{X_1-\hat{X}_1,y-\hat{y}} &= S^w_{X_1-\hat{X}_1,\beta_1'\left\{X_1-\hat{X}_1\right\}+\left[\beta_2-\bar{\beta}_2\right]'X_2+(\varepsilon-\hat{\varepsilon})} \\ &= S^w_{X_1-\hat{X}_1}\beta_1 + S^w_{X_1-\hat{X}_1,\left[\beta_2-\bar{\beta}_2\right]'X_2} + S^w_{X_1-\hat{X}_1,\varepsilon} + S^w_{X_1-\hat{X}_1,\hat{\varepsilon}}. \end{aligned}$$

Finally,

$$S^w_{X_1-\hat{X}_1,\left[\beta_2-\bar{\beta}_2\right]'X_2} = S^w_{\xi-\hat{\xi},\left[\beta_2-\bar{\beta}_2\right]'X_2} + S^w_{V,\left[\beta_2-\bar{\beta}_2\right]'X_2} - S^w_{\hat{V},\left[\beta_2-\bar{\beta}_2\right]'X_2},$$
$$S^w_{X_1-\hat{X}_1,\varepsilon} = S^w_{\xi-\hat{\xi},\varepsilon} + S^w_{V,\varepsilon} - S^w_{\hat{V},\varepsilon}, \quad S^w_{X_1-\hat{X}_1,\hat{\varepsilon}} = S^w_{\xi-\hat{\xi},\hat{\varepsilon}} + S^w_{V,\hat{\varepsilon}} - S^w_{\hat{V},\hat{\varepsilon}}.$$

It follows from Lemmas B.1-B.6 that $S^w_V \to^P \Sigma_w$, $\sqrt{n} S^w_{V,\varepsilon} \to^d N(0, \Phi_w)$, while all others of the above terms are negiglible. This yields the desired result.

**Proof of eq. (A.4):** Observe that

$$\begin{aligned} \tilde{\beta}^w_1 &= \beta_1 + \left(S^{\hat{w}}_{X_1-\hat{X}_1}\right)^{-1}\left[S^{\hat{w}}_{X_1-\hat{X}_1,\left[\beta_2-\bar{\beta}_2\right]'X_2} + S^{\hat{w}}_{X_1-\hat{X}_1,\varepsilon} + S^{\hat{w}}_{X_1-\hat{X}_1,\hat{\varepsilon}}\right], \\ \bar{\beta}^w_1 &= \beta_1 + \left(S^{w}_{X_1-\hat{X}_1}\right)^{-1}\left[S^{w}_{X_1-\hat{X}_1,\left[\beta_2-\bar{\beta}_2\right]'X_2} + S^{w}_{X_1-\hat{X}_1,\varepsilon} + S^{w}_{X_1-\hat{X}_1,\hat{\varepsilon}}\right], \end{aligned}$$

c.f. the proof of eq. (A.3). We therefore have

$$\begin{aligned} \tilde{\beta}^w_1 - \bar{\beta}^w_1 &= \left(S^{\hat{w}}_{X_1-\hat{X}_1}\right)^{-1}\left[S^{\hat{w}}_{X_1-\hat{X}_1,\left[\beta_2-\bar{\beta}_2\right]'X_2} - S^{w}_{X_1-\hat{X}_1,\left[\beta_2-\bar{\beta}_2\right]'X_2}\right] \\ &\quad + \left(S^{\hat{w}}_{X_1-\hat{X}_1}\right)^{-1}\left[S^{\hat{w}}_{X_1-\hat{X}_1,\varepsilon} - S^{w}_{X_1-\hat{X}_1,\varepsilon}\right] \\ &\quad + \left(S^{\hat{w}}_{X_1-\hat{X}_1}\right)^{-1}\left[S^{\hat{w}}_{X_1-\hat{X}_1,\hat{\varepsilon}} - S^{w}_{X_1-\hat{X}_1,\hat{\varepsilon}}\right] \\ &\quad + \left[\left(S^{\hat{w}}_{X_1-\hat{X}_1}\right)^{-1} - \left(S^{w}_{X_1-\hat{X}_1}\right)^{-1}\right]\left[S^{w}_{X_1-\hat{X}_1,\left[\beta_2-\bar{\beta}_2\right]'X_2} + S^{w}_{X_1-\hat{X}_1,\varepsilon} + S^{w}_{X_1-\hat{X}_1,\hat{\varepsilon}}\right] \\ &=: \hat{B}^{-1}\left(\hat{A}_1 - A_1\right) + \hat{B}^{-1}\left(\hat{A}_2 - A_2\right) + \hat{B}^{-1}\left(\hat{A}_3 - A_3\right) + \left(\hat{B}^{-1} - B^{-1}\right)A_4 \end{aligned}$$

First note that, from the proof of eq. (A.3), $A_4 = O_P(1/\sqrt{n})$. Next, write

$$\hat{B}^{-1} = B^{-1} - B^{-1}\left[\hat{B} - B\right]B^{-1} + O\left(\left\|\hat{B} - B\right\|^2\right),$$

where, with $\hat{\Delta} := \sup_{a \leq \tau \leq 1-a} |\hat{w}(\tau) - w(\tau)|$,

$$\begin{aligned}
\left\|\hat{B} - B\right\| &\leq \frac{1}{n}\sum_{t=1}^{n} \mathbb{I}_t(a)\left|\hat{w}_t^2 - w_t^2\right|\left(X_1 - \hat{X}_1\right)'\left(X_1 - \hat{X}_1\right) \\
&\leq \hat{\Delta} \times \frac{1}{n}\sum_{t=1}^{n} \mathbb{I}_t(a)\left(X_1 - \hat{X}_1\right)'\left(X_1 - \hat{X}_1\right) \\
&= \hat{\Delta} \times \operatorname{tr}\left\{S_{X_1 - \hat{X}_1}\right\} \\
&= O_P\left(\hat{\Delta}\right).
\end{aligned}$$

Here, the last equality follows from the fact that, by the same reasoning as in the proof of eq. (A.3) (with $w_t = 1$), $S_{X_1 - \hat{X}_1} = O_P(1)$. This implies that $\left\|\hat{B}^{-1} - B^{-1}\right\| = O_P\left(\hat{\Delta}\right)$ and that $\hat{B}^{-1} = B^{-1} + o_P(1)$.

Similarly, employing the same the arguments as in the proofs of Lemmas B.2 and B.5,

$$\begin{aligned}
\left\|\hat{A}_1 - A_1\right\| &\leq \hat{\Delta} \times \sup_{a \leq \tau \leq 1-a}\left\|\beta_{2,t} - \bar{\beta}_{2,t}\right\| \times \frac{1}{n}\sum_{t=1}^{n} \mathbb{I}_t(a)\left\|X_{1,t} - \hat{X}_{1,t}\right\|\left\|X_{2,t}\right\| \\
&= \hat{\Delta} \times \sup_{a \leq \tau \leq 1-a}\left\|\beta_{2,t} - \bar{\beta}_{2,t}\right\| \times O_P(1),
\end{aligned}$$

Next,

$$\hat{A}_2 - A_2 = \frac{1}{n}\sum_{t=1}^{n} \mathbb{I}_t(a)\left(\hat{w}_t^2 - w_t^2\right)\left(X_1 - \hat{X}_1\right)'\varepsilon_t = \frac{1}{\sqrt{n}}\left\{Z_n(\hat{w}) - Z_n(w)\right\},$$

where

$$Z_n(f) := \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \mathbb{I}_t(a) f(t/n)\left(X_1 - \hat{X}_1\right)'\varepsilon_t \quad f \in \mathcal{F},$$

and $\mathcal{F} = \left\{f : [0,1] \mapsto \mathbb{R} | \sup_{0 \leq \tau \leq 1} |f(\tau)| \leq F\right\}$ is a compact function space for some fixed bound $F > 0$. From Lemmas B.2, B.4 and B.6, $Z_n(f) \to^d Z(f)$ for any $f \in \mathcal{F}$, where $\{Z(f) : f \in \mathcal{F}\}$ is a Gaussian process. Furthermore, for any $f, g \in \mathcal{F}$,

$$Z_n(f) - Z_n(g) = Z_n(f - g) = S_{X_1 - \hat{X}_1, \varepsilon}^{f-g}.$$

By the same arguments employed in the proofs of Lemmas B.2, B.4 and B.6, one can show that for some constant $C > 0$,

$$E\left[\left\|Z_n(f) - Z_n(g)\right\|^2\right] = E\left[\left\|S_{X_1 - \hat{X}_1, \varepsilon}^{f-g}\right\|^2\right] \leq C \times \sup_{0 \leq \tau \leq 1}|f(\tau) - g(\tau)|^2,$$

which implies that $Z_n(f)$ is stochastically equicontinuous. It now follows that $Z_n(\cdot) \to^d Z(\cdot)$ on $\mathcal{F}$, c.f. Pollard (1990, Theorem 10.2), which in turn implies that

$$Z_n(\hat{w}) - Z_n(w) = \{Z_n(\hat{w}) - Z(\hat{w})\} - \{Z_n(w) - Z(w)\} + \{Z(w) - Z(\hat{w})\} = o_P(1).$$

This shows that $\hat{A}_2 - A_2 = o_P(1/\sqrt{n})$. By similar arguments, it can be shown that $\hat{A}_3 - A_3 = o_P(1/\sqrt{n})$. ■

**Proof of Theorem 3.3.** Let $\bar{F}_n$ denote the test statistic with $w_t$ known, and define for any sequence $\beta = \{\beta_t\}$ the corresponding SSR,

$$SSR^w(\beta) = \sum_{t=1}^{n} \mathbb{I}_t(a) w_t \left(y_t - \beta_t' X_t\right)^2. \tag{A.8}$$

We note for future use that the following expansion holds:

$$\begin{aligned} SSR^w(\beta) - SSR^w(\beta_0) &= \frac{\partial SSR^w(\beta_0)}{\partial \beta}(\beta - \beta_0) + \frac{1}{2}(\beta - \beta_0)' \frac{\partial^2 SSR^w(\beta_0)}{\partial \beta \partial \beta'}(\beta - \beta_0) \\ &= -2\sum_{t=1}^{n} \mathbb{I}_t(a) w_t \varepsilon_t X_t'(\beta_t - \beta_{0,t}) + \sum_{t=1}^{n} \mathbb{I}_t(a) w_t (\beta_t - \beta_{0,t})' X_t X_t' (\beta_t - \beta_{0,t}) \end{aligned} \tag{A.9}$$

In particular, $SSR^w(\hat{\beta})/n = \int_0^1 w(\tau)\sigma^2(\tau)\,d\tau + o_P(1)$, c.f. Lemma B.7, such that

$$\bar{F}_n = \frac{n}{2}\frac{SSR^w(\tilde{\beta}) - SSR^w(\hat{\beta})}{SSR^w(\hat{\beta})} = \frac{SSR^w(\tilde{\beta}) - SSR^w(\hat{\beta})}{2\int_0^1 w(\tau)\sigma^2(\tau)\,d\tau + o_P(1)},$$

where

$$\begin{aligned} SSR^w(\tilde{\beta}) - SSR^w(\hat{\beta}) &= \left\{SSR^w(\tilde{\beta}) - SSR^w(\beta_0)\right\} - \left\{SSR^w(\hat{\beta}) - SSR^w(\beta_0)\right\} \\ &= :\Delta SSR_1 - \Delta SSR_2. \end{aligned}$$

