Econometrics Journal (2012), volume **00**, pp. 1–42.
Article No. ectj??????

Nonparametric Detection and Estimation of Structural Change

Dennis Kristensen[†] [‡] [*]

[†] *Department of Economics, UCL, Gower Street, London WC1E 6BT, UK.*
E-mail: `d.kristensen@ucl.ac.uk`

[‡] *Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 7 Ridgmount Street, London WC1*

[*] *Center for Research in Econometric Analysis of Time Series, Aarhus University, Building 1322, Bartholins All 10, DK-*

Received: July 2012

Summary We propose a semi-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 time-varying (nonparametric) and constant (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.

Keywords: *Structural Change, Time Series Regression, Nonparametric, Estimation, Testing, Generalized Likelihood Ratio, Time-varying, Locally stationary.*

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 asset returns (Ang and Kristensen, 2012). 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 misspecified 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, β_t, in the following linear regression model,

$$y_t = \beta_t' X_t + \sigma_t z_t, \tag{1.1}$$

where σ_t is the potentially time-varying volatility. We impose no parametric structure on the time variation in β_t and σ_t and instead estimate both components nonparametrically. This way, the risk of misspecification 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 either generalized F or Wald test statistics. We also show how the proposed methods can be used as guidance in the search for 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 embed a rich class of data generating processes, including random walk type dynamics in the parameters. 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 have good finite sample properties under random walk, smooth transition and structural break specifications. Moreover, the tests have precise size properties and exhibit strong power against all the different alternatives. In fact, the nonparametric tests do not trail far behind parametric structural break tests under correct specification of the time-variation, and clearly dominate when the time-varying coefficients are smoothly changing.

The usefulness of the proposed methodology is demonstrated through two empirical applications: In the first one, we investigate structural instabilities in US productivity and find strong evidence of such. We test two parametric structural breaks models against the nonparametric alternative, and find strong support for a three-breaks model while there is mixed evidence for only one break in the sample. 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 loadings 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 explained by underlying macro factors.

There is a large literature on estimation and testing parametric forms of structural change such as deterministic breaks (Andrews and Ploberger, 1994; Bai and Perron 1998), smooth transition (Lin and Teräsvirta, 1994) and hidden Markov models (Hamilton, 1989; Hansen, 1992; Nyblom, 1989). In an influential paper, Elliott and Müller (2006) show that under regularity conditions, efficient parametric tests for a class of breaking processes are equivalent. This seems to indicate that not much can be gained from introducing new tests. However, they restrict themselves to breaking processes that (after scaling) are asymptotically well-approximated by Brownian motions. While the considered class is broad, it still rules out a number of relevant alternatives. In particular, so-called high-frequency alternatives are excluded. A well-known limitation of parametric tests is that they are not able to detect all such alternative; see e.g. Eubank and LaRiccia (1992) and Fan et al (2001). In contrast, our tests can detect such high-frequency alternatives at an optimal rate. On the other hand, within the class of breaking processes of Elliott and Müller (2006), our test will in general be less powerful compared to parametric tests. Thus, in terms of power, there is a trade-off between our test and existing parametric tests. Finally, even if a given parametric test is able to detect structural changes, it will only deliver a consistent estimator of the process β_t if the parametric model of the time-variation is correctly specified. Thus, parametric procedures will in general deliver inconsistent estimators. An exception is the procedure of Müller and Petalas (2010); they however restrict themselves to moderate time-variations such that the magnitude of the instability decreases as the sample size increases.

A number of nonparametric tests have been proposed that involve integration/summation over the changing parameters; e.g, the CUSUM test of Brown et al. and the test of Chu et al. (1995). These tests suffer from the same problem as parametric tests, namely that they cannot detect high-frequency alternatives very well. Moreover, they do not provide estimates of the breaking process under the alternative and has non-standard asymptotic distributions. Instead, our approach is based on idea originating from Robinson (1989), who propose nonparametric estimators of β_t and σ_t, and extended in Cai (2007). However, these studies focus solely on the estimation of changing parameters, and do not consider estimation and testing of constant parameters.

Our testing approach is related to the work by Chen and Hong (2012), Hidalgo (1995), Gao, Gijbels and van Bellegem (2008) and Juhl and Xiao (2005) who also develop kernel-based tests for stability in regression models. Hidalgo (1995) and Gao et al (2008) focus on structural change in fully nonparametric regressions, and so are more robust compared to our tests since no asumptions are made about how the covariates enter the model. On the other hand, their methods cannot handle models with a large number of regressors due to the well-known curse of dimensionality of fully nonparametric estimators, while our procedure has no such isues. Juhl and Xiao (2005) and Chen and Hong (2012) consider linear regression models, but Juhl and Xiao (2005) focus on the case where only the intercept is time-varying, while Chen and Hong (2012) only develop estimators and tests under the null of *all* regression coefficients being constant. We here extend the results of

Chen and Hong (2012) 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, 2012). Secondly, we allow for heteroskedastic errors and modify estimators and test statistics to handle this. Thirdly, we accommodate 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 (2012).

The remains of the paper are organized as follows: In the next section, we introduce our model and develop the proposed estimators and test statistics. Section 3 contains our theoretical results, while Section 4 gives some extensions. Bandwidth selection and bootstrapping is discussed in Section 5. The results of a simulation study and the empirical applications are presented in Section 6 and 7 respectively. 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 eq. (1.1) with $y_t \in \mathbb{R}$ and $X_t \in \mathbb{R}^m$. The error normalized error term $z_t \in \mathbb{R}$ satisfies

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

As such, $\sigma_t^2 > 0$ represents the conditional variance of y_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 coefficients 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,

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

We throughout maintain the hypothesis that $\beta_{2,t}$ is potentially 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. One popular way of modelling β_t is through deterministic breaks, see e.g. Andrews and Ploberger (1994), Bai (1999) and Bai and Perron (1998). In the simplest case, with one break, the dynamics of the regression coefficients are modelled as $\beta_t = \bar{\beta}_1 \mathbb{I}\{t \leq [\pi n]\} + \bar{\beta}_2 \mathbb{I}\{t > [\pi n]\}$ for some (unknown) $\pi \in (0, 1)$ and $\bar{\beta}_1, \bar{\beta}_2 \in \mathbb{R}^m$. 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 assume deterministic changes, another approach is to model β_t as a stochastic process; see e.g. Hamilton (1989), Hansen (1992) and Nyblom (1989).

We here take an alternative approach and do not impose any such parametric restrictions on the nature of the time-variation. However, some additional restrictions has to be imposed on the type of time-variation in order to make any further progress since, at the current level of generality, we have as many parameters as observations. To obtain nonparametric identification, we impose the following rescaling on the parameters,

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

for some functions $\beta : [0,1] \mapsto \mathbb{R}^m$ and $\sigma^2 : [0,1] \mapsto \mathbb{R}_+$ We here use β and σ^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. Under eq. (2.4), the alternative to H_0 is then given by

$$H_A : \beta_{1,t} = \beta_1\left(t/n\right) \text{ for some non-constant function } \beta_1\left(\cdot\right).$$

The assumption in eq. (2.4) 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 eq. (2.4) is rich enough to include many of the parametric models discussed earlier, including structural break and smooth transition models. Moreover, stochastic specifications of β_t can be approximated by eq. (2.4) by choosing the function $\beta\left(\cdot\right)$ as the corresponding continuous-time equivalent. For example, the (rescaled) 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) 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 H_A. In order to motivate our estimators, suppose that $z_t|X_t \sim N\left(0,1\right)$; 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* log-likelihood at τ 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, $K_h\left(z\right) = K\left(z/h\right)/h$ with $K\left(\cdot\right)$ being a kernel and $h > 0$ a window width. The kernel weights $K_h\left(t/n - \tau\right)$, $t = 1, ..., n$, determine how we use information around the time point τ 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 τ 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 τ,

$$(\hat{\beta}(\tau), \hat{\sigma}^2(\tau)) = \arg \max_{\beta \in \mathbb{R}^m, \sigma^2 > 0} L_n^{\text{local}}\left(\beta, \sigma^2 | \tau\right).$$

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, for example, Fan and Huang (2005). The volatility estimator is akin to the one 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. If in fact $\sigma_t^2 = \sigma^2(t/n, X_t)$, the above estimator $\hat{\sigma}^2(\tau)$ will not be consistent and we should instead smooth the residuals w.r.t. both t/n and X_t.

Next, we consider estimation of the parametric (β_1) and nonparametric ($\beta_{2,t}$) components under H_0. We propose to estimate the time-varying and constant coefficients by partitioned regression akin to the semi-nonparametric estimators of Fan and Huang (2005) and Robinson (1988). 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 $b > 0$. 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 β_1, treat β_1 as known and estimate $\beta_2(\tau)$ by

$$\tilde{\beta}_2(\tau, \beta_1) = \arg \max_{\beta_2 \in \mathbb{R}^{m_2}} L_n^{\text{local}}\left(\beta_1, \beta_2, \sigma^2 | \tau\right),$$

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 bandwidth b. It is easily shown that the estimator can be written as $\tilde{\beta}_{2,t}(\beta_1) = \hat{M}_{y,t} - \hat{M}_{X_1,t}' \beta_1$, where for any sequence A_t

$$\hat{M}_{A,t} := \left[\sum_{s=1}^{n} K_b(s/n - \tau) X_{2,s} X_{2,s}'\right]^{-1} \left[\sum_{s=1}^{n} K_b(s/n - \tau) X_{2,s} A_s'\right]. \tag{2.8}$$

Plugging $\tilde{\beta}_{2,t}(\beta_1)$ into the global log-likelihood together with some preliminary estimator of σ_t^2 (for example, the unconstrained estimator, $\hat{\sigma}_t^2$), a natural estimator would be the maximizer of this w.r.t. β_1.

However, due to bias problems with our kernel estimators for τ close to the two boundaries, $\tau = 0$ and 1, we first redefine the 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 negligible. 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.

We estimate β_1 by maximizing the trimmed version of the profile log-likelihood, $\tilde{\beta}_1 = \arg\max_{\beta_1} L_n(\beta_1, \tilde{\beta}_2, \hat{\sigma}^2)$. It is easily shown that $\tilde{\beta}_1$ is given as

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

where

$$\hat{y}_t = y_t - \hat{M}_{y,t}' X_{2,t}, \quad \hat{X}_{1,t} = X_{1,t} - \hat{M}_{X_1,t}' X_{2,t}.$$

We can substitute $\tilde{\beta}_1$ back into the expression of $\tilde{\beta}_{2,t}(\beta_1)$ to obtain an estimator of $\beta_{2,t}$ under the null:

$$\tilde{\beta}_{2,t} = \hat{M}_{y,t} - \hat{M}_{X_1,t}' \tilde{\beta}_1, \tag{2.10}$$

where $\hat{M}_{y,t}$ and $\hat{M}_{X_1,t}'$ are evaluated using the "nonparametric" 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 replacing $\varepsilon_t^2(\hat{\beta}_t)$ with $\varepsilon_t^2(\tilde{\beta}_t)$ in eq. (2.7).

