Time-Varying Instrumental Variable Estimation *

Liudas Giraitis[†] George Kapetanios [‡§] Massimiliano Marcellino[¶]

September 1, 2020

Abstract

We develop non-parametric instrumental variable estimation and inferential theory for econometric models with possibly endogenous regressors whose coefficients can vary over time either deterministically or stochastically, and the time-varying and uniform versions of the standard Hausman exogeneity test. After deriving the asymptotic properties of the proposed procedures, we assess their finite sample performance by means of a set of Monte Carlo experiments, and illustrate their application by means of an empirical example on the Phillips curve.

JEL classification: C14, C26, C51
Keywords: Instrumental variables, Time-varying parameters, endogeneity, Hausman test, Non-parametric methods, Phillips curve

*We would like to the thank the Editor, Oliver Linton, the Associate Editor and two Referees for very constructive comments on an earlier version of the paper.

[†]Queen Mary University of London
[‡]King's College, London
[§]Corresponding author.
Email address: l george.kapetanios@kcl.ac.uk (G. Kapetanios).
[¶]Bocconi University, IGIER and CEPR

1 Introduction

The investigation of structural change in econometric models has received increasing attention in the literature over the past couple of decades. This development is not surprising. Assuming, wrongly, that the model structure remains fixed over time has clear adverse implications, such as inconsistency of parameter estimators and associated test-statistics, and major forecast failures.

Various methods have been proposed to identify and handle structural change. In the early contributions, changes were supposed to be deterministic, to occur rarely, and to be abrupt. A number of tests for the presence of parameter breaks of that form exist in the literature, starting with the ground-breaking work of Chow (1960), who assumed knowledge of the point in time at which the structural change occurred. The usual fix was the inclusion of proper dummy variables as additional explanatory variables. Other tests relax the underlying assumptions, e.g., Brown, Durbin, and Evans (1974), Ploberger and Kramer (1992) and many others. In this context, it is worth noting that little is being said about the cause of structural breaks in either statistical or economic terms. The work by Kapetanios and Tzavalis (2010) provides a possible avenue for modelling structural breaks and, thus, addresses partially this issue.

A more recent strand of the literature takes an alternative approach and allows the coefficients of parametric models to evolve randomly over time. Evolution can be either discrete, as in Markov Switching models (e.g., Hamilton (1989)) or threshold models (e.g., Tong (1990)), or continuous. In turn, continuous parameter evolution can be driven either by observable variables, as in smooth transition models (e.g., Terasvirta (1998)), or by unobservable shocks combined with time series models for the parameters, as in random coefficient models (e.g., Nyblom (1989)).

Even more recently, there has been a growing interest in random coefficient models that also allow for stochastic volatility, e.g., Cogley and Sargent (2005) or Primiceri (2005), who study the question of whether it was changes in coefficients or in the variance of shocks - policy or otherwise - that gave rise to the period of macroeconomic calmness ("Great Moderation") after 1985. In these papers, and a vast amount of subsequent work, parameters typically evolve as random walks or autoregressive processes.

Yet another strand of the vast structural change literature returns to the assumption of deterministic breaks but allows for a smooth evolution of the parameters. Such modelling attempts have a long pedigree in statistics, starting with the work of Priestley (1965). Priestley's paper suggested that processes may have time-varying spectral densities which change slowly over time. The context of such modelling is nonparametric and has, more recently, been followed up by Robinson (1989), Robinson (1991), Dahlhaus (1997), Chen and Hong (2012) and others, some of whom refer to such processes as locally stationary processes. This approach, while popular in statistics, has not really been influential in applied macroeconometrics where, as mentioned, random coefficient models dominate. Kapetanios and Yates (2008) is an exception, using this type of models to revisit the study of the evolution of inflation persistence, in Cogley and Sargent (2005).

Finally, it is worth noting the work of Muller and Watson (2008) and Muller and Petalas (2010), who also examine structural change and consider both deterministic and stochastic time-varying parameters. While both approaches can be used for the same modelling purposes, the underlying models have very distinct properties. Building on this work, Giraitis, Kapetanios, and Yates (2014) (GKY) have developed a framework for the estimation of random coefficient models using kernel methods. They have provided theoretical, Monte Carlo and empirical results that justify their estimation method and showed that it performs very well in small samples and has trivial computational cost.