Combining the expansion in eq. (A.9) (with $\beta = \hat{\beta}$) and the representation given in eq. (A.1), we obtain

$$\begin{aligned} \Delta SSR_2 &= -\frac{2}{n}\sum_{t=1}^{n}\sum_{u=1}^{n} \mathbb{I}_t(a) w_t \varepsilon_t X_t' \left[\Lambda_t^{-1} K_{t,u} X_u' \varepsilon_u + O_P(h^r)\right] \\ &\quad + \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\sum_{u=1}^{n} \mathbb{I}_t(a) w_t \left[\varepsilon_s X_s' K_{s,t}\Lambda_t^{-1} + O_P(h^r)\right] X_t X_t' \left[\Lambda_t^{-1} K_{t,u} X_u' \varepsilon_u + O_P(h^r)\right] \\ &= :-\Delta SSR_{2,1} + \Delta SSR_{2,2}, \end{aligned}$$

By Lemma C.4,

$$O_P(h^r) \times \sum_{t=1}^{n} \mathbb{I}_t(a) w_t \varepsilon_t X_t' = O_P\left(h^r \sqrt{n}\right) = o_P\left(1/\sqrt{h}\right).$$

Thus, we can ignore this term, and decompose the remaining terms in $\Delta SSR_{2,1}$ into

$$\Delta SSR_{2,1} \simeq \frac{2K(0)}{nh}\sum_{t=1}^{n} \mathbb{I}_t(a) w_t \varepsilon_t^2 X_t' \Lambda_t^{-1} X_t + \frac{2}{n}\sum_{t \neq u} \mathbb{I}_t(a) w_t \varepsilon_t X_t' K_{t,u} \Lambda_t^{-1} X_u \varepsilon_u,$$

where the average in the first term satisfies

$$E\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t\left(a\right)w_t\varepsilon_t^2X_t'\Lambda_t^{-1}X_t\right]=\frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t\left(a\right)w_t\sigma_t^2E\left[X_t'\Lambda_t^{-1}X_t\right]=m\int_0^1 w\left(s\right)\sigma^2\left(s\right)ds+o\left(1\right),$$

and, due to the mixing conditions, for some small $\epsilon>0$,

$$\text{Var}\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t\left(a\right)w_t\varepsilon_t^2X_t'\Lambda_t^{-1}X_t\right]=O\left(\frac{1}{n^{1-\epsilon}}\right).$$

The terms in $\Delta SSR_{2,2}$ involving $O_P\left(h^r\right)$ are again of lower order and can be ignored, while the remaining terms can be written as

$$\begin{aligned}\Delta SSR_{2,2} &\simeq \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\mathbb{I}_t\left(a\right)w_t\varepsilon_s^2X_s'K_{s,t}^2\Lambda_t^{-1}X_tX_t'\Lambda_t^{-1}X_s\\ &\quad+\frac{1}{n^2}\sum_{s\neq u}\sum_{t=1}^{n}\varepsilon_sX_s'K_{s,t}\mathbb{I}_t\left(a\right)w_t\Lambda_t^{-1}X_tX_t'\Lambda_t^{-1}K_{t,u}X_u\varepsilon_u\\ &=:\Delta SSR_{2,21}+\Delta SSR_{2,22}.\end{aligned}\tag{A.10}$$

By Lemma B.8,

$$\Delta SSR_{2,21}\simeq\frac{m\kappa_2}{h}\int_0^1 w\left(s\right)\sigma^2\left(s\right)ds,$$

while $\Delta SSR_{2,22}$ is a third-order U-statistic. We proceed as in Aït-Sahalia et al (2009, Proof of Theorem 1, Claim (a)): By standard symmetrization arguments and Hoeffding's decomposition, we can write

$$\Delta SSR_{2,22}=\frac{1}{n}\sum_{s\neq u}\mathbb{I}_s\left(a\right)w_s\varepsilon_sX_s'\Lambda_s^{-1}\left(K*K\right)_{s,u}X_u\varepsilon_u+\frac{1}{n^2}\sum_{s\neq t\neq u}\Phi_{s,t,u},$$

where $\Phi_{s,t,u}=\varphi_{s,t,u}^*-\varphi_{s,t}^*-\varphi_{s,u}^*-\varphi_{t,u}^*$, $\varphi_{s,t,u}^*=\varphi_{s,t,u}+\varphi_{s,u,t}+\varphi_{t,s,u}+\varphi_{t,u,s}+\varphi_{u,t,s}+\varphi_{s,u,t}$ is a symmetric kernel,

$$\varphi_{s,t,u}=\varepsilon_sX_s'K_{s,t}\mathbb{I}_t\left(a\right)w_t\Lambda_t^{-1}X_tX_t'\Lambda_t^{-1}K_{t,u}X_u\varepsilon_u,$$

and $\varphi_{s,t}^*=\int\varphi_{s,t,u}dF_t$. Here, $F_t$ is the marginal distribution of $\left(X_t^*,\varepsilon_t^*\right)$ which is an independent copy of $\left(X_t,\varepsilon_t\right)$. By verifying the conditions of Gao and King (2004, Lemma C.2), we obtain under the mixing and moment conditions imposed that

$$\frac{1}{n^2}\sum_{s\neq t\neq u}\Phi_{s,t,u}=O_P\left(\frac{1}{\sqrt{nh^5}}\right)=o_P\left(\frac{1}{\sqrt{h}}\right).$$

In total,

$$\begin{aligned}\Delta SSR_2 &\simeq \frac{m}{h}\left[\kappa_2-2K\left(0\right)\right]\int_0^1 w\left(\tau\right)\sigma^2\left(\tau\right)d\tau\\ &\quad+\frac{1}{n}\sum_{t\neq u}\mathbb{I}_s\left(a\right)w_t\varepsilon_tX_t'\Lambda_t^{-1}\left[\left(K*K\right)_{t,u}-2K_{t,u}\right]X_u\varepsilon_u\end{aligned}$$

To analyze $\Delta SSR_1$, first note that by Theorem 3.2, $\tilde{\beta}_1 - \beta_1 = O_P\left(1/\sqrt{n}\right)$ such that we can replace the estimator with $\beta_1$ in $\Delta SSR_1$. Moreover, by the same arguments used to show eq. (A.1),

$$\tilde{\beta}_{2,t} - \beta_{2,t} = \Lambda_{22,t}^{-1}\frac{1}{n}\sum_{s=1}^{n} K_{s,t}X_{2,s}\varepsilon_s + O_P\left(h^r\right),$$

uniformly in $t$, and we obtain

$$\begin{aligned}\Delta SSR_1 &\simeq -\frac{1}{n}\sum_{t=1}^{n}\sum_{u=1}^{n}\mathbb{I}_t\left(a\right)w_t\varepsilon_t X_{2,t}'\Lambda_{22,s}^{-1}X_{2,u}'K_{s,u}\varepsilon_u \\ &\quad +\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\sum_{u=1}^{n}\mathbb{I}_t\left(a\right)w_t\varepsilon_s X_{2,s}'K_{s,t}\Lambda_{22,t}^{-1}X_{2,t}X_{2,t}'\Lambda_{22,t}^{-1}K_{t,u}X_{2,u}'\varepsilon_u \\ &= : -\Delta SSR_{1,1} + \Delta SSR_{1,2}.\end{aligned} \tag{A.11}$$

The two terms $\Delta SSR_{1,1}$ and $\Delta SSR_{2,2}$ are on a similar form as $\Delta SSR_{2,1}$ and $\Delta SSR_{2,2}$, and by the same arguments as before,

$$\begin{aligned}\Delta SSR_1 &\simeq \frac{m_2}{h}\left[\kappa_2 - 2K\left(0\right)\right]\int_0^1 w\left(s\right)\sigma^2\left(s\right)ds \\ &\quad +\frac{1}{n}\sum_{s\neq u}\mathbb{I}_s\left(a\right)w_s\varepsilon_s X_{2,s}'\left[\left(K*K\right)_{s,u} - 2K_{s,u}\right]\Lambda_{22,u}^{-1}X_{2,u}\varepsilon_u.\end{aligned}$$

Combining the expressions of $\Delta SSR_1$ and $\Delta SSR_2$, we now have

$$\bar{F}_n \simeq \frac{m_1\left[K\left(0\right) - \frac{1}{2}\kappa_2\right]}{h} + \frac{n^{-1}\sum_{s\neq t}\phi_{1,n}\left(u_s,u_t\right)}{2\int_0^1 w\left(\tau\right)\sigma^2\left(\tau\right)d\tau},$$

where

$$\phi_{1,n}\left(u_s,u_t\right) := w_t\varepsilon_t\varepsilon_u\left[2K_{t,u} - \left(K*K\right)_{t,u}\right]\bar{X}_t'\bar{\Lambda}_t^{-1}\bar{X}_t, \tag{A.12}$$

with $u_t = \left(t/n, \varepsilon_t, X_t\right)$, and

$$\bar{X}_t := X_{1,t}' - \omega_t X_{2,t}, \quad \Lambda_{11.2,t} := \Lambda_{11,t} - \omega_t\Lambda_{21,t}, \quad \omega_t := \Lambda_{12,t}\Lambda_{22,t}^{-1}.$$

It now follows by Lemma B.9 that

$$\frac{\bar{F}_n - \mu_n^F}{\nu_n^F} \simeq \frac{n^{-1}\sum_{s\neq t}\phi_{1,n}\left(u_s,u_t\right)}{V_{1,n}} \to^d N\left(0,1\right),$$

Finally, we demonstrate that the estimation of $w_t$ does not affect the result. We write

$$\frac{F_n - \mu_n^F}{\sqrt{\nu_n^F}} = \frac{\bar{F}_n - \mu_n^F}{\sqrt{\nu_n^F}} + \frac{F_n - \bar{F}_n}{\sqrt{\nu_n^F}},$$

where

$$F_n - \bar{F}_n = \frac{\tilde{F}_n(\hat{w} - w)}{2\int_0^1 w\left(\tau\right)\sigma^2\left(\tau\right)d\tau + o_P\left(1\right)}, \quad \tilde{F}_n(f) = SSR^f(\tilde{\beta}) - SSR^f(\hat{\beta}),$$

and $SSR^{f}(\beta)$ is the SSR with weighting function $f$. Since the limiting distribution of $\bar{F}_n$ was derived for any given continuous function $w$, we know that for any continuous function $f:[0,1]\mapsto\mathbb{R}$, $\sqrt{h}\left[\tilde{F}_{1,n}(f)-q_n(f)\right]=O_P(1)$, where

$$q_{1,n}(f):=\frac{m}{h}\left[\kappa_2-2K(0)\right]\int_0^1 f(\tau)\,\sigma^2(\tau)\,d\tau.$$