While the above GLS estimator $\tilde{\beta}_1$ is asymptotically efficient (see Section 3), one may worry about its precision in samples of small and moderate 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,

$$\tilde{\beta}_1^w = \left[\sum_{t=1}^{n} \mathbb{I}_t(a)\, \hat{w}_t \hat{X}_{1,t} \hat{X}_{1,t}' \right]^{-1} \sum_{t=1}^{n} \mathbb{I}_t(a)\, \hat{w}_t \hat{X}_{1,t} \hat{y}_t', \tag{2.11}$$

where $\hat{w}_t = \hat{w}(t/n)$ for some (potentially estimated) weighting function $\hat{w} : [0,1] \mapsto \mathbb{R}_+$. With $\hat{w}_t = \hat{\sigma}_t^{-2}$, 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 β_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 ω satisfying $\int_0^1 \omega(\tau)\,d\tau = 1$. Ang and Kristensen (2012) 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 ω, 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

We propose to test H_0 by comparing the unrestricted and restricted fit of the model. We consider two ways of doing this: The first test is a F type test that compares the sums of squared residuals (SSR's) associated with the unrestricted and restricted model, while the second is a Wald type test that directly compares the restricted and unrestricted estimator of $\beta_1(\tau)$.

To obtain the F statistic, we first define the residuals under H_0 and H_A respectively,

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

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

$$SSR_0^w = \sum_{t=1}^n \mathbb{I}_t(a)\,\hat{w}_t \tilde{\varepsilon}_t^2, \quad SSR_A^w = \sum_{t=1}^n \mathbb{I}_t(a)\,\hat{w}_t \hat{\varepsilon}_t^2, \tag{2.12}$$

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 (2001) for varying-coefficient models. In particular, with $\hat{w}_t = \hat{\sigma}_t^{-2}$, 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 $\beta_1(\tau)$:

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

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 (2012), 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,

$$y_{n,t} = \beta_{n,t}' X_{n,t} + \sigma_{n,t} z_{n,t}, \quad t = 1, ..., n, \tag{3.13}$$

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 (2000), 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. Kristensen (2011), 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

$$\Lambda_{n,t} = \begin{bmatrix} \Lambda_{n,11,t} & \Lambda_{n,12,t} \\ \Lambda_{n,21,t} & \Lambda_{n,22,t} \end{bmatrix} \in \mathbb{R}^{m \times m}, \tag{3.14}$$

where $\Lambda_{n,kl,t} \equiv E\left[X_{n,k,t} X_{n,l,t}'\right] \in \mathbb{R}^{m_k \times m_l}$ for $k, l \in \{1, 2\}$. 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 β-mixing where the mixing coefficients,

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

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 $E\left[z_{n,t}^2 | X_{n,t}\right] = 1$ and $\lambda_{n,t} := E[(z_{n,t}^2 - 1)^2] < \infty$.

A.3 The sequences $\beta_{n,t}$, $\Lambda_{n,t}$ and $\sigma_{n,t}^2$ satisfy $\beta_{n,t} = \beta(t/n) + o(1)$, $\Lambda_{n,t} = \Lambda(t/n) + o(1)$, $\sigma_{n,t}^2 = \sigma^2(t/n) + o(1)$ 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 $\tau \in [0,1]$, $\Lambda(\tau) > 0$ and $\sigma^2(\tau) > 0$.

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

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

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

A.5 The covariance matrices Φ_w and Σ_w as defined below are non-singular:

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

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

where $\Lambda_{11|2,t} \equiv \Lambda_{11,t} - \Lambda_{12,t}\Lambda_{22,t}^{-1}\Lambda_{21,t}$ with $\Lambda_{kl,t}$, $k,l = 1,2$, defined in eq. (3.14).

A.6 The "semiparametric" bandwidth b satisfies $nb^{2r} \to 0$, $\log^2(n)/\left(nb^2\right) \to 0$ 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 β-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 α-mixing, but for the semiparametric estimation and testing, we will rely on U-statistics results that only exist for β-mixing sequences. The assumption of geometrically decaying mixing coefficients is only imposed to make proofs and remaining conditions simpler, and could most likely be weakened. 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 polynomial $\theta(\tau, z) = \beta_1(\tau)\, z + ... + \beta_q(\tau)\, z^q$ 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 and heteroskedasticity on the form $\sigma_t^2 = \sigma^2(X_t)$. We conjecture that our results can be extended to allow for autocorrelation and more general heteroskedasticity, 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 jumps in the coefficients. Following the arguments in Gijbels (2003), it is easily shown that if $\beta(\tau)$ jumps at some $\bar{\tau}$ such that $\beta(\bar{\tau}_-) \neq \beta(\bar{\tau}_+)$, then the proposed estimator is biased asymptotically at this point, $E[\beta(\bar{\tau})] = (\beta(\bar{\tau}_-) + \beta(\bar{\tau}_+))/2 + o(1)$ as $h \to 0$. 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 semi-parametric estimators and the test statistics remain valid when a finite number of jumps are present since these happen with measure zero. The assumption of r times differentiability is assumed for technical convenience, and could most likely be weakened to the assumption that the functions are Lipschitz; see 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 in the case $\hat{w}(\tau) = \hat{\sigma}^{-2}(\tau)$ since this satisfies $\hat{\sigma}^2(\tau) = \sigma^2(\tau) + O_P(h^r) + O_P(\sqrt{\log(n)/(nh)})$ uniformly on $\tau \in (1, a-1)$, c.f., Fan and Yao (1998).

The rank condition in Assumption (A.5) is employed to ensure identification and asymptotic normality of β_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 is a generalization of the result in Robinson (1989, Eq. 15.12) to allow for time varying volatility and non-stationary regressors. In particular, we allow X_t to include lagged values of y_t. The result tells us how pointwise confidence bands of the regression coefficients can be computed. A simple estimator of the asymptotic variance can be obtained by substituting in $\hat{\sigma}^2(\tau)$ together with:

$$\hat{\Lambda}(\tau) = \sum_{i=1}^{n} K_h(t/n - \tau) X_t X_t'. \tag{3.15}$$

The confidence bands 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.

The asymptotic distributional result in Theorem 3.1 is standard for nonparametric estimators. It reveals an important advantage of our estimation and testing strategy over alternative methods as proposed in Gao et al (2008) and Hidalgo (1995), 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. On the other hand, our estimators will be inconsistent if the regression model is not linear in X_t.

The next theorem states the asymptotic distribution of the estimators under H_0:

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

$$\sqrt{n}(\tilde{\beta}_1^w - \beta_1) \to^d N\left(0, \Sigma_w^{-1}\Phi_w\Sigma_w^{-1}\right), \tag{3.16}$$

where Φ_w and Σ_w are given in (A.5). Moreover, for any $\tau \in (0,1)$, as $h \to 0$, $nh \to \infty$ and $nh^{1+2r} \to 0$:

$$\sqrt{nh}(\tilde{\beta}_2(\tau) - \beta_2(\tau)) \to^d N\left(0, \|K\|^2 \Lambda_{22}^{-1}(\tau)\sigma^2(\tau)\right), \tag{3.17}$$

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 time-varying volatility. The theorem and its proof reveals that our estimator is first-order asymptotically equivalent to the weighted least-squares estimator of the infeasible regression $\bar{y}_t = \beta_1'\bar{X}_{1,t} + \varepsilon_t$, where $\bar{y}_t = y_t - M_{y,t}'X_{2,t}$, $\bar{X}_{1,t} := X_{1,t} - M_{X_1,t}'X_{2,t}$, and $M_{A,t}$ is the asymptotic limit of $\hat{M}_{A,t}$ defined in eq. (2.8). The asymptotic variance terms can be consistently estimated by

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

The asymptotic variance of $\tilde{\beta}_2(\tau)$ can be estimated using estimators similar to those given in eq. (2.7).

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 $\tau_t := t/n$ as i.i.d. draws from a uniform distribution which is independent of (X_t, z_t), 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,

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

Substituting these into the expression of $\mathcal{I}_0$, we obtain that $\mathcal{I}_0 = \int \sigma^{-2}(\tau)\Lambda_{11|2}(\tau)\,d\tau$, which matches up with Σ_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.

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

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

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

For the power analysis, consider the following class of local alternatives: $H_{A,n} : \beta_{1,n}(\tau) = \beta_1 + g_n(\tau)$, where $g_n(\tau)$ is a smooth function. For simplicity, assume that the distribution of X_t does not change under $H_{A,n}$; this rules out autoregressive models, but we conjecture that our results extend to this type of regression models. By inspection of Fan et al (2001, Proof of Theorem 7), it is easily seen that their arguments can be extended to a time series setting in the same way that Theorem 3.3 extends Fan et al (2001, Theorem 6). We are therefore able to conclude that the GLR test statistic is asymptotically optimal in the sense that it can detect the above class of local alternatives with optimal rate.

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,

$$\frac{W_n - \mu_n^W}{\sqrt{\nu_n^W}} \xrightarrow{d} N(0,1), \tag{3.18}$$

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 Ω_t as $\Omega_t = \Lambda_{11,t}\sigma_t^{-2}$, we obtain $r_K^W := m_1\kappa_2/\|K * K\|^2$, 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 Ω_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 (2012) that W_n is asymptotically more efficient in the sense of Pitmann; see also Hong and Lee (2008). We conjecture that the minimum-distance statistic will share the optimality of the GLR statistic for local alternatives.

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:

$$y_t = \alpha' W_t + \gamma'_{1,t} Z_{1,t} + \gamma'_{2,t} Z_{2,t} + \sigma_t z_t. \tag{4.19}$$

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:

$$\hat{\alpha}^w = \left[\sum_{t=1}^{n} \mathbb{I}_t(a)\, \hat{w}_t \hat{W}_t \hat{W}'_{Z,t}\right]^{-1} \sum_{t=1}^{n} \mathbb{I}_t(a)\, \hat{w}_t \hat{W}_t \hat{y}'_t, \quad \hat{\gamma}_t = \hat{M}_{h,t}(y) - \hat{M}_{h,t}(W)' \hat{\alpha}, \tag{4.20}$$

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

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

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:

$$\frac{F_{1,n} - \mu_{1,n}^F}{\sqrt{\nu_{1,n}^F}} \to^d N(0,1), \quad \frac{W_{1,n} - \mu_{1,n}^W}{\sqrt{\nu_{1,n}^W}} \to^d N(0,1) \tag{4.21}$$

where $\mu_{1,n}^F$, ν_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 γ_1.

To develop parametric tests of the functional form of varying coefficients, we now wish to test $H_2 : \gamma_{1,t} = \gamma_1(t/n; \theta)$ against the maintained hypothesis that eq. (4.19) 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 θ, $\tilde{\theta}$, 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. (4.21).