Turning now to instrumental variable (IV) regression, it is clearly a major tool for the econometric analysis of many economic phenomena. However, a usual assumption made when carrying out IV regression on time series, or panel, data is that the entertained model is constant over time, implying that the parameter vector to be estimated does not change during the sample period used for estimation. This assumption is both crucial for the properties of IV regression and likely to be suspect in many cases. For example, the impact of education on GDP growth, or of banks' size on their profits, can change over time as a consequence of technological progress. Or the effects of a real forcing variable on inflation can depend on the business cycle phase. Moreover, the relationship between endogenous variables and instruments can also change over time. For example, an autoregressive model that links the current value of a real variable to its past values (the instruments) can exhibit instability over time. Finally, the endogeneity status of a variable could be also time-varying. Consider, for example, a regression of inflation on a short term interest rate over a sample where the central bank switched from exchange rate pegging to inflation targeting.

The problem of modifying IV regression to account for the possible presence of structural change has received limited attention in the literature. Hall, Han, and Boldea (2012) have developed inferential and estimation theory for regressions with endogenous regressors by extending the structural break framework of Bai and Perron (1998). In further work, Boldea and Hall (2013) and Antoine and Boldea (2018) have addressed a variety of other issues associated with this problem. While this work is useful, the structural break framework may be considered restrictive, in light of the above review of time-varying models. On the other hand, specific parametric assumptions on the model driving parameter evolution can be restrictive as well. A major contribution that partially resolves this issue is that of Chen (2015) who extends the deterministic locally stationary framework discussed above to the IV case. However, the deterministic nature of the assumed structural change remains an issue for economic and financial data.

Therefore, in this paper we propose non-parametric, kernel-based, estimation and inferential theory for time-varying IV regression, with either deterministic or random coefficients, extending the framework of GKY. We derive asymptotic distributions for time-varying IV (TV-IV) estimators, which turn out to be asymptotically equivalent but can differ in finite samples.

We also derive the asymptotic distribution of a time-varying version of the Hausman exogeneity test, which compares time-varying OLS and IV estimators, possibly also allowing for changes in the endogeneity status of the regressors over time. Besides a local Hausman test, we develop a uniform test for exogeneity with nice asymptotic properties.

As it is well known that IV estimators can have unpleasant finite sample properties, we evaluate bias and variance related measures for our time-varying IV estimators in an extensive Monte Carlo study, also in comparison with time-varying OLS. The results are rather encouraging, and can be also used to provide indications on the choice of the kernel bandwidth for empirical applications.

Finally, to illustrate in practice the use of time-varying IV, we estimate a simple Phillips curve for the USA, using unemployment as a forcing variable for inflation, as this topic has attracted quite a lot of attention in the academic literature (see, e.g., Stock and Watson (1989)) and it is also at the center of the policy debate, to understand whether decreases in the unemployment rate will eventually increase inflation and therefore require a tightening in monetary conditions. Interestingly, we find substantial fluctuations in the coefficient of unemployment, which remains negative over the entire sample but less so in the recent period, and statistically significant only since the early '90s. Moreover, the time-varying Hausman test suggests that unemployment was endogenous until the end of the 1970s and for a few years around 2000, while exogeneity is not rejected in the most recent period.

The rest of the paper is structured as follows. Section 2 describes our time-varying IV methods and derives the theoretical results. Section 3 discusses the Monte Carlo results. Section 4 presents the empirical application. Section 5 summarizes the main results and concludes. All proofs are relegated to an Appendix and an Online Supplement.

2 Theory

GKY introduced a non-parametric time-varying OLS estimation method that can be applied to a wide set of models, as it boils down to a kernel based generalisation of a rolling window. The main innovation of their work is to show the asymptotic validity of the estimation and the provision of confidence bands for the estimators in the presence of time-varying coefficients, which evolve as persistent stochastic processes. We would like to provide similar results in the IV regression context.

Let us consider the following model for univariate variable, y_t:

$$y_t = x_t'\beta_t + u_t, \quad t = 1, ..., T \tag{1}$$

$$x_t = \Psi_t' z_t + v_t, \tag{2}$$

where $x_t = (x_{1,t}, .., x_{p,t})'$ is a $p \times 1$ vector of random variables, $\beta_t = (\beta_{1,t}, .., \beta_{p,t})'$ is a $p \times 1$ parameter vector, u_t is a 1×1 noise. Subsequently, in (2), $z_t = (z_{1,t}, ..., z_{n,t})'$ is a $n \times 1$ vector of random variables, $\Psi_t' = (\psi_{\ell k,t})$ is a $p \times n$ parameter matrix and $v_t = (v_{1,t}, .., v_{p,t})'$ is a $p \times 1$ noise vector.