Moreover,

$$\sqrt{h}\left[\tilde{F}_n(f_1)-q_n(f_1)\right]-\sqrt{h}\left[\tilde{F}_n(f_2)-q_n(f_2)\right]=\sqrt{h}\left[\tilde{F}_n(f_1-f_2)-q_n(f_1-f_2)\right],$$

where, by using the same arguments as above

$$E\left[\left|\left[\tilde{F}_n(f_1-f_2)-q_n(f_1-f_2)\right]\right|^2\right]\leq\frac{C}{h}\sup_{\tau\in[0,1]}\left|f_1(\tau)-f_2(\tau)\right|^2.$$

Thus, using that $\sup_\tau|\hat{w}(\tau)-w(\tau)|=o_P\left(\sqrt{h}\right)$,

$$\begin{aligned}\frac{\left|F_n-\bar{F}_n\right|}{\sqrt{\nu_n^F}} &\leq C\sqrt{h}\left|\tilde{F}_n(\hat{w}-w)-q_n(\hat{w}-w)\right|+\sqrt{h}\left|q_n(\hat{w}-w)\right|\\ &\leq \sup_{\|f\|_\infty\leq C\sqrt{h}}\frac{\left|\tilde{F}_n(f)-q_n(f)\right|}{\sqrt{\nu_n^F}}+\sqrt{h}\left|q_n(\hat{w}-w)\right|\\ &= o_P(1).\end{aligned}$$

■

**Proof of Theorem 3.4.** Let $\bar{W}_n$ denote the test statistic with known $\Omega_t$. Write

$$\begin{aligned}\bar{W}_n &= \sum_{t=1}^n \mathbb{I}_t(a)\left(\left\{\tilde{\beta}_1^w-\beta_{1,t}\right\}-\left\{\hat{\beta}_{1,t}-\beta_{1,t}\right\}\right)'\Omega_t\left(\left\{\tilde{\beta}_1^w-\beta_{1,t}\right\}-\left\{\hat{\beta}_{1,t}-\beta_{1,t}\right\}\right)\\ &= \sum_{t=1}^n \mathbb{I}_t(a)\left(\hat{\beta}_{1,t}-\beta_{1,t}\right)'\Omega_t\left(\hat{\beta}_{1,t}-\beta_{1,t}\right)+\sum_{t=1}^n \mathbb{I}_t(a)\left(\tilde{\beta}_1^w-\beta_1\right)'\Omega_t\left(\tilde{\beta}_1^w-\beta_1\right)\\ &\quad -2\sum_{t=1}^n \mathbb{I}_t(a)\left(\tilde{\beta}_1^w-\beta_1\right)'\Omega_t\left(\hat{\beta}_{1,t}-\beta_{1,t}\right)\\ &= :\bar{W}_{1,n}+\bar{W}_{2,n}+\bar{W}_{3,n}.\end{aligned}$$

The second and third term satisfies

$$\bar{W}_{2,n}=\sqrt{n}\left(\tilde{\beta}_1^w-\beta_1\right)'\left\{\frac{1}{n}\sum_{t=1}^n\mathbb{I}_t(a)\,\Omega_t\right\}\sqrt{n}\left(\tilde{\beta}_1^w-\beta_1\right)=O_P(1),$$

$$\left|\bar{W}_{3,n}\right|\leq 2\sqrt{n}\left\|\tilde{\beta}_1^w-\beta_1\right\|'\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathbb{I}_t(a)\,\Omega_t\left(\hat{\beta}_{1,t}-\beta_{1,t}\right)\right\|$$

where,

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{I}_t(a)\,\Omega_t\left(\hat{\beta}_{1,t}-\beta_{1,t}\right) &= \frac{1}{n^{3/2}}\sum_{t=1}^{n}\sum_{u=1}^{n}\mathbb{I}_t(a)\,\Omega_t\left[\Lambda_t^{-1}K_{t,u}X_u'\varepsilon_u+O_P(h^r)\right] \\
&\simeq \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{I}_t(a)\,\Omega_t\Lambda_t^{-1}X_t'\varepsilon_t+O_P\left(\sqrt{n}h^r\right) \\
&= O_P(1)+O_P\left(\sqrt{n}h^r\right).
\end{aligned}
$$

Following the same arguments as in the proof of Theorem 3.3, the first term satisfies

$$
\begin{aligned}
\bar{W}_{1,n} &\simeq \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\sum_{u=1}^{n}\mathbb{I}_t(a)\,\varepsilon_sX_{1,s}'K_{s,t}\Lambda_{11,t}^{-1}\Omega_t\Lambda_{11,t}^{-1}K_{t,u}X_{1,u}\varepsilon_u \\
&\simeq \frac{1}{n^2}\sum_{t=1}^{n}\sum_{u=1}^{n}\mathbb{I}_t(a)\,\varepsilon_u^2X_{1,u}'\Lambda_{11,t}^{-1}\Omega_t\Lambda_{11,t}^{-1}K_{t,u}^2X_{1,u} \\
&\quad+\frac{1}{n^2}\sum_{s\neq u}\varepsilon_sX_{1,s}'\sum_{t=1}^{n}\left\{\mathbb{I}_t(a)\,K_{s,t}\Lambda_{11,t}^{-1}\Omega_t\Lambda_{11,t}^{-1}K_{t,u}\right\}X_{1,u}\varepsilon_u \\
&\simeq \mu_n^W+\frac{1}{n}\sum_{s\neq t}\phi_{2,n}(u_s,u_t),
\end{aligned}
$$

where

$$
\phi_{2,n}(u_s,u_t):=\varepsilon_sX_{1,s}'\Lambda_{11,s}^{-1}\Omega_s\Lambda_{11,s}^{-1}(K*K)_{s,t}X_{1,t}\varepsilon_t. \tag{A.13}
$$

It now follows by Lemma B.9 that

$$
\frac{\bar{W}_n-\mu_n^W}{\sqrt{\nu_n^F}}\simeq\frac{n^{-1}\sum_{s\neq u}\phi_{2,n}(W_s,W_u)}{V_{2,n}}\to^d N(0,1).
$$

One can show that $\bar{W}_n$ and $W_n$ have the same asymptotic distribution by the same arguments as in the proof of Theorem 3.3. ■

# B Lemmas

**Lemma B.1** *$S_{g-\hat{g}}^w=O_P\left(h^{2r}\right)+O_P(\log(n)/(nh))$ for $g=\beta_2'X_2$ and $\xi$.*

**Proof.** First consider $g=\beta_2'X_2$:

$$
\begin{aligned}
S_{g-\hat{g}}^w &= \frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t(a)\,w_t\left[g_t-\hat{g}_t\right]^2 \\
&\leq \frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t(a)\,w_t\left[\bar{\beta}_{2.t}-\beta_{2,t}\right]'X_{2,t}X_{2,t}'\left[\bar{\beta}_{2,t}-\beta_{2,t}\right] \\
&\leq \frac{1}{n}\sum_{t=1}^{n}w_tX_{2,t}X_{2,t}'\times\sup_{a\leq\tau\leq1-a}\left|\bar{\beta}_2(\tau)-\beta_2(t)\right|^2 \\
&= O_P\left(h^{2r}\right)+O_P(\log(n)/(nh)),
\end{aligned}
$$

where we have used Lemma C.1. The proof for $g = \xi = M(X_1)' X_2$ is similar. ■

**Lemma B.2** *$\sqrt{n} S^w_{g-\hat{g},e} = O_P\left(h^{2r}\sqrt{n}\right) + O_P\left(\log(n)/(\sqrt{n}h)\right)$ for $g = \beta_2' X_2, \xi$ and $e = \varepsilon, V$.*

**Proof.** We only give a proof for the case where $g = \beta_2' X_2$ and $e = \varepsilon$. Define $m_t = \Lambda_{22,t}\beta_{2,t}$ and

$$\hat{m}_t = \sum_{s=1}^n K_{hn}(s-t) X_{2,s} X_{2,s}' \beta_{2,s}, \quad \hat{\Lambda}_{22,t} = \sum_{s=1}^n K_{hn}(s-t) X_{2,s} X_{2,s}',$$

such that $\bar{\beta}_{2,t}$ given in eq. (A.7) can be written as $\bar{\beta}_{2,t} = \hat{\Lambda}_{22,t}^{-1}\hat{m}_t$. Thus,

$$S^w_{g-\hat{g},e} = \frac{1}{n}\sum_{t=1}^n \mathbb{I}_t(a)\left[\bar{\beta}_{2,t} - \beta_{2,t}\right]' w_t X_{2,t}\varepsilon_t = \frac{1}{n}\sum_{t=1}^n \mathbb{I}_t(a)\left[\hat{\Lambda}_{22,t}^{-1}\hat{m}_t - \Lambda_{22,t}^{-1} m_t\right]' w_t X_{2,t}\varepsilon_t.$$

Use the following identity,

$$\frac{a}{b} - \frac{a_0}{b_0} = b_0^{-1}(a - a_0) - b_0^{-2} a_0 (b - b_0) - (b_0 b)^{-1}(b - b_0)\left[(a - a_0) - b_0^{-1} a_0 (b - b_0)\right],$$

to write

$$\begin{aligned}
S^w_{g-\hat{g},\varepsilon} &= \frac{1}{n}\sum_{t=1}^n \mathbb{I}_t(a)\left[\hat{\Lambda}_{22,t}^{-1}\hat{m}_t - \Lambda_{22,t}^{-1} m_t\right]' w_t X_{2,t}\varepsilon_t \\
&= \frac{1}{n}\sum_{t=1}^n \mathbb{I}_t(a)\left[\hat{m}_t - m_t\right]' w_t \Lambda_{22,t}^{-1} X_{2,t}\varepsilon_t - \frac{1}{n}\sum_{t=1}^n \mathbb{I}_t(a)\left[\hat{\Lambda}_{22,t} - \Lambda_{22,t}\right]\beta_{2,t}' w_t \Lambda_{22,t}^{-1} X_{2,t}\varepsilon_t \\
&\quad - \frac{1}{n}\sum_{t=1}^n \mathbb{I}_t(a)\left[(\hat{m}_t - m_t) - \beta_{2,t}\left(\hat{\Lambda}_{22,t} - \Lambda_{22,t}\right)\right]'\left(\hat{\Lambda}_{22,t} - \Lambda_{22,t}\right)\left(\Lambda_{22,t}\hat{\Lambda}_{22,t}\right)^{-1} w_t X_{2,t}\varepsilon_t \\
&=: A_{n,1} - A_{n,2} - A_{n,3}.
\end{aligned}$$

We now prove that for any $\epsilon > 0$:

$$\begin{aligned}
\text{Claim B.2.1} &: \quad A_{n,1} = O_P\left(n^{-1+\epsilon/2}\right) + O_P\left(n^{-1/2}h^r\right), \\
\text{Claim B.2.2} &: \quad A_{n,2} = O_P\left(n^{-1+\epsilon/2}\right) + O_P\left(n^{-1/2}h^r\right), \\
\text{Claim B.2.3} &: \quad A_{n,3} = O_P\left(h^{2r}\right) + O_P\left(\log(n)/(nh)\right).
\end{aligned}$$

This will complete our proof.