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.

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 justification for the proposed selection rule for b, but our simulation study reveals that it does an acceptable job. We focus on global bandwidth selection procedures. In situations where the type of time-variation in the parameters change over the sample, local bandwidth selection rules will probably perform better. The procedures suggested below are straightforward to adjust to allow for local bandwidths if that is of interest.

To estimate h, we employ a plug-in method. We here focus on the estimation of β_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

$$\mathrm{Bias}(\hat{\beta}(\tau)) = h^2 b(\tau) + o\left(h^2\right) \quad \text{with} \quad b(\tau) := \mu_2 \beta^{(2)}(\tau), \tag{5.22}$$

and

$$\mathrm{Var}(\hat{\beta}(\tau)) = \frac{1}{nh} v(\tau) + o(1/(Th)) \quad \text{with} \quad v(\tau) = \|K\|^2 \Lambda^{-1}(\tau)\sigma^2(\tau), \tag{5.23}$$

[1] It is outside of the scope of this paper to analyze this more general semiparametric estimation problem, but we 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.

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

$$h^* = \left[\frac{||V||}{||B||^2} \right]^{1/5} n^{-1/5}, \tag{5.24}$$

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]

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

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

Step 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 eqs. (2.7) and (3.15) with $h = \hat{h}_1$. We use these to obtain $\hat{V}_2 = \|K\|^2 \frac{1}{n} \sum_{t=1}^n \hat{\Lambda}_t^{-1} \otimes \hat{\sigma}_t^2$ and $\hat{B}_2 = \mu_2 \frac{1}{n} \sum_{t=1}^n \hat{\beta}_t^{(2)}$, 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:

$$\hat{h}_2 = \left[\frac{||\hat{V}_2||}{||\hat{B}_2||^2} \right]^{1/5} \times n^{-1/5}. \tag{5.25}$$

One could alternatively use 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 $E[||\tilde{\beta}_1^w - \beta_1||^2]$. 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

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

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 η_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), You and Chen (2006), 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. The main goal is to demonstrate that the developed tools indeed can detect and track time-varying parameters in most cases and are competitive with standard, parametric methods. We investigate the performance under four different DGP's for the time-varying parameters: We first consider the scenario where the coefficients either follow Random Walks (RW's) or a Smooth Transitions (ST's). Next, we compare the performance of our methods with parametric tests when the coefficients either exhibit breaks or are smooth. Importantly, the RW and break specifications are not covered by our asymptotic theory since the parameter paths in these cases are discontinuous and so one could worry that the nonparametric estimators and tests would perform poorly. As we shall see, however, they are still able to detect and the time variation and are competitive with parametric structural break tests. In fact, under smooth alternatives, our tests clearly outperform parametric structural break tests.

6.1. Performance under RW and ST Specifications

We here examine the performance of the proposed estimators and tests under the following two specifications:

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

where $v_\beta = 0.05$, $\alpha = 0.5$, and $\mu = \sigma = 1$. This allows us to investigate how smoothness of the parameter trajectories affect the estimators and tests. The volatility DGP is specified as a stochastic volatility model,

$$\log \sigma_t^2 = \log \sigma_{t-1}^2 + \eta_{\sigma,t}, \quad \eta_{\sigma,t} \sim \text{i.i.d.}N\left(0, v_\sigma^2\right), \quad \log \sigma_0^2 = 0.1,$$

where $v_\sigma = 0.05$. Finally, throughout we let the rescaled errors of the regression model be i.i.d. normally distributed, $z_t \sim N(0, 1)$. We note that the RW specification is ruled

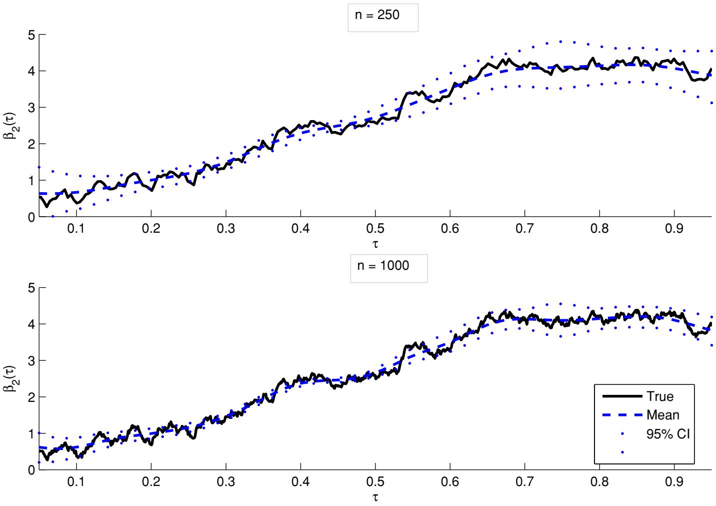

Figure 1. Simulation study, performance of estimator when $\beta_{2,t}$ follows a random walk, $n = 250$ and 1000.

out by Assumption A.3 and so the formal asymptotic results for the estimators and test statistics do not apply to this case. It is worth pointing out though that under the RW specification, $\beta_{2,[\tau n]}/\sqrt{n}$ weakly converge towards a Brownian motion whose trajectories are almost surely continuous, and so, in large samples, the RW specification satisfies A.3. Similarly for the RW specification of the log-volatility.

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 5. 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 σ_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 $\beta_{2,t}$, $\tilde{\beta}_{2,t}$ given in eq. (2.10), under the RW and ST specifications. From Figure 1 we see that while the estimator, by its nature, cannot completely track the discontinuous RW specification, 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 ($w_t = 1$) and weighted ($w_t = \tilde{\sigma}_t^{-2}$) semiparametric estimators of β_1. For comparison, we also report results for the infeasible OLS estimator which assumes knowledge

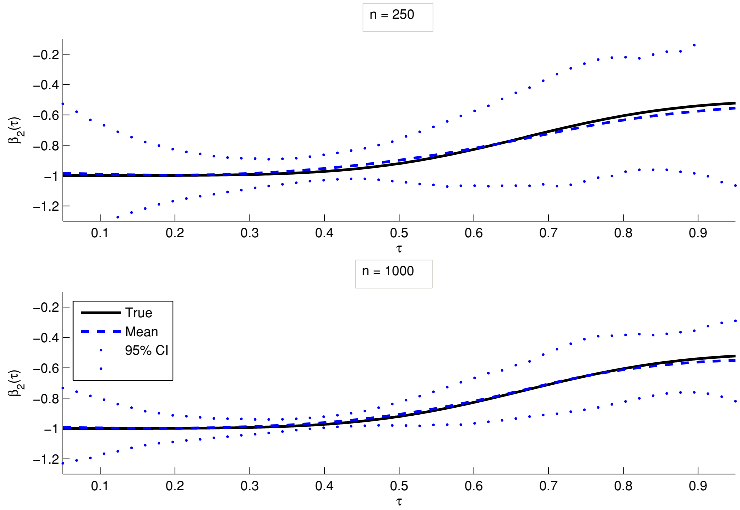

Figure 2. Simulation study, performance of estimator when $\beta_{2,t}$ follows a smooth transition, $n = 250$ and 1000.

of $\beta_{2,t}$ and σ_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, except for the ST specification with $n = 1000$. We have no good explanation for why the unweighted estimator performs best in this scenario.

	$\beta_{2,t}$ RW			$\beta_{2,t}$ ST		
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 deviation and RMSE of semiparametric estimators.
Notes: In each cell, bias, standard deviation 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 noteworthy, 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 implement the tests with both $\beta_{1,t}$ and $\beta_{2,t}$ are either random walks or smooth transitions. In Table 3, the rejection probabilites of the two tests are reported for the two different specification. We see that they have good power for moderate and large samples with the rejection rates increasing with sample size. As expected, the weighted test has significantly better power compared to the unweighted test. The power depends on the underlying data-generating mechanism and the tests do better at detecting 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).

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.

	$\beta_{2,t}$ 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 nonparametric F test 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 nonparametric F test using weighted and unweighted statistics.
Note: In each cell, the top and bottom number is power
of weighted and unweighted test respectively.

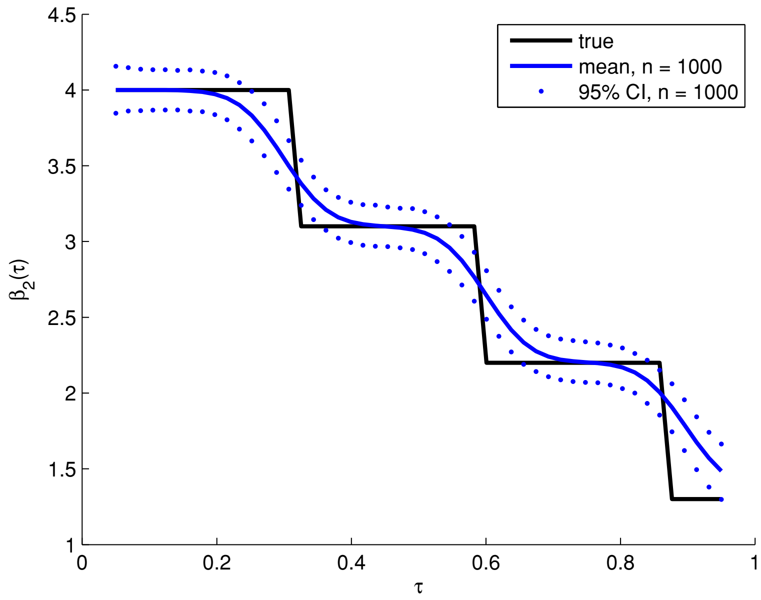

Figure 3. Simulation study, performance of $\tilde{\beta}_2(\tau)$ when $\beta_{2,t}$ follows the SB specification, $n = 1000$.

6.2. Comparison with Alternative Tests

The purpose of this part of the simulation study is to compare the size and power of our tests with those of parametric structural break (SB) test and the CUSUM test. We do this under the following two scenarios, where either the change is discontinuous or smooth:

$$\text{SB} : \beta_1(\tau) = -0.5 \times [3\tau] + 2, \quad \beta_2(\tau) = -0.9 \times [3\tau] + 4,$$
$$\text{SS} : \beta_1(\tau) = 16 \times \exp\left(-\frac{(\tau - 1/2)^2}{200}\right) + 2, \quad \beta_{2,t} = 20 \times \exp\left(-\frac{(\tau - 1/2)^2}{200}\right) + 4,$$

where $[z]$ denotes the nearest integer smaller than z. Since standard structural break tests do not accommodate for changing volatility, we here set $\sigma_t = \sigma$ constant.