As in the standard IV setting, we assume that endogenous variables x_t are correlated with u_t but there exist exogenous instruments z_t, uncorrelated with u_t and v_t:

$$Ez_t u_t = 0, \quad Ez_t v_t' = 0, \quad t \geq 1. \tag{3}$$

In this paper we discuss the IV setting (1)-(2) with time varying parameters β_t and Ψ_t' whose elements are either smoothly varying deterministic functions as in Assumption 2 or smoothly varying persistent stochastic processes as in Assumption 3.[1] In addition, we assume that the elements $z_{\ell,t} z_{k,t}$ of $z_t z_t'$, elements $z_{\ell,t} v_{k,t}$ of $z_t v_t'$ and elements $z_{\ell,t} u_{k,t}$ of $z_t u_t$ are α-mixing variables as specified in Assumption 1 below.

The objective of this paper is to construct consistent estimators of β_t and Ψ_t and derive asymptotic normality for the estimator of β_t.

Our main estimator of β_t is a kernel type estimate

$$\widetilde{\beta}_{1,t} = \Big(\sum_{j=1}^{T} b_{H,|j-t|} \widehat{\Psi}_j' z_j x_j'\Big)^{-1} \Big(\sum_{j=1}^{T} b_{H,|j-t|} \widehat{\Psi}_j' z_j y_j\Big) \tag{4}$$

[1] Of course, if the coefficients are actually constant then the time varying estimator we propose is still consistent albeit inefficient. Similarly, if some variables are actually exogenous then estimation using the IV estimator is again consistent but inefficient.

computed with kernel weights $b_{H,|j-t|}$ and bandwidth parameter H defined below in (6) where $\widehat{\Psi}_j$ is a consistent estimate of Ψ_j. In our case, $\widehat{\Psi}_t$ is just the kernel OLS estimator

$$\widehat{\Psi}_t = \Big(\sum_{j=1}^{T} b_{L,|j-t|} z_j z_j'\Big)^{-1}\Big(\sum_{j=1}^{T} b_{L,|j-t|} z_j x_j'\Big). \tag{5}$$

The bandwidth L used in (5) can be different from H, used for (4).

We consider the estimators (4) and (5) with kernel weights

$$b_{H,|j-t|} = K(\frac{|j-t|}{H}) \tag{6}$$

where $H \to \infty$, $H = o(T)$ is the bandwidth parameter and $K(x)$, $x \in (0,a)$ is a non-negative continuous function with a finite or infinite support such that for some $C > 0$ and $\nu > 3$,

$$K(x) \le C(1+x^{\nu})^{-1}, \quad |(d/dx)K(x)| \le C(1+x^{\nu})^{-1}, \quad x \in (0,a). \tag{7}$$

Standard examples of functions satisfying (7) are $K(x) = I(0 \le x \le 1)$, $K(x) \le C(1+x^{\nu})^{-1}$ with $\nu > 3$ and $K(x) = \exp(-cx^{\alpha})$ with $c > 0$ and $\alpha > 0$.

In addition to $\widetilde{\beta}_{1,t}$, we study the estimator

$$\widetilde{\beta}_{2,t} = \Big(\sum_{j=1}^{T} b_{H,|j-t|} \widehat{\Psi}_t' z_j x_j'\Big)^{-1}\Big(\sum_{j=1}^{T} b_{H,|j-t|} \widehat{\Psi}_t' z_j y_j\Big), \tag{8}$$

which, for $n = p$, and if $\widehat{\Psi}_t$ is full rank, further simplifies asymptotically to

$$\Big(\sum_{j=1}^{T} b_{H,|j-t|} z_j x_j'\Big)^{-1}\Big(\sum_{j=1}^{T} b_{H,|j-t|} z_j y_j\Big). \tag{9}$$

The estimators $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ are asymptotically equivalent. There are other possible estimators such as

$$\widetilde{\beta}_t = \Big(\big(\sum_{j=1}^{T} b_{H,|j-t|} x_j z_j'\big)\big(\sum_{j=1}^{T} b_{H,|j-t|} z_j z_j'\big)^{-1}\big(\sum_{j=1}^{T} b_{H,|j-t|} z_j x_j'\big)\Big)^{-1}$$
$$\times\big(\sum_{j=1}^{T} b_{H,|j-t|} x_j z_j'\big)\big(\sum_{j=1}^{T} b_{H,|j-t|} z_j z_j'\big)^{-1}\big(\sum_{j=1}^{T} b_{H,|j-t|} z_j y_j\big)$$