**Proof of Claim B.2.1:** Define $u_t = (\tau_t, X_{2,t}, \varepsilon_t)$, and

$$b(u_s, u_t) = \Lambda_{22,s}^{-1}\left[K_h(\tau_s - \tau_t) X_{2,s} X_{2,s}' \beta_{2,s} - m_t\right] w_t X_{2,t}\varepsilon_t,$$

where $\tau_t = t/n$, $t = 1, ..., n$. We can treat $\tau_t$ as i.i.d. uniformly distributed random variables drawn independently of all other random variables. We then see that $A_{n,1}$ can be written as a $V$-statistic, $A_{n,1} = \sum_{s=1}^n \sum_{t=1}^n b(u_s, u_t)/n^2$. Defining

$$\phi(u_s, u_t) := b(u_s, u_t) + b(u_t, u_s),$$

we can write $A_{n,1} = (n-1)/n \times U_n + \sum_{t=1}^{n} \phi(u_t, u_t)/n^2$, where $U_n = \sum_{s<t} \phi(u_s, u_t)/[n(n-1)]$ while $\sum_{t=1}^{n} \phi(u_t, u_t)/n^2 = O_P(n^{-1})$. We then obtain from the Hoeffding decomposition that $U_n = 2\sum_{t=1}^{n} \bar{\phi}(u_t)/n + R_n$, where $R_n$ is a remainder while

$$\bar{\phi}(u) = E[\phi(u, u_t)] = E[b(u, u_t)] + E[b(u_t, u)] = \Lambda_{22}^{-1}(\tau)[\bar{m}(\tau) - m(\tau)]' w(\tau) x_2 e,$$

with $u = (\tau, x_2, e)$, and

$$\bar{m}(\tau) = \int_0^1 K_h(\tau - s) \Lambda_{22}(s) \beta_2(s) ds + O\left(1/\sqrt{n}\right) = m(\tau) + O(h^\nu) + O\left(1/\sqrt{n}\right).$$

Here, we have used that $E[X_{2,t}\varepsilon_t] = 0$. By Lemma C.2, $R_n$ satisfies $R_n = O_P\left(n^{-1+\epsilon/2} s_{n,\delta}\right)$, where $s_{n,\delta} := \sup_{s,t} E\left[|\phi(u_s, u_t)|^{2+\delta}\right]^{1/(2+\delta)}$. Thus,

$$\begin{aligned} s_{n,\delta} &= \sup_{s,t} E\left[|\phi(\tau_s, X_{2,s}, \varepsilon_s, \tau_t, X_{2,t}, \varepsilon_t)|^{2+\delta}\right]^{1/(2+\delta)} \\ &= \sup_{s,t} E\left[E\left[|\phi(\tau_s, X_{2,s}, \varepsilon_s, \tau_t, X_{2,t}, \varepsilon_t)|^{2+\delta} |\tau_s, \tau_t\right]\right]^{1/(2+\delta)} \\ &= \sup_{s,t} \left\{\int_0^1 \int_0^1 E\left[|\phi(v, X_{2,s}, \varepsilon_s, w, X_{2,t}, \varepsilon_t)|^{2+\delta}\right] dv dw\right\}^{1/(2+\delta)}, \end{aligned}$$

where

$$\begin{aligned} & E\left[|\phi(v, X_{2,s}, \varepsilon_s, w, X_{2,t}, \varepsilon_t)|^{2+\delta}\right] \\ \leq\ & 2\left\|\Lambda_{22}^{-1}(v)\right\|^{2+\delta} w^{2+\delta}(v) E\left[\left(|K_h(v-w)| \left\|X_{2,s}X_{2,s}'\beta_2(v)\right\| + \|m(w)\|\right)^{2+\delta} \|X_{2,t}\|^{2+\delta} |\varepsilon_t|^{2+\delta}\right] \\ \leq\ & C\left\|\Lambda_{22}^{-1}(v)\right\|^{2+\delta} w^{2+\delta}(v) \left(|K_h(v-w)|^{2+\delta} E\left[\|X_{2,s}\|^{8+8\delta}\right]^{1/(4+4\delta)} \|\beta_2(v)\|^{2+\delta} + \|m(w)\|^{2+\delta}\right) \\ & \times E\left[\|X_{2,t}\|^{8+8\delta}\right]^{1/(4+4\delta)} E\left[|\varepsilon_t|^{8+8\delta}\right]^{1/(4+4\delta)} \\ \leq\ & C|K_h(v-w)|^{2+\delta}. \end{aligned}$$

Thus,

$$s_{n,\delta} \leq C\left\{\int_0^1 \int_0^1 |K_h(v-w)|^{2+\delta} dv dw\right\}^{1/(2+\delta)} = O\left(h^{-(1+\delta)/(2+\delta)}\right).$$

The other term in the Hoeffding decomposition, $\sum_{t=1}^{n} \bar{\phi}(u_t)/n$, satisfies

$$E\left[\left(\frac{1}{n}\sum_{t=1}^{n} \bar{\phi}(u_t)\right)^2\right] = B_{n,1} + B_{n,2},$$

where

$$\begin{aligned}
\|B_{n,1}\| &= n^{-1}E\left[\left\|\bar{\phi}\left(u_t\right)\right\|^2\right] \\
&\leq n^{-1}2\int_a^{1-a}\left\|\Lambda_{22}^{-1}\left(\tau\right)\right\|^2\left\|\Lambda_{22}\left(\tau\right)\right\|\left\|\bar{m}\left(\tau\right)-m\left(\tau\right)\right\|^2 w^2\left(\tau\right)\sigma^2\left(\tau\right)d\tau \\
&\leq n^{-1}2\int_0^1\left\|\Lambda_{22,t}^{-1}\right\|^2\left\|\Lambda_{22}\left(\tau\right)\right\|w^2\left(\tau\right)\sigma^2\left(\tau\right)d\tau\times\sup_{a\leq\tau\leq 1-a}\left\|\bar{m}\left(\tau\right)-m\left(\tau\right)\right\|^2 \\
&\leq Cn^{-1}h^{2r},
\end{aligned}$$

and, by Lemma C.3 together with $E\left[\bar{\phi}\left(u_t\right)\right]=0$,

$$\begin{aligned}
\|B_{n,2}\| &\leq n^{-2}\sum_{s=1}^{n}\sum_{t\neq s}\left\|E\left[\bar{\phi}\left(u_s\right)'\bar{\phi}\left(u_t\right)\right]\right\| \\
&= n^{-2}\sum_{s=1}^{n}\sum_{t\neq s}\left|E\left[\bar{\phi}\left(u_s\right)'\bar{\phi}\left(u_t\right)\right]-E\left[\bar{\phi}\left(u_s\right)\right]'E\left[\bar{\phi}\left(u_t\right)\right]\right| \\
&= 4M_n^{1/(1+\delta)}\times n^{-2}\sum_{s=1}^{n}\sum_{t\neq s}\beta\left(\left|t-s\right|\right)^{\delta/(1+\delta)},
\end{aligned}$$

where $M_n=\max\left\{M_{n,1},M_{n,2}\right\}$ with

$$M_{n,1}=E\left[\left|\bar{\phi}\left(u_s\right)\right|^{1+\delta}\right]E\left[\left|\bar{\phi}\left(u_t\right)\right|^{1+\delta}\right]\leq E\left[\left|\bar{\phi}\left(u_t\right)\right|^2\right]^2=O\left(h^{2r}\right)$$

and

$$M_{n,2}=E\left[\left|\bar{\phi}\left(u_s\right)\right|^{1+\delta}\left|\bar{\phi}\left(u_t\right)\right|^{1+\delta}\right]\leq E\left[\left|\bar{\phi}\left(u_s\right)\right|^2\right]^{(1+\delta)/2}E\left[\left|\bar{\phi}\left(u_s\right)\right|^2\right]^{(1+\delta)/2}=O\left(h^{2r}\right),$$

while

$$n^{-2}\sum_{s=1}^{n}\sum_{t\neq s}\beta\left(\left|t-s\right|\right)^{\delta/(1+\delta)}=A\times n^{-2}\sum_{s=1}^{n}\sum_{t\neq s}\left|t-s\right|^{-\beta\delta/(1+\delta)}=O\left(n^{-1}\right).$$

**Proof of Claim B.2.2:** We proceed as in Claim B.2.1: Define $u_t=(t/n,X_{2,t},z_t)$, and

$$\begin{aligned}
b\left(u_s,u_t\right) &:= \left[K_h\left(s/n-t/n\right)X_{2,s}X_{2,s}'-\Lambda_t\right]\beta_{2,t}'\Lambda_{22,t}^{-1}X_{2,t}w_t\sigma_t z_t, \\
\bar{\phi}\left(u_t\right) &= E\left[b\left(u_s,u\right)\right]_{u=u_t}=\left[\bar{\Lambda}_t-\Lambda_t\right]\beta_{2,t}'\Lambda_{22,t}^{-1}X_{2,t}w_t\sigma_t z_t
\end{aligned}$$

where $\bar{\Lambda}_{22,t}=\bar{\Lambda}_{22}\left(t/n\right)$ and

$$\bar{\Lambda}_{22}\left(\tau_0\right)=\int_0^1 K_h\left(\tau-\tau_0\right)\Lambda_{22}\left(\tau\right)d\tau=\Lambda_{22}\left(\tau_0\right)+O\left(h^r\right).$$

Then, $A_{n,2}$ can be written as a $U$-statistic and Hoeffding's decomposition applies:

$$A_{n,2} = \frac{1}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} b(u_s, u_t) = \frac{2}{n} \sum_{t=1}^{n} \bar{\phi}(u_t) + R_n.$$

By following the same arguements as in the proof of Claim B.2.1, it can now be shown that $A_{n,2}$ has the same rate as $A_{n,1}$.

**Proof of Claim B.2.3:** We have

$$|A_{n,3}| \leq \left\{ \max_t \mathbb{I}_t(a) \left\| \hat{\Lambda}_{22,t} - \Lambda_{22,t} \right\| \left( \|\hat{m}_t - m_t\| + \left\| \bar{\Lambda}_t - \Lambda_t \right\| \right) \right\} \times \frac{1}{n} \sum_{t=1}^{n} \left[ \left\| \Lambda_{22,t}^{-2} \right\| + o_P(1) \right] \left\| \beta_{2,t} \right\| w_t |\varepsilon_t|,$$

where $\sum_{t=1}^{n} \left[ \left\| \Lambda_{22,t}^{-2} \right\| + o_P(1) \right] \left\| \beta_{2,t} \right\| w_t |\varepsilon_t| / n = O_P(1)$, while by Lemma C.1,

$$\max_t \mathbb{I}_t(a) \left\| \hat{\Lambda}_{22,t} - \Lambda_{22,t} \right\| \left( \|\hat{m}_t - m_t\| + \left\| \bar{\Lambda}_{22,t} - \Lambda_{22,t} \right\| \right) = O_P\left(h^{2r}\right) + O_P\left(\log(n) / (nh)\right).$$

■

**Lemma B.3** $S^w_{\hat{\varepsilon}} = O_P\left(h^{2r}\right) + O_P\left(\log(n) / (nh)\right)$, $S^w_{\hat{V}} = O_P\left(h^{2r}\right) + O_P\left(\log(n) / (nh)\right)$, $S^w_{\hat{\varepsilon},\hat{V}} = O_P\left(h^{2\nu}\right) + O_P\left(\log(n) / (nh)\right)$.