We first examine how our semiparametric estimator of β_1 and nonparametric estimator of $\beta_2(\tau)$ are affected by the discontinuities in $\beta_2(\tau)$ under the SB specification when $H_0 : \beta_{1,t} = \beta_1$ is true and $n = 1000$ (similar results were obtained for $n = 250$ and 500). In Figure 3, we see that the estimator $\tilde{\beta}_2(\tau)$ is biased at the jump points but is still able to track the parameter paths fairly well. Moreover, the semiparametric estimator of β_1, $\tilde{\beta}_1^w$, does surprisingly well with a bias of 0.23×10^{-3} and a standard deviation of 16.99×10^{-3} which are comparable with the results reported in Table 1. We conclude that the semi-nonparametric estimators perform well even in the presence of jumps.

Next, we compare our nonparametric test with parametric structural break (SB) tests.

We focus on the null of both coefficients being constant, $H_0 : \beta_{1,t} = \beta_1$ and $\beta_{2,t} = \beta_2$, since the implementation of SB tests for partially time-varying coefficients can be quite delicate. Under the alternative, both $\beta_{1,t}$ and $\beta_{2,t}$ follows SB specifications. As noted earlier, our estimators are in this case inconsistent around the break points. At the same time, the SB specification falls within the framework of the structural break tests. Taking these two features together, we would expect the structural break tests to outperform our test. Table 3 and 4 report size and power of our tests relative to the $\sup F$ test of Andrews (1993), the $\exp F$ test of Andrews and Ploberger (1994) and the CUSUM test. First, in terms of size, all tests are comparable. Second, the power performance of the structural break tests is very good, but so is the one of our nonparametric tests which is only lagging slightly behind the parametric one. In contrast, the CUSUM test does a poor job; this is not surprising since it is well-known that the CUSUM test is only able to detect changes in the intercept, c.f. Ploberger and Krämer (1992). In conclusion, even under correct specification, our nonparametric test is competitive with SB tests, while CUSUM tests are not recommended.

	$\sup F$			$\exp F$			CUSUM			nonpar F		
n	1000	500	250	1000	500	250	1000	500	250	1000	500	250
$p = 1\%$	1	1.8	3.1	1.1	1.9	2.6	1.1	1.3	1.7	1.3	1.5	1.9
$p = 5\%$	5.6	7	8.9	5.1	7.5	8.8	4.4	5.9	7.2	6.1	6.1	6
$p = 10\%$	11.9	13.2	14.6	10.8	13.1	13.5	7.9	10.2	12.9	11.1	10.9	10.5

Table 3: Size comparison of tests under SB specification

	$\sup F$			$\exp F$			CUSUM			nonpar F		
n	1000	500	250	1000	500	250	1000	500	250	1000	500	250
$p = 1\%$	27.2	14.4	9.4	29.6	15.7	8.7	0.9	1.1	1.5	29.2	12.5	6.2
$p = 5\%$	50.5	32.1	21.3	55.8	33.9	24	4.1	5.3	6.7	51.4	30.3	16.9
$p = 10\%$	63.9	43.7	31.5	68.5	46.2	35	8.3	10.5	13.7	65.2	42.1	29.4

Table 4: Power comparison of tests under SB Specification

Next, we consider the SS specification. Under the null, the DGP is the same as in the SB case, and so the size results reported in Table 3 carry over to the SS specification. Next, we consider the power performance. Here, in contrast to the SB specification, the SS one favours our test since it is smooth and has a spiky shape. As is demonstrated in a number of different studies on power comparisons of nonparametric and parametric tests, the former are better at detecting spiky (high-frequency) alternatives (see Eubank and LaRiccia, 1992). This is confirmed by the power results reported in Tables 5: The power of the nonparametric F test is between 2-10 times better than the ones of the parametric tests, and so clearly dominates when the time variation falls outside of the SB framework.

	sup F			exp F			CUSUM			nonpar F		
n	1000	500	250	1000	500	250	1000	500	250	1000	500	250
$p = 1\%$	8.5	5.9	4.1	8.8	5.4	3.6	2.1	2.5	3.3	86.4	43.3	14.8
$p = 5\%$	26	16.8	11.5	27.3	17.8	12.4	5.4	9.2	11.9	95.4	68.8	36.4
$p = 10\%$	40.3	26.9	20.6	44.6	27.8	21.3	10.8	15.4	20	97.6	79.7	49.8

Table 5: Power comparison of tests under SS specification

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 (2012).

7.1. US Productivity

Hansen (2001) analyzed structural changes in US productivity within the framework of parametric structural break models. He found evidence of one significant break in 1992 with the possibility of two more breaks in 1963 and 1982 respectively. The aim here is to see whether these findings are 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 monthly 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 $H_0 : \rho_{2,t} = \rho_{3,t} = 0$ 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 μ_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 4-5, the nonparametric estimates of the parameter trajectories are plotted together with corresponding structural break estimates when allowing for one and three breaks, respectively. 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 4 shows the trajectory of the intercept, μ_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 ρ_t as plotted in Figure 5: 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

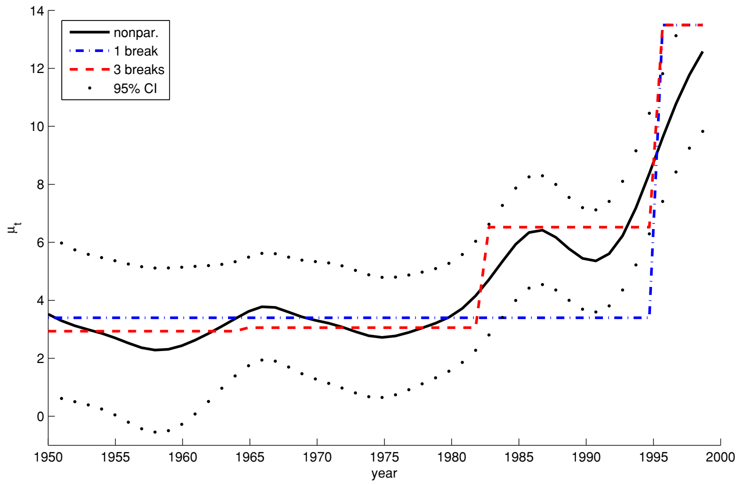

Figure 4. Estimates of structural change in μ, 1950-2000.

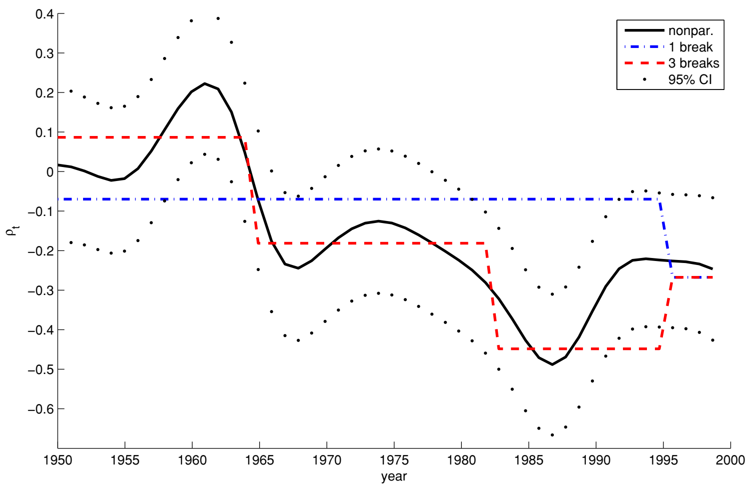

Figure 5. Estimates of time variation in ρ, 1950-2000.

well for the AR coefficient with much tighter confidence intervals (note different scale of the y-axis for the two figures).

To formally test the structural break models against the nonparametric alternative, 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 $F_n = 16.25$ 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 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) that a 3-breaks model adequately 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. Within this class of models, the short-term interest rate is driven by a linear combination of factors where the factors in turn solve a VAR model with constant coefficients; see Duffie and Kan (1996). However, there is amble empirical evidence that affine models are unstable over time. These instabilities have major implications for forecasting the yield curve and for bond pricing. Most studies examining time-variation in affine models 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 involved in estimating dynamic models with latent variables, most studies have confined themselves to single-factor models despite consensus 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.

To estimate the affine model, we first extract the factors from yield curve data. As demonstrated in Joslin, Singleton and Zhu (2011), for any set of d yields, $y_t(1), ..., y_t(d)$, the factors can be chosen as any linear combinations of those. With $Y_t \in \mathbb{R}^d$ denoting a given linear combination, the resulting affine model with time-varying coefficients take the form

$$r_t = b_{0,r} + b_{1,r}' Y_t, \quad \Delta Y_t = B_{0,t} + B_{1,t} Y_{t-1} + e_{Y,t},$$

with $e_{Y,t} \sim$i.i.d. $N(0, \Sigma_{Y,t})$. In our application, we estimate a three factor model using a data set of yields from the so-called Eurodollar term structure. The data set consists of

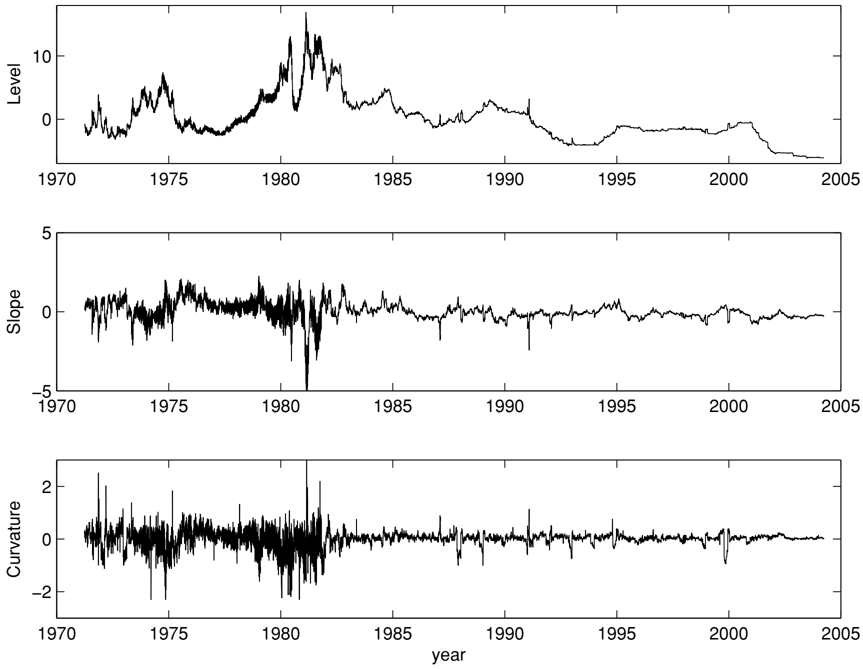

Figure 6. Time series plot of the level, slope and curvature factor, 1971-2004.

$n = 8669$ daily observations of the 1, 3, and 6 months Eurodollar yield for the period 1971-2004. We can freely rotate the yields and here choose our factors as the so-called level, slope and curvature of the yield curve. With $y_t(1)$, $y_t(3)$ and $y_t(6)$ denoting the three observed yields, the factors are constructed as $Y_{1,t} = y_t(1)$ ("level"), $Y_{2,t} = y_t(6) - y_t(1)$ ("slope"), and $Y_{3,t} = y_t(1) + y_t(6) - 2y_t(3)$ ("curvature"). This rotation helps to give a natural interpretation of the factors which at the same time are close to being orthogonal, see Litterman and Scheinkman (1991). The resulting demeaned factors are plotted in Figure 6. 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 their dynamics.