which is intuitive, in that it extends the standard IV estimator, based on covariances between regressors and instruments, to the time varying case. In the just-identified case $\widetilde{\beta}_t$ asymptotically coincides with $\widetilde{\beta}_{1,t}$, if $H = L$ is imposed. Another potential estimator is the two-stage least square estimator (2SLS) used by Chen (2015):

$$\widetilde{\beta}_{3,t} = \Big(\sum_{j=1}^{T} b_{H,|j-t|}\widehat{\Psi}'_j z_j z'_j \widehat{\Psi}_j\Big)^{-1}\Big(\sum_{j=1}^{T} b_{H,|j-t|}\widehat{\Psi}'_j z_j y_j\Big) \tag{10}$$

which is frequently used in applied work. Both $\widetilde{\beta}_t$ and $\widetilde{\beta}_{3,t}$ are asymptotically equivalent to $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$. Neither will be considered further in detail, in this paper. However, we include some comments on $\widetilde{\beta}_{3,t}$ in Lemma 2 and in the context of Theorem 4 showing that $\widetilde{\beta}_{1,t}$ might be preferable with respect to the 2SLS estimate $\widetilde{\beta}_{3,t}$.

We focus mainly on $\widetilde{\beta}_{1,t}$ for theoretical tractability. We note these different estimators to illustrate the possibilities for different procedures that arise from considering a time varying setting, as well as to emphasise that, in small samples, there may be materially important choices to make, as we explore in our Monte Carlo study, where we compare the finite sample performance of our estimators to that of the OLS estimator given by

$$\widehat{\beta}_t = \Big(\sum_{j=1}^{T} b_{H,|j-t|} x_j x'_j\Big)^{-1}\Big(\sum_{j=1}^{T} b_{H,|j-t|} x_j y_j\Big). \tag{11}$$

Next, we outline assumptions on z_t, v_t and u_t and time-varying parameters β_t and Ψ_t.

Assumption 1 *Elements of z_t, v_t and u_t have the following properties.*
(i) There exists $\theta > 8$ such that uniformly over ℓ and t,

$$E|z_{\ell,t}|^\theta,\ E|v_{\ell,t}|^\theta,\ E|u_t|^\theta \le C < \infty. \tag{12}$$

(ii) For any (ℓ, k), $(z_{\ell,t}z_{k,t} - Ez_{\ell,t}z_{k,t})$, $(z_{\ell,t}v_{k,t})$, $(z_{\ell,t}u_t)$, $(v_{k,t})$ and (u_t) are α-mixing (but not necessarily stationary) processes with mixing coefficients α_k such that for some $0 < \phi < 1$ and $c > 0$,

$$\alpha_k \le c\phi^k, \quad k \ge 1. \tag{13}$$

(iii) The matrices $\Sigma_{zz,t} = E[z_t z'_t]$ and $\Sigma_{vv,t} = E[v_t v'_t]$ are such that $\max_{t\ge1} ||\Sigma^{-1}_{zz,t}||_{sp} < \infty$, $\max_{t\ge1} ||\Sigma^{-1}_{vv,t}||_{sp} < \infty$.

We denote by $||A||_{sp}$ and $||A||$ the spectral and Frobenius norm of matrix A, respectively.

We assume that $\beta_t = \beta_{T,t}$ and $\Psi_t = \Psi_{T,t}$ are triangular arrays of vectors and matrices whose elements $(\beta_{\ell,t})$ and $(\psi_{\ell k,t})$ satisfy either Assumption 2 or Assumption 3.

Assumption 2 *$(\psi_{\ell k,t})$ and $(\beta_{\ell,t})$ are non-random sequences of real numbers uniformly bounded in t and satisfy the smoothness condition*

$$|\beta_{\ell,t} - \beta_{\ell,s}| \le C\frac{|t-s|}{T}, \quad |\psi_{\ell k,t} - \psi_{\ell k,s}| \le C\frac{|t-s|}{T}, \quad t, s, = 1, ..., T \tag{14}$$

where the positive constant, C, does not depend on ℓ, k, t, s and T.

Assumption 3 *$(\psi_{\ell k,t})$ and $(\beta_{\ell,t})$ are random processes that satisfy the smoothness condition*

$$|\beta_{\ell,t} - \beta_{\ell,s}| \le (\frac{|t-s|}{T})^{1/2} r_{\ell,ts}, \quad |\psi_{\ell k,t} - \psi_{\ell k,s}| \le (\frac{|t-s|}{T})^{1/2} q_{\ell k,ts}, \quad t, s, = 1, ..., T \tag{15}$$

and the distribution of variables $X = \beta_{\ell,t}$, $r_{\ell,ts}$, $\psi_{\ell k,t}$, $q_{\ell k,ts}$ has a thin tail:

$$P(|X| \ge \omega) \le \exp(-c_0|\omega|^{\alpha}), \quad \omega > 0 \tag{16}$$

for some $c_0 > 0$, $\alpha > 0$ that do not depend on ℓ, k, t, s and T.