**Proof.** Since $\max_t \mathbb{I}_t(a) \left\| \hat{\Lambda}_{22,t} - \Lambda_{22,t} \right\| = o_P(1)$ and the minimum eigenvalue of $\Lambda_{22,t}$, $\lambda_{\min}(t/n)$, is bounded away from zero,

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_t(a) w_t \left\| X'_{2,t} \hat{\Lambda}_{22,t}^{-1} \right\|^2 \overset{\text{w.p.a.1}}{\leq} \sup_{0 \leq \tau \leq 1} w(\tau) \frac{\left[\inf_\tau \lambda_{\min}(\tau)/2 + o(1)\right]^{-2}}{n} \sum_{t=1}^{n} \|X_{2,t}\|^2 = O_P(1).$$

Thus,

$$\begin{aligned}
\|S^w_{\hat{\varepsilon}}\| &\leq \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_t(a) w_t \hat{\varepsilon}_t^2 \\
&= \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_t(a) w_t \left| \bar{\varepsilon}_t \hat{\Lambda}_{22,t}^{-1} X_{2,t} \right|^2 \\
&\leq \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_t(a) w_t \left\| X'_{2,t} \hat{\Lambda}_{22,t}^{-1} \right\|^2 \times \sup_{a \leq \tau \leq 1-a} \|\bar{\varepsilon}(\tau)\|^2 \\
&\leq O_P(1) \times \sup_{a \leq \tau \leq 1-a} \|\bar{\varepsilon}(\tau)\|^2,
\end{aligned}$$

where $\bar{\varepsilon}(\tau) = n^{-1} \sum_{s=1}^{n} K_h(s/n - \tau) X_{2,s} \varepsilon_s \in \mathbb{R}^{m_2}$ satisfies

$$\sup_{a \leq \tau \leq 1-a} \|\bar{\varepsilon}(\tau)\| = O_P(h^r) + O_P\left(\log(n) / \sqrt{nh}\right),$$

c.f. Lemma C.1.

Similarly, we can show that $S^w_{\hat{V}} = O_P\left(h^{2r}\right) + O_P\left(\log(n)/(nh)\right)$. The cross-term $S^w_{\hat{\varepsilon},\hat{V}}$ satisfies, by Cauchy-Schwarz's inequality,

$$\left\|S^w_{\hat{\varepsilon},\hat{V}}\right\| \leq \sqrt{\left\|S^w_{\hat{\varepsilon}}\right\| \times \left\|S^w_{\hat{V}}\right\|} = O_P\left(h^{2r}\right) + O_P\left(\log(n)/(nh)\right).$$

■

**Lemma B.4** $\sqrt{n}S^w_{e,\hat{e}} = O_P\left(\sqrt{n}h^{2r}\right) + O_P\left(\log(n)/(\sqrt{n}h)\right)$*, for* $e = \varepsilon, V$*.*

**Proof.** We will only give a proof for $S^w_{\varepsilon,\hat{V}}$; one can follow the same strategy to show the claimed results for $S^w_{V,\hat{\varepsilon}}$, $S^w_{V,\hat{V}}$ and $S^w_{\varepsilon,\hat{\varepsilon}}$. With $\bar{V}_t = 1/n\sum_{s=1}^n K_h\left((s-t)/n\right)X_{2,s}V_s$,

$$\begin{aligned}
S^w_{U,\hat{V}} &= \frac{1}{n}\sum_{t=1}^n \mathbb{I}_t(a)\,w_t\varepsilon_t\bar{V}_t'\hat{\Lambda}^{-1}_{22,t}X_{2,t} \\
&= \frac{1}{n}\sum_{t=1}^n \mathbb{I}_t(a)\,w_t\varepsilon_t\bar{V}_t'\Lambda^{-1}_{22,t}X_{2,t} + \frac{1}{n}\sum_{t=1}^n \mathbb{I}_t(a)\,w_t\varepsilon_t\bar{V}_t'\left(\Lambda_t\hat{\Lambda}_t\right)^{-1}\left(\Lambda_{22,t} - \hat{\Lambda}_{22,t}\right)X_{2,t} \\
&=: A_{n,1} + A_{n,2}.
\end{aligned}$$

We claim that for any $\epsilon > 0$:

$$\begin{aligned}
\text{Claim B.4.1} &: \quad A_{n,1} = O_P\left(n^{-1+\epsilon/2}\right) + O_P\left(n^{-1/2}h^r\right), \\
\text{Claim B.4.2} &: \quad A_{n,2} = O_P\left(h^{2r}\right) + O_P\left(\log(n)/(nh)\right).
\end{aligned}$$

**Proof of Claim B.4.1:** $A_{n,1}$ takes the form of a $U$-statistic: Define $u_t = (\tau_t, X_{2,t}, \varepsilon_t)$, $\tau_t = t/n$, and

$$b(u_s, u_t) = \mathbb{I}_t(a)\,w_t\varepsilon_t K_h(s/n - t/n)\,V_s X_{2,s}'\Lambda^{-1}_{22,t}X_{2,t},$$

and note that $E\left[b(u_s,u)\right] = E\left[b(u,u_t)\right] = 0$. Then, by the same arguments as in the proof of Lemma B.2,

$$A_{n,1} = \frac{1}{n(n-1)}\sum_{s<t} b(u_s,u_t) + O_P(1/n) = R_n + O_P(1/n),$$

where $R_n = O_P\left(n^{-1+\epsilon/2}s_{n,\delta}\right)$ with $s_{n,\delta} \equiv \sup_{s,t} E\left[\left|b(W_s,W_t)\right|^{2+\delta}\right]^{1/(2+\delta)}$. Here,

$$\begin{aligned}
&E\left[\left|b(u_s,u_t)\right|^{2+\delta}|\tau_s,\tau_t\right]^{1/(2+\delta)} \\
\leq\ & E\left[w_t\left|U_t\right|^{2+\delta}\left|K_h(\tau_s-\tau_t)\right|^{2+\delta}\left|V_s\right|^{2+\delta}\left\|X_{2,s}'\Lambda^{-1}_{22,t}X_{2,t}\right\|^{2+\delta}\right]^{1/(2+\delta)} \\
\leq\ & \bar{w}\left[\inf_\tau \lambda_{\min}(\tau)\right]^{-2-\delta} \times E\left[\left|K_h(s/n-t/n)\right|^{2+\delta}\left|V_s\right|^{2+\delta}\left|\varepsilon_t\right|^{2+\delta}\left\|X_{2,t}\right\|^{4+2\delta}\right]^{1/(2+\delta)} \\
\leq\ & \bar{w}\left[\inf_\tau \lambda_{\min}(\tau)\right]^{-2-\delta} \times E\left[\left|K_h(s/n-t/n)\right|^{4+2\delta}\left|V_s\right|^{4+2\delta}\right]^{1/(4+4\delta)} \times E\left[\left|\varepsilon_t\right|^{4+2\delta}\left\|X_{2,t}\right\|^{8+4\delta}\right]^{1/(4+4\delta)} \\
\leq\ & C\left|K_h\left((s-t)/n\right)\right|
\end{aligned}$$

where $\bar{w} = \sup_\tau w(\tau)$. Thus, $s_{n,\delta} = O(1)$.

**Proof of Claim B.4.2:**

$$|A_{n,2}| \leq \frac{1}{n}\sum_{t=1}^{n} w_t |\varepsilon_t| \left\|\bar{V}_t\right\| \left\| \left(\Lambda_{22,t}\hat{\Lambda}_{22,t}\right)^{-1} \right\| \left\|\Lambda_{22,t} - \hat{\Lambda}_{22,t}\right\| \left\|X_{2,t}\right\|$$

where $n^{-1}\sum_{t=1}^{n} w_t |\varepsilon_t| \|X_{2,t}\| = O_P(1)$, $\max_t \left\| \left(\Lambda_{22,t}\hat{\Lambda}_{22,t}\right)^{-1} \right\| = O_P(1)$ and, by Lemma C.1,

$$\begin{aligned}\max_t \left\|\Lambda_{22,t} - \hat{\Lambda}_{22,t}\right\| &= O_P(h^r) + O_P\left(\sqrt{\log(n)/nh}\right), \\ \max_t \left\|\bar{V}_t\right\| &= O_P(h^r) + O_P\left(\log(n)/\sqrt{nh}\right).\end{aligned}$$

■

**Lemma B.5** $S^w_{g-\hat{g},\hat{U}} = o_P\left(n^{-1/2}\right)$ *for* $g = \beta_2' X_2, \xi$ *and* $U = \varepsilon, V$.

**Proof.** We have $\left\|S^w_{g-\hat{g},\hat{U}}\right\| \leq \left\|S^w_{g-\hat{g}}\right\|^{1/2} \times \left\|S^w_{\hat{U}}\right\|^{1/2}$. The result now follows from Lemmas B.1 and B.3. ■

**Lemma B.6** $S^w_V \to^P \Sigma_w$ *and* $\sqrt{n}S^w_{V,\varepsilon} \to^d N(0,\Phi_w)$.

**Proof.** We have

$$S^w_V - \Sigma_w = \frac{1}{n}\sum_{t=1}^{n} \mathbb{I}_t(a) w_t \left[V_t V_t' - \Lambda_{22|1,t}\right] + \frac{1}{n}\sum_{t=1}^{n} \left[\mathbb{I}_t(a) - 1\right] w_t \Lambda_{22|1,t},$$

where the first term is $o_P(1)$ according to the Law of Large Numbers of heterogeneous, mixing sequences (see e.g. Wooldridge and White, 1988) since

$$\lim_{n\to\infty} \sup_{1\leq t\leq n} E\left[\left\|\mathbb{I}_t(a) w_t V_t V_t'\right\|^2\right] \leq \lim_{n\to\infty} \sup_{1\leq t\leq n} \left[w(t/n) \left\|\Lambda_{22|1}(t/n)\right\|^2 + o(1)\right] < \infty.$$

The second term satisfies

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} \left|\mathbb{I}_t(a) - 1\right| w_t \left\|\Lambda_{22|1,t}\right\| &= \int_0^a w(s) \left\|\Lambda_{22|1}(s)\right\| ds + \int_{1-a}^1 w(s) \left\|\Lambda_{22|1}(s)\right\| ds + o(1/n) \\ &= O(a).\end{aligned}$$

We show the weak convergence result by applying the CLT of Lemma C.4: First,

$$\begin{aligned}\frac{1}{n}E\left[\sum_{t=1}^{n} \mathbb{I}_t(a) w_t^2 \varepsilon_t^2 V_t V_t'\right] &= \frac{1}{n}\sum_{t=1}^{n} \mathbb{I}_t(a) w_t^2 \sigma_t^2 \Lambda_{22|1,t} + o(1) \\ &= \frac{1}{n}\sum_{t=1}^{n} w_t^2 \sigma_t^2 \Lambda_{22|1,t} + \frac{1}{n}\sum_{t=1}^{n} \left[\mathbb{I}_t(a) - 1\right] w_t^2 \sigma_t^2 \Lambda_{22|1,t} \\ &= \Phi_w + o(1) + O(a).\end{aligned}$$