We estimate the VAR dynamics of the three factors allowing all coefficients to be time-varying. We focus in the following on dynamics of the level factor, $\Delta Y_{1,t} = b_{0,t} + b'_{1,t}Y_{t-1} + \varepsilon_t$; results for the two other factors are available upon request. The kernel estimates of the loadings $b_{1,t} = (b_{1,t}(1), b_{1,t}(2), b_{1,t}(3))$ are shown in Figures 7-9. For comparison, we also report the OLS estimates for each of the factor loadings in the three figures. These estimates deliver an informal rejection of the null of constant VAR coefficients with all three exhibiting substantial variation over time. In particular, the Fed Experiment changed the yield curve dynamics quite dramatically. 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 insignificant

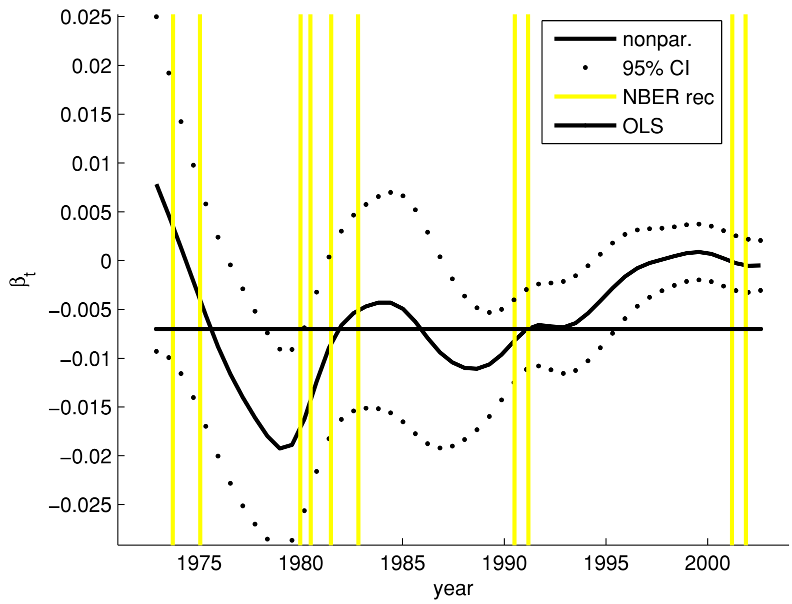

Figure 7. Estimated level factor loading with pointwise 95% confidence bands, 1971-2004.

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.

We 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 regression coefficients are constant; we strongly reject all three nulls with p-values well below 1%. Next, we test whether any of the coefficients individually is constant; again, we heavily reject the three nulls. Finally, the null of all coefficients being constant is strongly rejected. We conclude that there is strong evidence of time-variation in all regression coefficients. 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 coefficients 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 coefficients for the the level and slope variables, while the curvature factor's coefficients remains time-varying.

To see whether the estimated variation in the coefficients is driven by underlying macro factors, the reported NBER recessions within the sample period are shown in Figures 7-9 together with the estimates. There appears to be some correlation between whether the economy is in a recession and changes in the coefficients, 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 parameter 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

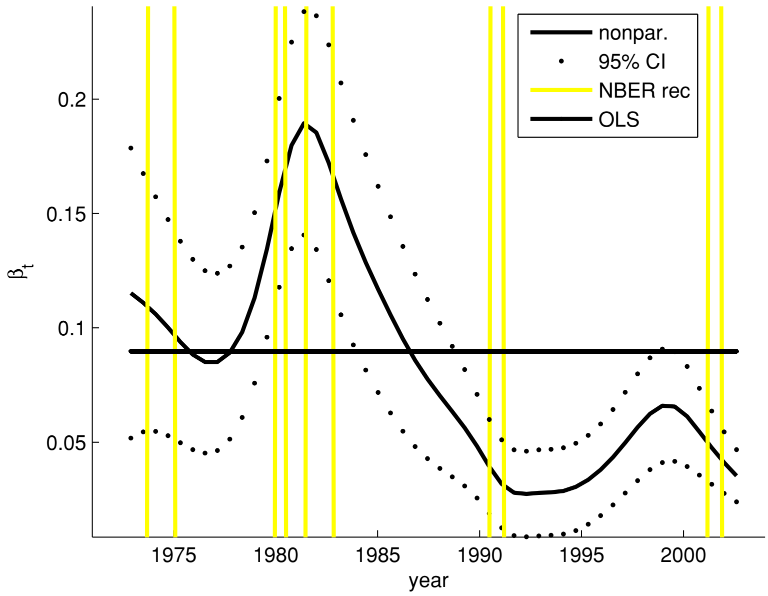

Figure 8. Estimated slope factor loading with pointwise 95% confidence bands, 1971-2004.

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 coefficients are explained by underlying macro factors. Looking at the individual regression 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.

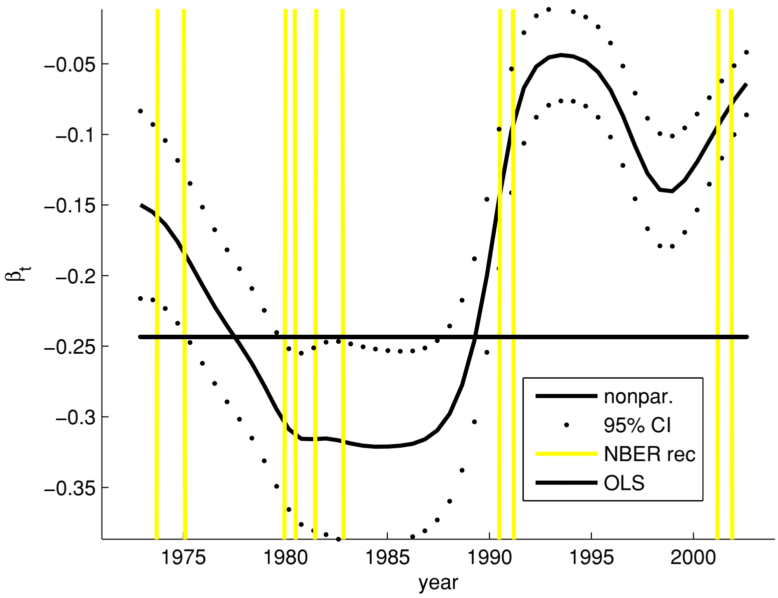

Figure 9. Estimated curvature factor loading with pointwise 95% confidence bands, 1971-2004.

	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 6: 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. The theory is however silent about how to choose bandwidths required for the semiparametric estimators and test statistics in finite samples. These need to vanish at non-standard rates and so standard methods cannot be used. Data-driven bandwidth selection procedures for these are currently not available, and it would be highly useful to develop and analyze such.

The proposed estimators are not able to consistently detect jumps in the parameter paths. By adjusting the estimators along the lines of Gijbels (2003) and Gijbels, Lambert and Qiu (2007) who develop jump-preserving kernel smoothers in a cross-sectional setting, we expect that the asymptotic theory developed in the present work will still go through with only minor modifications. The formal analysis of jump-preserving estimators and associated test statistics in a time series setting is left for future research.

We have throughout assumed that the regressors are (locally) stationary. However, it is well-known that it is difficult to separate the effects of structural change from those of unit root-type behaviour. It would therefore be relevant to examine how our estimators and tests perform under unit-root type dynamics to see if they remain consistent.

It would also be of interest to extend the results to nonlinear models with time-varying parameters. It is easily seen that the proposed estimators and test statistics are straightforward to extend to nonlinear models whose time-invariant 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; see Robinson (1991) and Fryzlewicz, Sapatinas, and Subba Rao (2008) for estimation in some particular models within this general class.

ACKNOWLEDGEMENTS

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, and Ju Hyun Kim for excellent research assistance. 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 Research Foundation through a grant to Center for Research in Econometric Analysis of Time Series, the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001 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.

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.

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 (2012). Testing Conditional Factor Models. *forthcoming in Journal of Financial Economics*.

Bai. J. (1999). Likelihood Ratio Tests for Multiple Structural Changes. *Journal of Econometrics 26*, 299–323.

Bai, J. and P. Perron (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica 66*, 47–79.

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

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.

Chen, B. and Y. Hong (2012). Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression. *Econometrica 80*, 1157–1183.

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

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

Duffie, D. and R. Kan (1996). A Yield-factor Model of Interest Rates. *Mathematical Finance 6*, 379–406.

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

Eubank, R.L. and V.N. LaRiccia (1992). Asymptotic comparison of Cramér-Von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Annals of Statistics 20*, 2071–2086.

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

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., C. Zhang and J. Zhang (2001). 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.

Fryzlewicz, P., T. Sapatinas, and S. Subba Rao (2008). Normalized Least-Squares Estimation in Time-Varying ARCH Models. *Annals of Statistics 36*, 742–786.

Gao, J., I. Gijbels and S. van Bellegem (2008). Nonparametric Simultaneous Testing for Structural Breaks. *Journal of Econometrics 143*, 123–142.

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 M.G. Akritas and D.N. Politis (Eds.) *Recent Advances and Trends in Nonparametric Statistics*, 183–202. Amsterdam: 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*, 235–272.

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. (2000). 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 53*, 173–187.

Hidalgo, J. (1995). A Nonparametric Conditional Moment Test for Structural Stability. *Econometric Theory 11*, 671–698.

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.

Joslin, S., K.J. Singleton and H. Zhu (2011). A New Perspective on Gaussian Dynamic Term Structure Models. *Review of Financial Studies 24*, 926–970.

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. (2012). Stationary Approximations of Time-inhomogenous Markov Chains with Applications. Manuscript, Dep. of Economics, UCL.

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.

Litterman, R., and J. Scheinkman (1991). Common Factors Affecting Bond Returns. *Journal of Fixed Income 3*, 54–61.

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.

Müller, U. and P. Petalas (2010). Efficient Estimation of the Parameter Path in Unstable Time Series Models. *Review of Economic Studies 77*, 1508–1539.

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

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

Ploberger, W. and Walter Krämer (1992). The CUSUM Test with OLS Residuals. *Econometrica 60*, 271–285.

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

Robinson, P.M. (1987). Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form. *Econometrica 55*, 875-891.

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

Robinson, P.M. (1989). Nonparametric Estimation of Time-Varying Parameters. In P. Hackl (Ed.), *Statistical Analysis and Forecasting of Economic Structural Changes*, 253–264. New York: Springer-Verlag.

Robinson, P.M. (1989). Nonparametric Estimation of Time-Varying Parameters. In P. Hackl and A. Westlund (Eds.), *Economic Structural Change in Analysis and Forecasting*, 179–190. New York: Springer-Verlag

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.