For example, a triangular array of deterministic parameters $\beta_{\ell,t} = g_\ell(t/T)$, $t = 1, ..., T$ where $g_\ell(x)$, $x \in [0, 1]$ have bounded derivatives, satisfy Assumption 2. In turn, we can specify an array of random processes as $\beta_{\ell,t} = T^{-1/2}u_{\ell,t}$, $t = 1, ..., T$ where $u_{\ell,t}$ are random walk processes with $u_{\ell,t} - u_{\ell,t-1} \sim NIID(0, 1)$, satisfying Assumption 3. Many other examples of allowable processes are provided in Giraitis, Kapetanios, and Yates (2014), as well as Section 2.4 of Giraitis, Kapetanios, and Yates (2018). In particular, time varying parameter processes can include deterministic and stochastic components, $\beta_{\ell,t} = T^{-1/2}u_{\ell,t} + g_\ell(t/T)$ and $u_{\ell,t}$ may be a general unit root process rather than random walk. For example, $\beta_{\ell,t} = T^{-1/2}u_{\ell,t}$ satisfies Assumption 3 if $u_{\ell,t}$ is a unit root process, $\omega_t = u_{\ell,t} - u_{\ell,t-1}$ is stationary mixing and has a thin tail distribution. Then $r_{\ell,ts} = (T/(t-s))^{1/2}(\beta_{\ell,t} - \beta_{\ell,s})$ satisfies $P(|r_{\ell,ts}| \ge x) = P((t-s)^{-1/2}|\sum_{j=1}^{t-s}\omega_j| \ge x)$ and has thin tail distribution, see Lemma 1 in Dendramis, Giraitis, and Kapetanios (2018).

Next, we assume that the bandwidth parameters H, and L satisfy

$$c_1 T^{1/(\theta/4-1)+\delta} \le H,\ L \le c_2 T^{1-\delta} \tag{17}$$

where $\theta > 8$ is as in Assumption 1, $c_1, c_2 > 0$ and $\delta > 0$ is arbitrarily small.

Denote

$$r_{T,H} = \sqrt{\frac{\log T}{H}} + \frac{H}{T}, \quad \bar{r}_{T,H,\alpha} = \sqrt{\frac{\log T}{H}} + (\frac{H}{T})^{1/2} \log^{2/\alpha} T. \tag{18}$$

In the next theorem we establish a uniform consistency rate of estimates $\widehat{\Psi}_t$ and $\widetilde{\beta}_{1,t}$. Remarkably, besides the standard IV assumptions $Ez_tu_t = 0$ and $Ez_tv_t' = 0$ it allows (u_t), (v_t), (z_tu_t), (z_tv_t') to be serially correlated sequences.

Theorem 1 *Suppose z_t, u_t and v_t satisfy Assumption 1 and Ψ_t satisfies either Assumption 2 or Assumption 3.*

(i) Then, as $T \to \infty$, the estimator $\widehat{\Psi}_t$ computed with L, satisfying (17), has the property that

$$\max_{t=1,\dots,T} ||\widehat{\Psi}_t - \Psi_t|| = \begin{cases} O_p(r_{T,L}) & \text{if } (\Psi_t) \text{ satisfies Assumption 2,} \quad (19) \\ O_p(\bar{r}_{T,L,\alpha}) & \text{if } (\Psi_t) \text{ satisfies Assumption 3.} \quad (20) \end{cases}$$

(ii) Suppose that the estimator $\widetilde{\beta}_{1,t}$ computed with the bandwidth H uses the estimator $\widehat{\Psi}_t$ computed with the bandwidth L and both H, L satisfy (17). Assume that $\Sigma_{\Psi zz\Psi,t} := \Psi_t'\Sigma_{zz,t}\Psi_t$ has the property

$$\max_{t=1,\dots,T} ||\Sigma^{-1}_{\Psi zz\Psi,t}||_{sp} \le \nu < \infty \quad a.s. \tag{21}$$

where ν does not depend on T. Then,

$$\max_{t=1,\dots,T} ||\widetilde{\beta}_{1,t} - \beta_t|| = \begin{cases} O_p(r_{T,H} + r_{T,L}) & \text{if } (\beta_t, \Psi_t) \text{ satisfy Assumption 2,} \quad (22) \\ O_p(\bar{r}_{T,H,\alpha} \log^{1/\alpha} T + \bar{r}_{T,L,\alpha}) & \text{if } (\beta_t, \Psi_t) \text{ satisfy Assumption 3.} \quad (23) \end{cases}$$