Thus, as $a \to 0$,

$$\sqrt{n} S^{w}_{V,\varepsilon} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbb{I}_t(a) w_t V_t \varepsilon_t' \to^d N(0, \Phi_w).$$

■

**Lemma B.7** *With $SSR^w(\beta)$ defined in eq. (2.13): $SSR^w(\hat{\beta})/n = \int_0^1 w(s)\sigma^2(s)\,ds + o_P(1)$ as $a, h \to 0$ and $\log(h)/(nh) \to \infty$.*

**Proof.** By the expansion in eq. (A.9),

$$\begin{aligned} \frac{1}{n} SSR^w(\hat{\beta}) - \frac{1}{n} SSR^w(\beta_0) &= -\frac{2}{n} \sum_{t=1}^{n} \mathbb{I}_t(a) w_t \left(\hat{\beta}_t - \beta_t\right)' X_t \varepsilon_t \\ &\quad + \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_t(a) w_t \left(\beta_t - \hat{\beta}_t\right)' X_t X_t' \left(\beta_t - \hat{\beta}_t\right) \\ &= : A_1 + A_2. \end{aligned}$$

Here,

$$E\left[\frac{1}{n} SSR^w(\beta_0)\right] = \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_t(a) w_t \sigma_t^2 \simeq \int_0^1 w(s)\sigma^2(s)\,ds - 2a,$$

and, using Lemma C.3,

$$\begin{aligned} \text{Var}\left(\frac{1}{n} SSR^w(\beta_0)\right) &\leq \frac{1}{n^2} \sum_{s,t=1}^{n} w_t^2 \sigma_t^4 \text{Cov}\left(z_s^2, z_t^2\right) \\ &= \frac{1}{n^2} \sum_{s,t=1}^{n} w_t^2 \sigma_t^4 E\left[\left(z_s^2 - 1\right)\left(z_t^2 - 1\right)\right] \\ &= \frac{1}{n} E\left[\left(z_t^2 - 1\right)^2\right] \times \frac{1}{n} \sum_{s,t=1}^{n} w_t^2 \sigma_t^4 + \frac{4C}{n^2} \sum_{s,t=1}^{n} b_n\left(|s-t|\right)^{\delta/(1+\delta)} \\ &= O(1/n). \end{aligned}$$

The two remainder terms satisfy

$$\|A_2\| \leq \left\{\frac{1}{n} \sum_{t=1}^{n} w_t X_t X_t'\right\} \times \sup_{a \leq \tau \leq 1-a} \left\|\beta(\tau) - \hat{\beta}(\tau)\right\|^2 = o_P(1),$$

$$\|A_3\| \leq \frac{2}{n} \sum_{t=1}^{n} w_t \|X_t \varepsilon_t\| \times \sup_{a \leq \tau \leq 1-a} \left\|\beta(\tau) - \hat{\beta}(\tau)\right\| = o_P(1).$$

■

**Lemma B.8** *The term $\Delta SSR_{2,21}$ defined in eq. (A.10) satisfies*

$$\Delta SSR_{2,21} = \frac{m\kappa_2}{h} \times \int w(\tau)\sigma(\tau)\,d\tau \times \left[1 + O_P(a) + O_P(h^r)\right] + O_P\left(\frac{1}{\sqrt{n^{1-\epsilon/2} h^{(2+2\delta)/(2+\delta)}}}\right).$$

**Proof.** We proceed as in the proof of Lemma B.2: Define $u_t = (\tau_t, X_t, \varepsilon_t)$, with $\tau_t = t/n$, $t = 1, ..., n$, and let $u_t^*$ is an i.i.d. copy of $u_t$. We then introduce

$$\phi(u_s, u_t) := b(u_s, u_t) + b(u_t, u_s) - 2\mu,$$

$$b(u_s, u_t) = \mathbb{I}_t(a) w_t \varepsilon_s^2 X_s' K_{s,t}^2 \Lambda_t^{-1} X_t X_t' \Lambda_t^{-1} X_s,$$

and $\mu = E[b(u_t^*, u_s^*)]$. With these definitions, we can write

$$\Delta SSR_{2,21} = \mu + \frac{n-1}{n} U_n + \sum_{t=1}^{n} \phi(u_t, u_t)/n^2,$$

where $U_n = \sum_{s<t} \phi(u_s, u_t)/[n(n-1)]$. The term $\sum_{t=1}^{n} \phi(u_t, u_t)/n^2$ is of lower order, while $U_n = 2\sum_{t=1}^{n} \bar{\phi}(u_t)/n + R_n$, where $R_n$ is a remainder term and

$$\bar{\phi}(u) = E[\phi(u, u_t)] = E[b(u, u_t)] + E[b(u_t, u)] - 2\mu.$$

Here, with $u = (\tau, x, e)$,

$$\begin{aligned} E[b(u, u_t)] &= e^2 x' \left( \frac{1}{h^2} \int_a^{1-a} K^2\left(\frac{\tau - t}{h}\right) w(t) \Lambda^{-1}(t)\, dt \right) x \\ &= e^2 x' \left( \frac{1}{h} \int K^2(z) w(\tau + zh) \Lambda^{-1}(\tau + zh)\, dt + O(a) \right) x \\ &= \frac{\kappa_2}{h} \times w(\tau) e^2 x' \Lambda^{-1}(\tau) x \times [1 + O(a) + O(h^r)], \end{aligned}$$

and

$$E[b(u_t, u)] = \frac{\kappa_2}{h} \times \mathbb{I}_t(a) w(\tau) \sigma^2(\tau) x' \Lambda^{-1}(\tau) x \times [1 + O(a) + O(h^r)].$$

Thus, $\mu \simeq \frac{m\kappa_2}{h} \times \int w(\tau) \sigma(\tau)\, d\tau$. By Lemma C.2, $R_n$ satisfies $R_n = O_P\left(n^{-1+\epsilon/2} s_{n,\delta}\right)$, where $s_{n,\delta} := \sup_{s,t} E\left[|\phi(u_s, u_t)|^{2+\delta}\right]^{1/(2+\delta)}$. This moment can be written as

$$s_{n,\delta} = \sup_{s,t} \left\{ \int_0^1 \int_0^1 E\left[ |\phi(v, X_{2,s}, \varepsilon_s, w, X_{2,t}, \varepsilon_t)|^{2+\delta} \right] dv dw \right\}^{1/(2+\delta)},$$

where

$$E\left[ |\phi(v, X_{2,s}, \varepsilon_s, w, X_{2,t}, \varepsilon_t)|^{2+\delta} \right] \le C |K_h(v - w)|^{4+2\delta}.$$

Thus,

$$s_{n,\delta} \le C \left\{ \int_0^1 \int_0^1 |K_h(v - w)|^{4+2\delta} dv dw \right\}^{1/(2+\delta)} = O\left(h^{-(2+2\delta)/(2+\delta)}\right).$$

Finally, by the same arguments as in the proof of Lemma B.2,

$$E\left[ \left( \frac{1}{n} \sum_{t=1}^{n} \bar{\phi}(u_t) \right)^2 \right] = O\left(n^{-1+\epsilon/2} h^{-(2+2\delta)/(2+\delta)}\right).$$

■

**Lemma B.9** *With $\phi_{1,n}(u_s,u_t)$ and $\phi_{2,n}(u_s,u_t)$ defined in eq. (A.12) and (A.13):*

$$\frac{n^{-1}\sum_{s\neq t}\phi_{i,n}(u_s,u_t)}{\sqrt{V_{i,n}}}\rightarrow^d N(0,1), \quad i=1,2,$$

*where*

$$\begin{aligned} V_{1,n} &= \frac{m_1}{h}\int w^2(\tau)\sigma^4(\tau)\,d\tau\times\left\|K-\tfrac{1}{2}(K*K)\right\|^2, \\ V_{2,n} &= \frac{2}{h}\int\sigma^4(\tau)\,tr\left\{\Omega(\tau)\Lambda_{11}^{-1}(\tau)\Omega(\tau)\Lambda_{11}^{-1}(\tau)\right\}d\tau\times\|K*K\|^2. \end{aligned}$$

**Proof.** We show the two results by verifying in each case the assumptions (A1)-(A3) stated in Fan and Li (1999). We note that Fan and Li (1999) restrict themselves to the case where $W_t$ is stationary and ergodic, but by inspection of their proof one can check that their main result still holds for non-stationary $W_t$ as long as the mixing conditions remain satisfied. In particular, we can appeal to the martingale CLT stated in Lemma C.4. The only difference is that their conditions (A1)-(A3) now has to hold uniformly over $t$ and $n$. This is akin to the generalization of the results of Hansen (2008) to the case of non-stationary sequences as developed in Kristensen (2009).

First, we write

$$n^{-1}\sum_{s\neq t}\phi_{1,n}(u_s,u_t)=\sum_{s<t}H_{1,n}(u_s,u_t), \tag{B.1}$$

where $u_t=(\tau_t,\varepsilon_t,X_t)$ and

$$H_{1,n}(u_s,u_t)=\frac{1}{n}\left[\phi_{1,n}(u_s,u_t)+\phi_{1,n}(u_s,u_t)\right]=\frac{1}{n}\varepsilon_s\varepsilon_t\bar{K}_{s,t}\left[a_{s,t}+a_{t,s}\right].$$

Here, we have for notational convenience introduced $a_{s,t}:=w_s\bar{X}_s'\bar{\Lambda}_s^{-1}\bar{X}_t$, $\bar{K}_{s,t}=2K_{s,t}-(K*K)_{s,t}$. Using the MGD property of $z_t$,

$$E\left[H_{1,n}(u_s,u)\right]=E\left[H_{1,n}(u,u_t)\right]=0.$$

Thus, the kernel is degenerate. Next, we verify that the moments defined on p. 248 in Fan and Li (1999) satisfy their conditions (A1)-(A3).