You, J. and G. Chen (2006). Wild Bootstrap Estimation in Partially Linear Models with Heteroscedasticity. *Statistics & Probability Letters 76*, 340–348.

APPENDIX 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: With $K_{t,\tau} = K_h(t/n - \tau)$, we have

$$\hat{\beta}(\tau) - \beta(\tau) = \left[\sum_{t=1}^{n} K_{t,\tau} X_t X_t'\right]^{-1} \sum_{t=1}^{n} K_{t,\tau} X_t X_t' \{\beta_t - \beta(\tau)\} + \left[\sum_{t=1}^{n} K_{t,\tau} X_t X_t'\right]^{-1} \sum_{t=1}^{n} K_{t,\tau} X_t \varepsilon_t. \tag{A.1}$$

By Lemma B.10, we obtain $n^{-1}\sum_{t=1}^{n} K_{t,\tau} X_i X_i' = \Lambda(\tau) + o_P(1)$, while

$$E\left[\left\|\frac{1}{n}\sum\nolimits_{t=1}^{n} K_{t,\tau} X_t X_t' \{\beta_t - \beta(\tau)\}\right\|\right] \leq C \sup_{t:|t/n-\tau|<Bh} \|\beta(t/n) - \beta(\tau)\| = O(h^r), \tag{A.2}$$

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

$$\hat{\beta}(\tau) - \beta(\tau) = \Lambda^{-1}(\tau) \frac{1}{n} \sum_{t=1}^{n} K_{t,\tau} X_t \varepsilon_t + O_P(h^r) \tag{A.3}$$

To complete the proof, we show that $\sum_{t=1}^{n} u_{n,t}/\sqrt{n} \to^{d} N\left(0, \left\|K^2\right\| \Lambda(\tau)\sigma^2(\tau)\right)$ as $nh \to \infty$, where $u_{n,t} = \sqrt{h}K_{t,\tau}X_t\varepsilon_t$ is a MGD w.r.t. $\mathcal{F}_t = \mathcal{F}(X_t, z_t, X_{t-1}, z_{t-1}, ...)$. This is done by verifying the conditions of Lemma B.13: As $nh \to \infty$,

$$\frac{1}{n}\sum_{t=1}^{n} E\left[u_{n,t}u_{n,t}'\right] = \frac{h}{n}\sum_{t=1}^{n} K_{t,\tau}^2\sigma_t^2\Lambda_t + o(1) \tag{A.4}$$

$$= \frac{1}{h}\int K^2\left(\frac{s-\tau}{h}\right)\sigma^2(s)\Lambda(s)\,ds + o(1) \tag{A.5}$$

$$= \left\|K^2\right\|\sigma^2(\tau)\Lambda(\tau) + o(1), \tag{A.6}$$

$$\frac{1}{n^{1+\delta/2}}\sum_{t=1}^{n} E\left[\left\|u_{n,t}\right\|^{2+\delta}\right] = \frac{h^{1+\delta/2}}{n^{1+\delta/2}}\sum_{t=1}^{n} K_{t,\tau}^{2+\delta}\sigma_t^{2+\delta}E\left[\left\|X_t\right\|^{2+\delta}\left|z_t\right|^{2+\delta}\right] \tag{A.7}$$

$$= C\frac{1}{(nh)^{\delta/2}}\sigma^{2+\delta}(\tau)\int K^{2+\delta}(z)\,dz = o(1). \tag{A.8}$$

□

Proof of Theorem 3.2: First note that, once we have shown eq. (3.16), we can treat β_1 as known, when deriving eq. (3.17) since it converges with $\sqrt{n}$-rate. But we can then employ the same arguments as in the proof of Theorem 3.1 to obtain eq. (3.17).

To show eq. (3.16), define for any two sequences A_t and B_t and any weighting function w,

$$S_{A,B}^{w} = \frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t(a)\,w_tA_tB_t', \tag{A.9}$$

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

$$\tilde{\beta}_1^w = (S_{X_1-\hat{X}_1}^{\hat{w}})^{-1}S_{X_1-\hat{X}_1,y-\hat{y}}^{\hat{w}}, \quad \bar{\beta}_1^w = (S_{X_1-\hat{X}_1}^{w})^{-1}S_{X_1-\hat{X}_1,y-\hat{y}}^{w}. \tag{A.10}$$

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

$$\sqrt{n}\left\{\bar{\beta}_1^w - \beta_1\right\} \to^{d} N\left(0, \Sigma_w^{-1}\Phi_w\Sigma_w^{-1}\right), \tag{A.11}$$

$$\sqrt{n}\{\tilde{\beta}_1^w - \bar{\beta}_1^w\} = o_P(1). \tag{A.12}$$

Proof of eq. (A.11): Define $V_t := X_{1,t} - \xi_t$, where $\xi_t := M_{X_1,t}'X_{2,t}$, and, for any random sequence A_t, $M_{A,t} := E\left[X_{2,t}X_{2,t}'\right]^{-1}E\left[X_{2,t}A_t'\right]$. We then have

$$E\left[X_{2,t}y_t'\right] = E\left[X_{2,t}X_{1,t}'\right]\beta_1 + E\left[X_{2,t}X_{1,t}'\right]\beta_{2,t}, \tag{A.13}$$

such that $M_{y,t} = M_{X_1,t}\beta_1 + \beta_{2,t}$ and

$$y_t - M_{y,t}'X_{2,t} = \beta_1'\left[X_{1,t} - M_{X_1,t}'X_{2,t}\right] + \varepsilon_t = \beta_1'V_t + \varepsilon_t \tag{A.14}$$

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

$$\bar{\beta}_{2,t} = \left[\sum_{s=1}^{n}K_{s,t}X_{2,s}X_{2,s}'\right]^{-1}\sum_{s=1}^{n}K_{s,t}X_{2,s}X_{2,s}'\beta_{2,t}, \tag{A.15}$$

such that $X_{1,t} - \hat{X}_{1,t} = \{\xi_t - \hat{\xi}_t\} + \{V_t - \hat{V}_t\}$ and

$$y_t - \hat{y}_t = \beta_1' \left[X_{1,t} - \hat{X}_{1,t}\right] + \left[\beta_{2,t} - \bar{\beta}_{2,t}\right]' X_{2,t} + \varepsilon_t - \hat{\varepsilon}_t. \quad \text{(A.16)}$$

In total,

$$S^w_{X_1-\hat{X}_1} = S^w_V + S^w_{\hat{V}} + S^w_{\xi-\hat{\xi}} + 2S^w_{V,\xi-\hat{\xi}} + 2S^w_{\hat{V},\xi-\hat{\xi}} + 2S^w_{V,\hat{V}} \quad \text{(A.17)}$$

$$S^w_{X_1-\hat{X}_1,y-\hat{y}} = S^w_{X_1-\hat{X}_1}\beta_1 + S^w_{X_1-\hat{X}_1,[\beta_2-\bar{\beta}_2]'X_2} + S^w_{X_1-\hat{X}_1,\varepsilon} + S^w_{X_1-\hat{X}_1,\hat{\varepsilon}}, \quad \text{(A.18)}$$

where

$$S^w_{X_1-\hat{X}_1,[\beta_2-\bar{\beta}_2]'X_2} = S^w_{\xi-\hat{\xi},[\beta_2-\bar{\beta}_2]'X_2} + S^w_{V,[\beta_2-\bar{\beta}_2]'X_2} - S^w_{\hat{V},[\beta_2-\bar{\beta}_2]'X_2}, \quad \text{(A.19)}$$

$$S^w_{X_1-\hat{X}_1,\varepsilon} = S^w_{\xi-\hat{\xi},\varepsilon} + S^w_{V,\varepsilon} - S^w_{\hat{V},\varepsilon}, \quad S^w_{X_1-\hat{X}_1,\hat{\varepsilon}} = S^w_{\xi-\hat{\xi},\hat{\varepsilon}} + S^w_{V,\hat{\varepsilon}} - S^w_{\hat{V},\hat{\varepsilon}}. \quad \text{(A.20)}$$

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 negligible. This yields the desired result.

Proof of eq. (A.12): Observe that

$$\tilde{\beta}^w_1 = \beta_1 + (S^{\hat{w}}_{X_1-\hat{X}_1})^{-1}\left[S^{\hat{w}}_{X_1-\hat{X}_1,[\beta_2-\bar{\beta}_2]'X_2} + S^{\hat{w}}_{X_1-\hat{X}_1,\varepsilon} + S^{\hat{w}}_{X_1-\hat{X}_1,\hat{\varepsilon}}\right], \quad \text{(A.21)}$$

$$\bar{\beta}^w_1 = \beta_1 + (S^{\hat{w}}_{X_1-\hat{X}_1})^{-1}\left[S^{w}_{X_1-\hat{X}_1,[\beta_2-\bar{\beta}_2]'X_2} + S^{w}_{X_1-\hat{X}_1,\varepsilon} + S^{w}_{X_1-\hat{X}_1,\hat{\varepsilon}}\right], \quad \text{(A.22)}$$

c.f. the proof of eq. (A.11). We therefore have

$$\begin{aligned}
\tilde{\beta}^w_1 - \bar{\beta}^w_1 = {} & (S^{\hat{w}}_{X_1-\hat{X}_1})^{-1}\left[S^{\hat{w}}_{X_1-\hat{X}_1,[\beta_2-\bar{\beta}_2]'X_2} - S^{w}_{X_1-\hat{X}_1,[\beta_2-\bar{\beta}_2]'X_2}\right] && \text{(A.23)}\\
& +(S^{\hat{w}}_{X_1-\hat{X}_1})^{-1}\left[S^{\hat{w}}_{X_1-\hat{X}_1,\varepsilon} - S^{w}_{X_1-\hat{X}_1,\varepsilon}\right] && \text{(A.24)}\\
& +(S^{\hat{w}}_{X_1-\hat{X}_1})^{-1}\left[S^{\hat{w}}_{X_1-\hat{X}_1,\hat{\varepsilon}} - S^{w}_{X_1-\hat{X}_1,\hat{\varepsilon}}\right] && \text{(A.25)}\\
& + \left[(S^{\hat{w}}_{X_1-\hat{X}_1})^{-1} - (S^{w}_{X_1-\hat{X}_1})^{-1}\right]\left[S^{w}_{X_1-\hat{X}_1,[\beta_2-\bar{\beta}_2]'X_2} + S^{w}_{X_1-\hat{X}_1,\varepsilon} + S^{w}_{X_1-\hat{X}_1,\hat{\varepsilon}}\right] && \text{(A.26)}\\
=: {} & \hat{B}^{-1}(\hat{A}_1 - A_1) + \hat{B}^{-1}(\hat{A}_2 - A_2) + \hat{B}^{-1}(\hat{A}_3 - A_3) + (\hat{B}^{-1} - B^{-1})A_4 && \text{(A.27)}
\end{aligned}$$

First note that, from the proof of eq. (A.11), $A_4 = O_P(1/\sqrt{n})$. Next, by a second order Taylor expansion of $\hat{B}^{-1}$ around B^{-1}, $\hat{B}^{-1} = B^{-1} - B^{-1}(\hat{B} - B)B^{-1} + O(||\hat{B} - B||^2)$, where, with $\hat{\Delta} := \sup_{a\leq\tau\leq 1-a}|\hat{w}(\tau) - w(\tau)|$,

$$||\hat{B} - B|| \leq \hat{\Delta} \times \frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t(a)(X_{1,t} - \hat{X}_{1,t})'(X_{1,t} - \hat{X}_{1,t}) = \hat{\Delta} \times \mathrm{tr}\{S_{X_1-\hat{X}_1}\} = O_P(\hat{\Delta}), \quad \text{(A.28)}$$

where we have used that, by the same reasoning as in the proof of eq. (A.11) (with $w_t = 1$), $S_{X_1-\hat{X}_1} = O_P(1)$. This implies that $||\hat{B}^{-1} - B^{-1}|| = O_P(\hat{\Delta})$ and $\hat{B}^{-1} = B^{-1} + o_P(1)$.