To analyze the asymptotic properties of the estimates $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$, we impose the following assumptions on bandwidth parameters $L \ge H$ used in $\widehat{\Psi}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$. They guarantee that $r_{T,L} = o_P(H^{-1/2})$, $\bar{r}_{T,L,\alpha} = o_P(H^{-1/2})$.

Assumption 4 *The estimator $\widehat{\Psi}_t$ is computed with bandwidth L and $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ with bandwidth H, which have the following properties.*

If (β_t, Ψ_t) satisfy Assumption 2, then $H = o(L/\log T)$, $\quad L = o(T^{2/3})$.

If (β_t, Ψ_t) satisfy Assumption 3, then $H = o(L/(\log T)^{\max(1,4/\alpha)})$, $\quad L = o(T^{1/2})$.

Lemma 2 establishes the main term of the time-varying IV estimator $\widetilde{\beta}_{1,t}$ and its equivalence to $\widetilde{\beta}_{2,t}$ and to the 2SLS estimator $\widetilde{\beta}_{3,t}$. Theorem 3 analyses convergence properties and asymptotic distribution of the IV estimator $\widetilde{\beta}_{1,t}$. The proofs are reported in the Appendix.

Lemma 2 *Suppose that the assumptions of Theorem 1(ii) are satisfied and Assumption 4* *holds. Then, for any sequence $t = t_T \in 1, ..., T$, as $T \to \infty$,*

$$\widetilde{\beta}_{1,t} - \beta_t = \Big(\sum_{j=1}^{T} b_{H,|j-t|}\Sigma_{\Psi zz\Psi,j}\Big)^{-1}\Big(\Psi_t' \sum_{j=1}^{T} b_{H,|j-t|} z_j u_j\Big) + o_P(H^{-1/2}), \tag{24}$$

$$\widetilde{\beta}_{1,t} - \widetilde{\beta}_{2,t} = o_p(H^{-1/2}), \quad \widetilde{\beta}_{1,t} - \widetilde{\beta}_{3,t} = o_p(H^{-1/2}). \tag{25}$$

To obtain the asymptotic normality of $\widetilde{\beta}_{1,t}$ we will use the result (24). We assume for simplicity that the variables $z_j u_j$ are uncorrelated

$$E[z_j u_j z_k' u_k] = 0, \quad k \neq j \tag{26}$$

and impose on the terms of (24) two additional standard assumptions: as $T \to \infty$,

$$K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|}\Sigma_{\Psi zz\Psi,j} \to_p \Sigma_{\Psi zz\Psi,t}, \tag{27}$$

$$K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} z_j u_j \to_D \mathcal{N}(0, \Sigma_{zu,t}) \tag{28}$$

where $K_t = \sum_{j=1}^{T} b_{H,|j-t|}$, $K_{2,t} = \sum_{j=1}^{T} b_{H,|j-t|}^2$ and $\Sigma_{zu,t} = \mathrm{var}(z_t u_t)$ is the variance-covariance matrix of $z_t u_t$ (which may vary with t). Convergence (27) holds under additional smoothness conditions on the change of $\mathrm{var}(z_t)$ similar to those in Assumption 2. Convergence (28) is achieved by imposing standard additional mixing or martingale difference type assumptions on the $n \times 1$ vector $z_t u_t$. For the sake of brevity, we do not provide a more detailed analysis of these convergence relations and the underlying assumptions, but refer the reader to previous work and in particular Giraitis, Kapetanios, and Yates (2014), as well as Theorems 2.2 and 2.3 of Giraitis, Kapetanios, and Yates (2018) and their associated proofs. Denote $\Sigma_{\Psi zu\Psi,t} = \Psi_t' \Sigma_{zu,t} \Psi_t$.