With $u_t^*$ denoting an i.i.d. copy of $u_t$, observe that

$$E\left[\left(a_{s,t}^*\right)^2\right]=w_s^2\mathrm{tr}\left\{\bar{\Lambda}_s^{-1}E\left[\bar{X}_t^*\left(\bar{X}_t^*\right)'\right]\bar{\Lambda}_s^{-1}E\left[\bar{X}_s^*\left(\bar{X}_s^*\right)'\right]\right\}=w_s^2\mathrm{tr}\left\{\bar{\Lambda}_s^{-1}\bar{\Lambda}_t\right\},$$

and

$$E\left[a_{s,t}^*a_{t,s}^*\right]=w_sw_tE\left[\mathrm{tr}\left\{\bar{\Lambda}_s^{-1}\bar{X}_t^*\left(\bar{X}_t^*\right)'\bar{\Lambda}_s^{-1}\bar{X}_s^*\left(\bar{X}_s^*\right)'\right\}\right]=w_sw_t\mathrm{tr}\left\{\bar{\Lambda}_s^{-1}\bar{\Lambda}_t\right\},$$

such that the variance term $\sigma_n^2$ of the $U$-statistic satisfies

$$\begin{aligned}
\sigma_n^2 \quad : \quad &= \frac{1}{n^2}\sum_{s,t} E\left[H_{1,n}^2\left(u_s^*, u_t^*\right)\right] \\
&= \frac{1}{n^4}\sum_{s,t} E\left[\left(\varepsilon_s^*\right)^2\left(\varepsilon_t^*\right)^2 \bar{K}_{s,t}^2\left(\left(a_{s,t}^*\right)^2+\left(a_{t,s}^*\right)^2+2a_{s,t}^* a_{t,s}^*\right)\right] \\
&\simeq \frac{1}{n^2h^2}\int \sigma^2(v)\,\sigma^2(z)\,\bar{K}^2\left(\frac{v-z}{h}\right)\left[w^2(v)+w^2(z)+2w(v)\,w(z)\right]\operatorname{tr}\left\{\bar{\Lambda}^{-1}(v)\,\bar{\Lambda}(z)\right\} dvdz \\
&\simeq \frac{4m_1}{n^2h}\int \sigma^4(v)\,w^2(v)\,dv \times \int \bar{K}^2(z)\,dz.
\end{aligned}$$

By similar arguments and using that $\sup_t E\left[\|X_t\|^4\right] < \infty$,

$$\begin{aligned}
\mu_{n_4} \quad : \quad &= \frac{1}{n^2}\sum_{s,t} E\left[H_{1,n}^4\left(u_s^*, u_t^*\right)\right] = \frac{1}{n^6}\sum_{s,t} E\left[\left(\varepsilon_s^*\right)^4\left(\varepsilon_t^*\right)^4 \bar{K}_{s,t}^4\left(\left(a_{s,t}^*\right)^2+\left(a_{t,s}^*\right)^2+2a_{s,t}^* a_{t,s}^*\right)^2\right] \\
&= O\left(\frac{1}{n^4h^3}\right)
\end{aligned}$$

Next, with $s_1 \neq t_1$ and $s_2 \neq t_2$,

$$\begin{aligned}
&\left|E\left[H_{1,n}\left(u_{s_1}, u_{t_1}\right)^2 H_{1,n}\left(t_{s_2}, u_{t_2}\right)^2\right]\right| \\
\leq\ & \frac{C}{n^4} \times \sup_{s_1\neq t_1, s_2\neq t_2} E\left[\left|z_{s_1}\right|^4\left|z_{t_1}\right|^4\left|z_{s_2}\right|^4\left|z_{t_2}\right|^4\right]^{1/2} \\
&\times \sup_{s_1\neq t_1, s_2\neq t_2} E\left[\left\|X_{s_1}\right\|^2\left\|X_{s_2}\right\|^2\left\|X_{t_1}\right\|^2\left\|X_{t_2}\right\|^2\right]^{1/2} \\
&\times \int\int\int\int \frac{1}{h^4}\bar{K}^2\left(\frac{v_1-z_1}{h}\right)\bar{K}^2\left(\frac{v_2-z_2}{h}\right)\sigma^2(v_1)\,\sigma^2(z_1)\,\sigma^2(v_2)\,\sigma^2(z_2)\,dv_1dz_1dv_2dz_2 \\
=\ & O\left(\frac{1}{n^4h^2}\right).
\end{aligned}$$

where we have utilized the moment conditions imposed on regressors and errors. Similarly,

$$\left|E\left[H_{1,n}\left(u_{s_1}, u_{t_1}\right) H_{1,n}\left(u_{s_2}, u_{t_2}\right)\right]\right| = O\left(\frac{1}{n^2}\right),$$

$$\left|E\left[H_{1,n}\left(u_{s_1}, u_{t_1}\right) H_{1,n}\left(u_{s_2}, u_{t_2}\right)^3\right]\right| = O\left(\frac{1}{n^4h^2}\right).$$

Using the MGD property of $\varepsilon_t$,

$$E\left[H_{1,n}\left(u_s^*, u_t\right) H_{1,n}\left(u_s^*, u_v\right) | u_s^*\right] = 0,$$

such that $\tilde{\gamma}_{n14} = 0$. Using that $\sup_{t\geq 1} E\left[z_t^4\right] < \infty$ and $\sup_{t\geq 1} E\left[\|X_t\|^4\right] < \infty$,

$$\begin{aligned}
\tilde{\gamma}_{n22} &= E\left[H_{1,n}^2\left(u_s^*, u_t^*\right) H_{1,n}^2\left(u_s^*, u_v^*\right)\right] \\
&\leq \frac{C}{n^7}\sum_{s,t,v} \bar{K}_{s,t}^2 \bar{K}_{s,v}^2 \sigma_s^4 \sigma_t^2 \sigma_v^2 \\
&\simeq \frac{C}{n^4h^4}\int_0^1\int_0^1\int_0^1 \sigma^4(u)\,\sigma^2(w)\,\sigma^2(z)\,\bar{K}^2\left(\frac{u-w}{h}\right)\bar{K}^2\left(\frac{u-z}{h}\right) dudwdz \\
&= O\left(\frac{1}{n^4h^2}\right).
\end{aligned}$$

We conclude that $\gamma_n := \max\{\gamma_{n11}, \tilde{\gamma}_{n22}, \tilde{\gamma}_{n14}\} = O\left(1/\left(n^4h^2\right)\right)$, and $\nu_n := \max\{\gamma_{n22}, \gamma_{n13}\} = O\left(1/\left(n^4h^2\right)\right)$. Thus, with $m \propto \log n$ and $r = n^{1/4}$, (A1) then holds if:

$$\begin{aligned}
\frac{\mu_{n_4} r^2 m}{n^2\sigma_n^4} &= O\left(\frac{\log n}{n^{3/2}h^2}\right) \to 0, \quad \frac{\gamma_n m^2}{\sigma_n^2} = O\left(\frac{\log(n)^2}{n^2 h}\right) \to 0, \\
\frac{\nu_n r m^5}{n^2\sigma_n^4} &= O\left(\frac{\log(n)^5}{n^{3/2}}\right) \to 0.
\end{aligned}$$

To verify (A2), first note that:

$$\begin{aligned}
G(u_t, u_v) &:= E\left[H_{1,n}\left(u_s^*, u_t\right) H_{1,n}\left(u_s^*, u_v\right) | u_t, u_v\right] \\
&= \frac{1}{n^3}\sum_s E\left[\left(\varepsilon_s^*\right)^2\right] \varepsilon_t \varepsilon_v \bar{K}_{s,t}\bar{K}_{s,v} \times \\
&\quad E\left[\left[w_s\left(\bar{X}_s^*\right)\bar{\Lambda}_t^{-1}\bar{X}_t + w_t\bar{X}_t\bar{\Lambda}_s^{-1}\bar{X}_s^*\right]\left[w_s\left(\bar{X}_s^*\right)\bar{\Lambda}_v^{-1}\bar{X}_u + w_u\bar{X}_v\bar{\Lambda}_s^{-1}\bar{X}_s^*\right]\right] \\
&= \frac{1}{n^2}\varepsilon_t\varepsilon_v\frac{1}{h^2}\int \bar{K}\left(\frac{s-t/n}{h}\right)\bar{K}\left(\frac{s-v/n}{h}\right)\sigma^2(s) \times \\
&\quad \left\{w^2(s)\,\bar{X}_t'\bar{\Lambda}_t^{-1}\bar{\Lambda}(s)\,\bar{\Lambda}_u^{-1}\bar{X}_v + w(s)\,tr\left\{\bar{\Lambda}_t^{-1}\bar{X}_t w_u \bar{X}_u\right\}\right. \\
&\quad \left.+w(s)\,w_t\bar{X}_t\bar{\Lambda}_v^{-1}\bar{X}_v + w_t w_v \bar{X}_t'\bar{\Lambda}^{-1}(s)\,\bar{X}_v\right\} ds \\
&= \frac{1}{n^2}\left(\bar{K}*\bar{K}\right)_{t,v}\varepsilon_t\varepsilon_v A_{t,v},
\end{aligned}$$

where

$$\begin{aligned}
A_{t,v} &:= \bar{X}_t'\bar{\Lambda}_t^{-1}\left(\int\sigma^2(s)\,w^2(s)\,\bar{\Lambda}(s)\,ds\right)\bar{\Lambda}_v^{-1}\bar{X}_v + \left(\int\sigma^2(s)\,w(s)\,ds\right)tr\left\{\bar{\Lambda}_t^{-1}\bar{X}_t w_v\bar{X}_v\right\} \\
&\quad + \left(\int\sigma^2(s)\,w(s)\,ds\right)w_t\bar{X}_t\bar{\Lambda}_v^{-1}\bar{X}_v + w_t w_v\bar{X}_t'\left(\int\sigma^2(s)\,\bar{\Lambda}^{-1}(s)\,ds\right)\bar{X}_v.
\end{aligned}$$

Then,

$$\sigma_G^2 := E\left[G^2(u_t, u_t)\right] = \left(\bar{K}*\bar{K}\right)^2(0)\,\frac{1}{n^5h^2}\sum_t \sigma_t^4 E\left[z_t^4\right] E\left[A_{t,t}^2\right] = O\left(\frac{1}{n^4h^2}\right),$$

and, with $s \neq t$,

$$\begin{aligned} E\left[G^2(u_s,u_t)\right] &= \frac{1}{n^6}\sum_{s,t}\left(\bar{K}*\bar{K}\right)^2_{s,t}\sigma_s^2\sigma_t^2 E\left[A^2_{s,t}\right] \\ &\leq \frac{C}{n^4}\frac{1}{h^2}\int\int\left(\bar{K}*\bar{K}\right)\left(\frac{v-w}{h}\right)dvdw \\ &= O\left(\frac{1}{n^4 h}\right) \end{aligned}$$

which shows that $\mu_{nG2} = O\left(1/(n^4h)\right)$. By similar arguments, $\gamma_{nG11} = O\left(1/(n^4h)\right)$. Assumption (A2) then holds if:

$$\begin{aligned} \frac{\mu_{n,G2}m^4}{\sigma_n^4} &= O\left(h\log(n)^4\right)\to 0, \quad \frac{\gamma_{nG11}m^4}{\sigma_n^4} = O\left(h\log(n)^4\right)\to 0, \\ \frac{\sigma_G^2 m}{n\sigma_n^4} &= O\left(\frac{\log(n)}{n}\right)\to 0. \end{aligned}$$

Finally, (A3) is easily shown to hold along the same lines as in Fan and Li (1999, Proof of Theorem 3.1). It now follows by Theorem 2.1 of Fan and Li (1999) that

$$\frac{n^{-1}\sum_{s\neq t}\phi_{1,n}(u_s,u_t)}{n\left(\sigma_n/\sqrt{2}\right)} = \frac{\sqrt{2}\sum_{t<u}H_{1,n}(u_s,u_t)}{n\sigma_n} \to^d N(0,1).$$