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

$$\begin{aligned}
||\hat{A}_1 - A_1|| &\leq \hat{\Delta} \times \sup_{a\leq\tau\leq 1-a}\left\|\beta_{2,t} - \bar{\beta}_{2,t}\right\| \times \frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t(a)||X_{1,t} - \hat{X}_{1,t}||\,\|X_{2,t}\| && \text{(A.29)}\\
&= \hat{\Delta} \times \sup_{a\leq\tau\leq 1-a}\left\|\beta_{2,t} - \bar{\beta}_{2,t}\right\| \times O_P(1), && \text{(A.30)}
\end{aligned}$$

Next,

$$\hat{A}_2 - A_2 = \frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t(a)\left(\hat{w}_t^2 - w_t^2\right)(X_{1,t} - \hat{X}_{1,t})'\varepsilon_t = \frac{1}{\sqrt{n}}\left\{Z_n(\hat{w}) - Z_n(w)\right\}, \tag{A.31}$$

where

$$Z_n(f) := \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{I}_t(a) f(t/n)(X_{1,t} - \hat{X}_{1,t})'\varepsilon_t \quad f \in \mathcal{F}, \tag{A.32}$$

and $\mathcal{F} = \left\{f : [0,1] \mapsto \mathbb{R} | \sup_{0\leq\tau\leq1}|f(\tau)| \leq F\right\}$ 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, $Z_n(f) - Z_n(g) = Z_n(f-g) = S^{f-g}_{X_1-\hat{X}_1,\varepsilon}$ for any $f, g \in \mathcal{F}$. By the same arguments employed in the proofs of Lemmas B.2, B.4 and B.6, one can show that for some constant $C > 0$,

$$E\left[\|Z_n(f) - Z_n(g)\|^2\right] = E[\|S^{f-g}_{X_1-\hat{X}_1,\varepsilon}\|^2] \leq C \times \sup_{0\leq\tau\leq1}|f(\tau) - g(\tau)|^2, \tag{A.33}$$

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

$$Z_n(\hat{w}) - Z_n(w) = \{Z_n(\hat{w}) - Z(\hat{w})\} - \{Z_n(w) - Z(w)\} + \{Z(w) - Z(\hat{w})\} = o_P(1). \tag{A.34}$$

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,

$$SSR^w(\beta) = \sum_{t=1}^{n}\mathbb{I}_t(a) w_t(y_t - \beta_t' X_t)^2. \tag{A.35}$$

We note for future use that the following expansion holds:

$$\begin{aligned} SSR^w(\beta) - SSR^w(\beta_0) &= \frac{\partial SSR^w(\beta_0)}{\partial\beta}(\beta - \beta_0) + \frac{1}{2}(\beta - \beta_0)'\frac{\partial^2 SSR^w(\beta_0)}{\partial\beta\partial\beta'}(\beta - \beta_0) \qquad \text{(A.36)} \\ &= -2\sum_{t=1}^{n}\mathbb{I}_t(a) w_t\varepsilon_t X_t'(\beta_t - \beta_{0,t}) + \sum_{t=1}^{n}\mathbb{I}_t(a) w_t(\beta_t - \beta_{0,t})' X_t X_t'(\beta_t - \beta_{0,t}) \end{aligned}$$

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

$$\bar{F}_n = \frac{n}{2}\frac{SSR^w(\tilde{\beta}) - SSR^w(\hat{\beta})}{SSR^w(\hat{\beta})} = \frac{SSR^w(\tilde{\beta}) - SSR^w(\hat{\beta})}{2\int_0^1 w(\tau)\sigma^2(\tau)\,d\tau + o_P(1)}, \tag{A.37}$$

where

$$\begin{aligned} SSR^w(\tilde{\beta}) - SSR^w(\hat{\beta}) &= \left\{SSR^w(\tilde{\beta}) - SSR^w(\beta_0)\right\} - \left\{SSR^w(\hat{\beta}) - SSR^w(\beta_0)\right\} && \text{(A.38)} \\ &=: \Delta SSR_1 - \Delta SSR_2. && \text{(A.39)} \end{aligned}$$

Combining the expansion in eq. (A.36) (with $\beta = \hat{\beta}$) and the representation given in eq.

(A.3), we obtain

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

By Lemma B.13, $O_P\left(h^r\right)\times\sum_{t=1}^{n}\mathbb{I}_t\left(a\right)w_t\varepsilon_t X_t' = O_P\left(h^r\sqrt{n}\right) = o_P(1/\sqrt{h})$. Thus, we can ignore this term, and decompose the remaining terms in $\Delta SSR_{2,1}$ into

$$\Delta SSR_{2,1} \simeq \frac{2K\left(0\right)}{nh}\sum_{t=1}^{n}\mathbb{I}_t\left(a\right)w_t\varepsilon_t^2 X_t'\Lambda_t^{-1}X_t + \frac{2}{n}\sum_{t\neq u}\mathbb{I}_t\left(a\right)w_t\varepsilon_t X_t' K_{t,u}\Lambda_t^{-1}X_u\varepsilon_u, \quad (A.40)$$

where the average in the first term satisfies

$$E\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t\left(a\right)w_t\varepsilon_t^2 X_t'\Lambda_t^{-1}X_t\right] = \frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t\left(a\right)w_t\sigma_t^2 E\left[X_t'\Lambda_t^{-1}X_t\right] = m\int_0^1 w\left(s\right)\sigma^2\left(s\right)ds + o\left(1\right), \quad (A.41)$$

and, due to the mixing conditions, $\mathrm{Var}\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{I}_t\left(a\right)w_t\varepsilon_t^2 X_t'\Lambda_t^{-1}X_t\right] = O\left(n^{-1+\epsilon}\right)$ for some small $\epsilon > 0$. 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

$$\begin{aligned}
\Delta SSR_{2,2} &\simeq \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\mathbb{I}_t\left(a\right)w_t\varepsilon_s^2 X_s' K_{s,t}^2\Lambda_t^{-1}X_tX_t'\Lambda_t^{-1}X_s \\
&\quad +\frac{1}{n^2}\sum_{s\neq u}\sum_{t=1}^{n}\varepsilon_s X_s' K_{s,t}\mathbb{I}_t\left(a\right)w_t\Lambda_t^{-1}X_tX_t'\Lambda_t^{-1}K_{t,u}X_u\varepsilon_u \\
&=: \Delta SSR_{2,21} + \Delta SSR_{2,22}. \qquad (A.42)
\end{aligned}$$

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

$$\Delta SSR_{2,22} = \frac{1}{n}\sum_{s\neq u}\mathbb{I}_s\left(a\right)w_s\varepsilon_s X_s'\Lambda_s^{-1}\left(K*K\right)_{s,u}X_u\varepsilon_u + \frac{1}{n^2}\sum_{s\neq t\neq u}\Phi_{s,t,u}, \quad (A.43)$$

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,

$$\varphi_{s,t,u} = \varepsilon_s X_s' K_{s,t}\mathbb{I}_t\left(a\right)w_t\Lambda_t^{-1}X_tX_t'\Lambda_t^{-1}K_{t,u}X_u\varepsilon_u, \quad (A.44)$$

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

$$\frac{1}{n^2}\sum_{s\neq t\neq u}\Phi_{s,t,u} = O_P\left(\frac{1}{\sqrt{nh^5}}\right) = o_P\left(\frac{1}{\sqrt{h}}\right). \quad (A.45)$$

In total,

$$\Delta SSR_2 \simeq \frac{m}{h}\left[\kappa_2 - 2K\left(0\right)\right]\int_0^1 w\left(\tau\right)\sigma^2\left(\tau\right)d\tau + \frac{1}{n}\sum_{t\neq u}\mathbb{I}_s\left(a\right)w_t\varepsilon_t X_t'\Lambda_t^{-1}\left[\left(K*K\right)_{t,u} - 2K_{t,u}\right]X_u\varepsilon_u. \tag{A.46}$$

To analyze ΔSSR_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 β_1 in ΔSSR_1. Moreover, by the same arguments used to show eq. (A.3), $\tilde{\beta}_{2,t} - \beta_{2,t} = \Lambda_{22,t}^{-1}\frac{1}{n}\sum_{s=1}^n K_{s,t}X_{2,s}\varepsilon_s + O_P\left(h^r\right)$ uniformly in t, and so

$$\begin{aligned}\Delta SSR_1 &\simeq -\frac{1}{n}\sum_{t=1}^n\sum_{u=1}^n \mathbb{I}_t\left(a\right)w_t\varepsilon_t X_{2,t}'\Lambda_{22,s}^{-1}X_{2,u}'K_{s,u}\varepsilon_u \qquad (A.47)\\ &+\frac{1}{n^2}\sum_{s=1}^n\sum_{t=1}^n\sum_{u=1}^n \mathbb{I}_t\left(a\right)w_t\varepsilon_s X_{2,s}'K_{s,t}\Lambda_{22,t}^{-1}X_{2,t}X_{2,t}'\Lambda_{22,t}^{-1}K_{t,u}X_{2,u}'\varepsilon_u\\ &=: -\Delta SSR_{1,1} + \Delta SSR_{1,2}.\end{aligned}$$

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,

$$\begin{aligned}\Delta SSR_1 &\simeq \frac{m_2}{h}\left[\kappa_2 - 2K\left(0\right)\right]\int_0^1 w\left(s\right)\sigma^2\left(s\right)ds \qquad (A.48)\\ &+\frac{1}{n}\sum_{s\neq u}\mathbb{I}_s\left(a\right)w_s\varepsilon_s X_{2,s}'\left[\left(K*K\right)_{s,u} - 2K_{s,u}\right]\Lambda_{22,u}^{-1}X_{2,u}\varepsilon_u. \qquad (A.49)\end{aligned}$$