Theorem 3 *Let assumptions of Theorem 1(ii) be satisfied and Assumption 4, (26), (27) and (28) hold. Then, for any sequence $t = t_T \in 1, ..., T$, $\widetilde{\beta}_{1,t}$ has the following asymptotic properties.*

(i) If (β_t, Ψ_t) are deterministic and satisfy Assumption 2, then

$$\frac{K_t}{K_{2,t}^{1/2}} \Sigma_{\Psi zz\Psi,t}(\widetilde{\beta}_{1,t} - \beta_t) \to_D \mathcal{N}(0, \Sigma_{\Psi zu\Psi,t}). \tag{29}$$

(ii) If (β_t, Ψ_t) are random, satisfy Assumption 3, $n = p$, and Ψ_t are invertible, then

$$\frac{K_t}{K_{2,t}^{1/2}} \Psi_t^{-1} \Sigma_{\Psi zz\Psi,t} (\widetilde{\beta}_{1,t} - \beta_t) \to_D \mathcal{N}(0, \Sigma_{zu,t}). \tag{30}$$

The normal approximations of Theorem 3 follow from the approximation (24) of Lemma 2 using (27) and (28). Note that we only report a straightforward asymptotic result for the case $n = p$. Of course, convergence occurs for the more general case but then Ψ_t is not square, and therefore, not invertible, and appears in the limiting distribution, making inference complex. The above inference results can be operationalised by replacing unknown quantities with their time varying estimates.

In her original work, Chen (2015) has introduced the idea of IV estimation and inference for parameter stability in a time-varying setting. We refer to Chen's paper for the background literature and intuition behind the IV estimates. For deterministic parameters the main novelty of our work is establishing uniform consistency rates under similar assumptions to Chen (2015). However, we focus on estimation of the path of parameters β_t, Ψ_t, $t = 1, ..., T$ under conditions on their increments rather then estimating functions $\beta(t/T)$, $\Psi(t/T)$. Imposing smoothness conditions on β_t, Chen (2015) obtains a slightly better optimal rate $o(T^{2/5})$ in normal approximation than ours, $o(T^{1/3})$. Differently from Chen's work, for simplicity, we present asymptotic distribution only in the case of uncorrelated noise $z_j u_j$.

Our paper develops further the ideas of Chen (2015) by showing that time varying IV estimation and inference can be applied to a large class of observables y_t, x_t and instrumental variables z_t. Our setting allows for stochastic time-varying persistent parameters β_t, Ψ_t that may be dependent on (x_t, z_t) and the noise (u_t, v_t). Consequently, a wide class of (non) stationary processes y_t and x_t can be sufficiently well approximated by $\Psi_t' z_t$, $x_t \beta_t'$ so that residual noises u_t and v_t can be expected to be standard mixing processes. Methodologically, this justifies the use of IV estimation and the Hausman test we develop and present below, in empirical work when it is not realistic to impose multiple restrictions on u_t, v_t and on $y_t - u_t, x_t - v_t$. Such restrictions are not required in our work, except for the Hausman test where, under the null hypothesis, we assume independence of (u_t) from $(y_t - u_t, x_t - v_t)$. By introducing such an assumption we aim at simplifying the derivation of the asymptotic distribution of the Hausman type test, which makes it attractive for empirical work. We develop a local test for exogeneity at the point t and a uniform test for exogeneity over period $(T_0, T_1]$.

In Theorem 4 we obtain an asymptotic approximation for the OLS estimator $\widehat{\beta}_t$, which is used for the time-varying version of the Hausman test. In the first part of the theorem, we derive a generic approximation for $\widehat{\beta}_t - \widetilde{\beta}_{1,t}$ that is valid irrespectively of whether x_j is exogenous or endogenous, i.e. when $Ev_ju_j \neq 0$ or $Ev_ju_j = 0$. In the second part of the theorem we derive asymptotic normality for the difference $\widehat{\beta}_t - \widetilde{\beta}_{1,t}$ of the OLS and IV estimates $\widehat{\beta}_t$ and $\widetilde{\beta}_{1,t}$ when x_t are exogenous variables. For simplicity, we assume that the variables u_j are independent.

Assumption 5 *(i)* (u_t) *are independent random variables,* $Eu_t = 0$, $Eu_t^2 = \sigma_{u,t}^2$,
(ii) (u_t) *and* $(x_t, x_t - v_t)$ *are mutually independent,*
(iii) the elements of $\Sigma_{vv,t}$ *and* $\sigma_{u,t}^2$ *satisfy Assumption 2,* $\sigma_{u,t}^2 + \sigma_{u,t}^{-2}$ *is bounded,*
(iv) for endogenous x_j, $E[v_ju_j]$ *satisfies Assumption 2.*