To prove the second claim, define

$$H_{2,n}(u_s,u_t) = \frac{1}{n}\left[\phi_{2,n}(u_s,u_t)+\phi_{2,n}(u_s,u_t)\right] = \varepsilon_s\varepsilon_t\bar{K}_{s,t}\left[a_{s,t}+a_{t,s}\right],$$

where $a_{s,t} := X_{1,s}'\Lambda_{11,s}^{-1}\Omega_s\Lambda_{11,s}^{-1}X_{1,t}$ and $\bar{K}_{s,t} = (K*K)_{s,t}$. Thus, the $U$-statistic is on the same form as before, and the assumptions (A1)-(A3) of Fan and Li (1999) can be verified as before. In particular,

$$\begin{aligned} E\left[\left(a^*_{s,t}\right)^2\right] &= \operatorname{tr}\left\{E\left[X^*_{1,s}\left(X^*_{1,s}\right)'\right]\Lambda_{11,s}^{-1}\Omega_s\Lambda_{11,s}^{-1}E\left[X^*_{1,t}\left(X^*_{1,t}\right)'\right]\Lambda_{11,s}^{-1}\Omega_s\Lambda_{11,s}^{-1}\right\} \\ &= \operatorname{tr}\left\{\Omega_s\Lambda_{11,s}^{-1}\Lambda_{11,t}\Lambda_{11,s}^{-1}\Omega_s\Lambda_{11,s}^{-1}\right\} \end{aligned}$$

and

$$E\left[a^*_{s,t}a^*_{t,s}\right] = \operatorname{tr}\left\{\Omega_s\Lambda_{11,s}^{-1}\Lambda_{11,t}\Lambda_{11,t}^{-1}\Omega_t\Lambda_{11,t}^{-1}\right\}.$$

Thus, the variance of the $U$-statistic takes the form

$$\begin{aligned} \sigma_n^2 &= \frac{1}{n^4}\sum_{s,t}E\left[\left(\varepsilon_s^*\right)^2\left(\varepsilon_t^*\right)^2\bar{K}^2_{s,t}\left(\left(a^*_{s,t}\right)^2+\left(a^*_{t,s}\right)^2+2a^*_{s,t}a^*_{t,s}\right)\right] \\ &\simeq \frac{4}{n^2h^2}\int\sigma^2(v)\,\sigma^2(w)\,\bar{K}^2\left(\frac{v-w}{h}\right)\operatorname{tr}\left\{\Omega(v)\,\Lambda_{11}^{-1}(v)\,\Lambda_{11}(w)\,\Lambda_{11}^{-1}(v)\,\Omega(v)\,\Lambda_{11}^{-1}(v)\right\}dvdw \\ &\simeq \frac{4}{n^2h}\int\sigma^4(\tau)\operatorname{tr}\left\{\Omega(\tau)\,\Lambda_{11}^{-1}(\tau)\,\Omega(\tau)\,\Lambda_{11}^{-1}(\tau)\right\}d\tau\times\int\bar{K}^2(z)\,dz, \end{aligned}$$

and the result follows from Fan and Li (1999, Theorem 2.1). ■

# C Auxiliary Lemmas

Let in the following $\{u_{n,t}\}$ be an absolutely regular triangular array with mixing coefficients $\beta_n(t)$ that satisfy $\beta_n(t) \leq Bt^{-\beta}$ for some $B, \beta > 0$.

**Lemma C.1** *Assume that there exists a function $m \in C^r([0,1])$ such that $E[u_{n,t}] = m(t/n) + o(1)$ and that $\sup_{n\geq 1}\sup_{1\leq t\leq n} E[\|u_{n,t}\|^s] < \infty$ for some $s > 2$. Then $\hat{m}(\tau) = \sum_{t=1}^n K_h(t/n - \tau)u_{n,t}$ satisfies*

$$\sup_{a\leq\tau\leq 1-a}|\hat{m}(\tau) - m(\tau)| = O_P(h^r) + O_P\left(\sqrt{\log(n)}/\sqrt{(nh)}\right)$$

*for any sequence $a \to 0$ satisfying $h/a \to 0$.*

**Proof.** We have by Kristensen (2009, Theorem 1) that $\sup_{a\leq\tau\leq 1-a}|\hat{m}(\tau) - E[\hat{m}(\tau)]| = O_P\left(\sqrt{\log(n)}/\sqrt{(nh)}\right)$. That $\sup_{a\leq\tau\leq 1-a}|E[\hat{m}(\tau)] - m(\tau)| = O(h^r)$ is a consequence of the following argument: First, uniformly over $\tau \in [0,1]$,

$$\begin{aligned} E\left[\frac{1}{n}\sum_{t=1}^n K_h(t/n-\tau)u_{n,t}\right] &= \frac{1}{nh}\sum_{t=1}^n K\left(\frac{t/n-\tau}{h}\right)[m(t/n)+o(1)] \\ &= \int_0^1 K_h(s-\tau)m(s)\,ds + O(1/(nh)), \end{aligned}$$

where we have used the mean-value theorem. Next, by change of variable and using standard Taylor expansion arguments, for any $\tau \in [a, 1-a]$

$$\begin{aligned} \int_0^1 K_h(s-\tau)m(s)\,ds &= \sum_{k=0}^{r-1} h^k\left\{\int_{-\tau/h}^{(1-\tau)/h} K(u)u^k du\right\} m^{(k)}(\tau) \\ &\quad + h^r\int_{-\tau/h}^{(1-\tau)/h} K(u)u^r m^{(r)}\left(\tau + u\bar{h}\right)du, \end{aligned}$$

for some $0 \leq \bar{h} \leq h$, where, for $1 \leq k \leq r$,

$$\begin{aligned} \sup_{\tau\in[a,1-a]}\left|\int_{-\tau/h}^{(1-\tau)/h} K(u)u^k du\right| &\leq \sup_{\tau\in[a,1-a]}\left|\int_{-\infty}^{-\tau/h} K(u)u^k du\right| + \sup_{\tau\in[a,1-a]}\left|\int_{(1-\tau)/h}^{+\infty} K(u)u^k du\right| \\ &\leq \int_{-\infty}^{-a/h}|K(u)||u|^k du + \int_{(1-a)/h}^{+\infty}|K(u)||u|^k du \\ &\to 0, \end{aligned}$$

as $a/h \to +\infty$, and similarly $\int_{-\tau/h}^{(1-\tau)/h} K(u)\,du \to 1$ uniformly. In particular,

$$\begin{aligned} &\sup_{\tau\in[a,1-a]}\int_{-\tau/h}^{(1-\tau)/h}|K(u)||u|^r\left\|m^{(r)}\left(\tau+u\bar{h}\right)\right\|du \\ \leq\ &\left\{\sup_{\tau\in[a,1-a]}\int_{-\tau/h}^{(1-\tau)/h}|K(u)||u|^r du\right\}\left\{\sup_{v\in[0,1]}\left\|m^{(r)}(v)\right\|\right\} \\ <\ &\infty. \end{aligned}$$

■

**Lemma C.2** *Assume that* $\beta > (2-\epsilon)(2+\delta)/\delta$ *for some* $\delta, \epsilon > 0$*. Then for any symmetric function* $\phi_n(u_{n,s}, u_{n,t})$*, the following decomposition holds:*

$$\frac{1}{n^2}\sum_{s,t=1}^{n}\phi_n(u_{n,s}, u_{n,t}) = \theta_n + \frac{2}{n}\sum_{t=1}^{n}\left[\bar{\phi}_n(u_{n,t}) - \theta_n\right] + R_n,$$

*where* $\theta_n = \sum_{s<t} E\left[\phi_n(u_{n,s}, u_{n,t})\right]/n^2$, $\bar{\phi}_n(u) = E\left[\phi_n(u, u_{n,t})\right]$*, and the remainder term satisfies* $E\left[R_n^2\right]^{1/2} = O\left(n^{-1+\epsilon/2} \times s_{n,\delta}\right)$ *with* $s_{n,\delta} = \sup_{s\neq t} E\left[\left|\phi_n(u_{n,s}, u_{n,t})\right|^{2+\delta}\right]^{1/(2+\delta)}$.

**Proof.** See Denker and Keller (1983, Proof of Proposition 2) for the proof in the case where $u_{n,t} = u_t$ is stationary and ergodic. By inspection of their proof, one easily sees that their result extends to the non-stationary case as long as the mixing conditions are maintained. ■

**Lemma C.3** *For any function* $\phi$ *with* $E\left[\left|\phi(u_{n,s}, u_{n,t})\right|^{1+\delta}\right] < \infty$*:*

$$\left|E\left[\phi(u_{n,s}, u_{n,t})\right] - E\left[\phi\left(u_{n,s}^*, u_{n,t}^*\right)\right]\right| \leq 4\max\left\{M_{n,1}, M_{n,2}\right\}\left|\beta_n(|s-t|)\right|^{\delta/(1+\delta)},$$

*where* $u_{n,t}^*$ *is an independent sequence with same marginal distribution as* $u_{n,t}$ *and*

$$M_{n,1} = E\left[\left|\phi(u_{n,s}, u_{n,t})\right|^{1+\delta}\right], \quad M_{n,2} = E\left[\left|\phi\left(u_{n,s}^*, u_{n,t}^*\right)\right|^{1+\delta}\right].$$

**Proof.** See Denker and Keller (1983, Proof of Lemma 6) for the proof in the case where $u_{n,t} = u_t$ is stationary and ergodic. Again, it can be checked that the proof still goes through in the non-stationary (but mixing) case. ■

**Lemma C.4** *Assume that* $u_{n,t}$ *is a MGD satisfying* $n^{-1}\sum_{t=1}^{n} E\left[u_{n,t}^2\right] \to \sigma^2 > 0$ *and, for some* $\delta > 0$, $n^{-1-\delta/2}\sum_{t=1}^{n} E\left[\left|u_{n,t}\right|^{2+\delta}\right] \to 0$*. Then* $\sum_{t=1}^{n} u_{n,t}/\sqrt{n} \to^P N\left(0, \sigma^2\right)$.

**Proof.** This is a straightforward implication of McLeish (1974, Theorem 2.3). ■

# Research Papers 2011

2010-75: Laurent A.F. Callot: A Bootstrap Cointegration Rank Test for Panels of VAR Models

2010-76: Peter R. Hansen, Asger Lunde and James M. Nason: The Model Confidence Set

2011-01: Cristina Amado and Timo Teräsvirta: Modelling Volatility by Variance Decomposition

2011-02: Timo Teräsvirta: Nonlinear models for autoregressive conditional heteroskedasticity

2011-03: Roxana Halbleib and Valeri Voev: Forecasting Covariance Matrices: A Mixed Frequency Approach

2011-04: Mark Podolskij and Mathieu Rosenbaum: Testing the local volatility assumption: a statistical approach

2011-05: Michael Sørensen: Prediction-based estimating functions: review and new developments

2011-06: Søren Johansen: An extension of cointegration to fractional autoregressive processes

2011-07: Tom Engsted and Stig V. Møller: Cross-sectional consumption-based asset pricing: The importance of consumption timing and the inclusion of severe crises

2011-08: Tommaso Proietti and Stefano Grassi: Bayesian stochastic model specification search for seasonal and calendar effects

2011-09: Matt P. Dziubinski: Option valuation with the simplified component GARCH model

2011-10: Tim A. Kroencke, Felix Schindler and Andreas Schrimpf: International Diversification Benefits with Foreign Exchange Investment Styles

2011-11: Eduardo Rossi and Paolo Santucci de Magistris: Estimation of long memory in integrated variance

2011-12: Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Generalized Jackknife Estimators of Weighted Average Derivatives

2011-13: Dennis Kristensen: Nonparametric Detection and Estimation of Structural Change