Combining the expressions of ΔSSR_1 and ΔSSR_2, we now have

$$\bar{F}_n \simeq \frac{m_1\left[K\left(0\right) - \frac{1}{2}\kappa_2\right]}{h} + \frac{n^{-1}\sum_{s\neq t}\phi_{1,n}\left(u_s,u_t\right)}{2\int_0^1 w\left(\tau\right)\sigma^2\left(\tau\right)d\tau}, \tag{A.50}$$

where $u_t = \left(t/n, \varepsilon_t, \bar{X}_t\right)$ with $\bar{X}_t := X_{1,t}' - \omega_t X_{2,t}$, $\Lambda_{11.2,t} := \Lambda_{11,t} - \omega_t\Lambda_{21,t}$ and $\omega_t := \Lambda_{12,t}\Lambda_{22,t}^{-1}$, while

$$\phi_{1,n}\left(u_s,u_t\right) := w_t\varepsilon_t\varepsilon_u\left[2K_{t,u} - \left(K*K\right)_{t,u}\right]\bar{X}_t'\bar{\Lambda}_t^{-1}\bar{X}_t. \tag{A.51}$$

It now follows by Lemma B.9 that

$$\frac{\bar{F}_n - \mu_n^F}{\nu_n^F} \simeq \frac{n^{-1}\sum_{s\neq t}\phi_{1,n}\left(u_s,u_t\right)}{V_{1,n}} \to^d N\left(0,1\right), \tag{A.52}$$

Finally, we demonstrate that the estimation of w_t does not affect the result. We write $F_n - \mu_n^F = \{\bar{F}_n - \mu_n^F\} + \{F_n - \bar{F}_n\}$, where

$$F_n - \bar{F}_n = \frac{\tilde{F}_n(\hat{w} - w)}{2\int_0^1 w\left(\tau\right)\sigma^2\left(\tau\right)d\tau + o_P\left(1\right)}, \quad \tilde{F}_n(f) = SSR^f(\tilde{\beta}) - SSR^f(\hat{\beta}), \tag{A.53}$$

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 $f : [0,1] \mapsto \mathbb{R}$, $\sqrt{h}\left[\tilde{F}_{1,n}(f) - q_n\left(f\right)\right] = O_P\left(1\right)$, where $q_{1,n}\left(f\right) :=$

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

$$\sqrt{h}\left[\tilde{F}_n(f_1) - q_n(f_1)\right] - \sqrt{h}\left[\tilde{F}_n(f_2) - q_n(f_2)\right] = \sqrt{h}\left[\tilde{F}_n(f_1 - f_2) - q_n(f_1 - f_2)\right], \tag{A.54}$$

where, by using the same arguments as above

$$E\left[\left|\left[\tilde{F}_n(f_1 - f_2) - q_n(f_1 - f_2)\right]\right|^2\right] \leq \frac{C}{h}\sup_{\tau\in[0,1]}\left|f_1(\tau) - f_2(\tau)\right|^2. \tag{A.55}$$

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

$$\frac{\left|F_n - \bar{F}_n\right|}{\sqrt{\nu_n^F}} \leq C\sqrt{h}\left|\tilde{F}_n(\hat{w} - w) - q_n(\hat{w} - w)\right| + \sqrt{h}\left|q_n(\hat{w} - w)\right| \tag{A.56}$$

$$\leq \sup_{\|f\|_\infty \leq C\sqrt{h}} \frac{\left|\tilde{F}_n(f) - q_n(f)\right|}{\sqrt{\nu_n^F}} + \sqrt{h}\left|q_n(\hat{w} - w)\right| = o_P(1). \tag{A.57}$$

□

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

$$\bar{W}_n = \sum_{t=1}^n \mathbb{I}_t(a)\left(\{\tilde{\beta}_1^w - \beta_1\} - \{\hat{\beta}_{1,t} - \beta_1\}\right)'\Omega_t\left(\{\tilde{\beta}_1^w - \beta_1\} - \{\hat{\beta}_{1,t} - \beta_1\}\right) \tag{A.58}$$

$$= \sum_{t=1}^n \mathbb{I}_t(a)\,(\hat{\beta}_{1,t} - \beta_1)'\Omega_t(\hat{\beta}_{1,t} - \beta_1) + \sum_{t=1}^n \mathbb{I}_t(a)\,(\tilde{\beta}_1^w - \beta_1)'\Omega_t(\tilde{\beta}_1^w - \beta_1) \tag{A.59}$$

$$-2\sum_{t=1}^n \mathbb{I}_t(a)\,(\tilde{\beta}_1^w - \beta_1)'\Omega_t(\hat{\beta}_{1,t} - \beta_1) \tag{A.60}$$

$$=: \bar{W}_{1,n} + \bar{W}_{2,n} + \bar{W}_{3,n}. \tag{A.61}$$

The second and third term satisfy

$$\bar{W}_{2,n} = \sqrt{n}(\tilde{\beta}_1^w - \beta_1)'\left\{\frac{1}{n}\sum_{t=1}^n \mathbb{I}_t(a)\,\Omega_t\right\}\sqrt{n}(\tilde{\beta}_1^w - \beta_1) = O_P(1), \tag{A.62}$$

$$\left|\bar{W}_{3,n}\right| \leq 2\sqrt{n}||\tilde{\beta}_1^w - \beta_1||\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^n \mathbb{I}_t(a)\,\Omega_t(\hat{\beta}_{1,t} - \beta_1)\right\| \tag{A.63}$$

where,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n \mathbb{I}_t(a)\,\Omega_t(\hat{\beta}_{1,t} - \beta_1) \simeq \frac{1}{\sqrt{n}}\sum_{t=1}^n \mathbb{I}_t(a)\,\Omega_t\Lambda_t^{-1}X_t'\varepsilon_t + O_P\left(\sqrt{n}h^r\right) = O_P(1) + O_P\left(\sqrt{n}h^r\right). \tag{A.64}$$

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

$$\bar{W}_{1,n} \simeq \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\sum_{u=1}^{n}\mathbb{I}_t(a)\,\varepsilon_s X_{1,s}' K_{s,t}\Lambda_{11,t}^{-1}\Omega_t\Lambda_{11,t}^{-1}K_{t,u}X_{1,u}\varepsilon_u \tag{A.65}$$

$$\simeq \frac{1}{n^2}\sum_{t=1}^{n}\sum_{u=1}^{n}\mathbb{I}_t(a)\,\varepsilon_u^2 X_{1,u}'\Lambda_{11,t}^{-1}\Omega_t\Lambda_{11,t}^{-1}K_{t,u}^2 X_{1,u} \tag{A.66}$$

$$+\frac{1}{n^2}\sum_{s\neq u}\varepsilon_s X_{1,s}'\sum_{t=1}^{n}\left\{\mathbb{I}_t(a)\,K_{s,t}\Lambda_{11,t}^{-1}\Omega_t\Lambda_{11,t}^{-1}K_{t,u}\right\}X_{1,u}\varepsilon_u \tag{A.67}$$

$$\simeq \mu_n^W + \frac{1}{n}\sum_{s\neq t}\phi_{2,n}(u_s,u_t), \tag{A.68}$$

where

$$\phi_{2,n}(u_s,u_t) := \varepsilon_s X_{1,s}'\Lambda_{11,s}^{-1}\Omega_s\Lambda_{11,s}^{-1}(K*K)_{s,t}X_{1,t}\varepsilon_t. \tag{A.69}$$

It now follows by Lemma B.9 that

$$\frac{\bar{W}_n-\mu_n^W}{\sqrt{\nu_n^F}} \simeq \frac{n^{-1}\sum_{s\neq u}\phi_{2,n}(W_s,W_u)}{V_{2,n}} \to^d N(0,1). \tag{A.70}$$

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. □

APPENDIX B: LEMMAS

The following lemmas are used in the proofs of the main theorems. Proofs of the results are available from the author upon request.

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

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

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

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

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

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

Lemma B.7. *With $SSR^w(\beta)$ defined in eq. (2.12): $SSR^w(\hat{\beta})/n = \int_0^1 w(s)\,\sigma^2(s)\,ds + o_P(1)$ as $a,h\to 0$ and $\log(h)/(nh)\to\infty$.*

LEMMA B.8. *The term $\Delta SSR_{2,21}$ defined in eq. (A.42) satisfies*

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

LEMMA B.9. *With $\phi_{1,n}(u_s, u_t)$ and $\phi_{2,n}(u_s, u_t)$ defined in eq. (A.51) and (A.69):*

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

where $V_{1,n} = \frac{m_1}{h}\int w^2(\tau)\sigma^4(\tau)\,d\tau \times \left\|K - \frac{1}{2}(K * K)\right\|^2$ *and*

$$V_{2,n} = \frac{2}{h}\int \sigma^4(\tau)\,tr\left\{\Omega(\tau)\Lambda_{11}^{-1}(\tau)\Omega(\tau)\Lambda_{11}^{-1}(\tau)\right\}d\tau \times \|K * K\|^2. \tag{B.3}$$

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 B.10. *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 for any sequence $a \to 0$ satisfying $h/a \to 0$:*

$$\sup_{a\leq\tau\leq 1-a}|\hat{m}(\tau) - m(\tau)| = O_P(h^r) + O_P\left(\sqrt{\log(n)}/\sqrt{(nh)}\right) \tag{B.4}$$

LEMMA B.11. *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:*

$$\frac{1}{n^2}\sum_{s,t=1}^n \phi_n(u_{n,s}, u_{n,t}) = \theta_n + \frac{2}{n}\sum_{t=1}^n\left[\bar{\phi}_n(u_{n,t}) - \theta_n\right] + R_n, \tag{B.5}$$

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

LEMMA B.12. *For any function ϕ with $E\left[|\phi(u_{n,s}, u_{n,t})|^{1+\delta}\right] < \infty$:*

$$\left|E[\phi(u_{n,s}, u_{n,t})] - E\left[\phi\left(u_{n,s}^*, u_{n,t}^*\right)\right]\right| \leq 4\max\{M_{n,1}, M_{n,2}\}\,|\beta_n(|s-t|)|^{\delta/(1+\delta)}, \tag{B.6}$$

where $u_{n,t}^*$ is an independent sequence with same marginal distribution as $u_{n,t}$, $M_{n,1} = E\left[|\phi(u_{n,s}, u_{n,t})|^{1+\delta}\right]$ and $M_{n,2} = E\left[\left|\phi\left(u_{n,s}^*, u_{n,t}^*\right)\right|^{1+\delta}\right]$.

LEMMA B.13. *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 $\delta > 0$, $n^{-1-\delta/2}\sum_{t=1}^n E\left[|u_{n,t}|^{2+\delta}\right] \to 0$. Then $\sum_{t=1}^n u_{n,t}/\sqrt{n} \to^P N(0, \sigma^2)$.*