The next theorem focuses on testing for x_t being exogenous, in settings relevant to applied work. It allows unrestricted mutual dependence of $(\beta_t,\ \Psi_t,\ x_t,\ v_t)$.
Set $S_{xx,t} = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}x_jx_j'$, $S_{\widehat{x}\widehat{x},t} = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}\widehat{x}_j\widehat{x}_j'$, $\widehat{x}_j = \widehat{\Psi}_jz_j$, $\Sigma_{xx,t} = \Psi_t'E[z_tz']\Psi_t + E[v_tv_t']$. Denote

$$V_{T,t} = (S_{\widehat{x}\widehat{x},t})^{1/2}(S_{xx,t})^{1/2}(\widehat{\beta}_t - \widetilde{\beta}_{1,t}), \tag{31}$$

$$\mathcal{H}_{T_0,T_1} = \frac{1}{\sqrt{T_1 - T_0}}\sum_{t=T_0+1}^{T_1} K_tK_T^{-1}\sigma_{u,t}^{-1}\Sigma_{vv,t}^{-1/2}V_{T,t}, \quad 0 \leq T_0 < T_1 \leq T. \tag{32}$$

The Hausman type statistic $V_{T,t}$ can be used for testing the null hypothesis $H_0:\ E[v_tu_t] = 0$ that x_j is exogenous at time t while $\mathcal{H}_{T_0,T_1}$ tests the global null hypothesis $H_0:\ E[v_tu_t] = 0,\ t = T_0 + 1, ..., T_1$ that x_j is exogenous over the period $t \in (T_0,\ T_1]$.

Theorem 4 *Let Assumption 1 hold,* (β_t, Ψ_t) *satisfy either Assumption 2 or Assumption 3, L satisfy (17) and*

$$c_1T^e \leq H \leq c_2T^{1-e} \tag{33}$$

for some small $0 < e < 1$ *and* $c_1, c_2 > 0$
Then, for any sequence $t = t_T \in 1, ..., T$*, the following holds.*

(i) For endogenous x_j*, under Assumption 5 (iii)-(iv):*

$$V_{T,t} = (1 - \Sigma_{vv,t}\Sigma_{xx,t}^{-1})^{1/2}E[v_tu_t] + o_p(1). \tag{34}$$

(ii) For exogenous x_j*, assume that* (u_j) *satisfy Assumption 5,* $H = o(T^{2/3})$ *if Assumption*

2 holds and $H = o(T^{1/2})$ if Assumption 3 holds. Then

$$K_t K_{2,t}^{-1/2} \sigma_{u,t}^{-1} \Sigma_{vv,t}^{-1/2} V_{T,t} \to_D \mathcal{N}(0, I). \tag{35}$$

If in addition $\widetilde{T} = T_1 - T_0$, H are such that $H = o(\widetilde{T})$, $H = o(T/\widetilde{T}^{1/2})$ if Assumption 2 holds and $H = o(T/\widetilde{T})$ if Assumption 3 holds, then

$$\mathcal{H}_{T_0,T_1} \to_D \mathcal{N}(0, I). \tag{36}$$

For endogenous x_t, when $E[v_t u_t] \neq 0$, the local statistic at time t diverges at the rate $\sqrt{H}$ while the statistic over period $(T_0, T_1]$ at the rate $\sqrt{T_1 - T_0} > \sqrt{H}$. For $(T_0, T_1] = [1, T]$, (36) requires $H = o(T^{1/2})$ under Assumption 2. If parameters are stochastic, (36) requires $T_1 - T_0 = o(T)$.

Properties (35) and (36) remain valid for $V_{T,t}$, $\mathcal{H}_{T_0,T_1}$ with $\widetilde{\beta}_{1,t}$ replaced by the 2SLS estimator $\widetilde{\beta}_{3,t}$, if Assumption 4 on H, L is also imposed. Hence, when testing for x_j being exogenous, the use of $\widetilde{\beta}_{1,t}$ has some theoretical advantages. The proof of Theorem 4 is relegated to an Online Supplement.

A common formulation of the Hausman (1978) test is as a quadratic form in the difference between the (time-varying in our case) OLS and IV estimators. Theorem 4 can be used to derive the (pointwise and uniform) asymptotic distribution of this version of the time-varying Hausman test. In particular, if $\Sigma_{v,t}^{-1}$ exists, then (35) implies that

$$\frac{K_t^2}{K_{2,t}} V_{T,t}' \widehat{\Sigma}_{\widehat{v}\widehat{v},t}^{-1} V_{T,t} \widehat{\sigma}_{\widehat{u},t}^{-2} \to_D \chi_p^2, \tag{37}$$

because the estimate $\widehat{\Sigma}_{\widehat{v}\widehat{v},t} := K_t^{-1} \sum_{j=1}^T b_{H,|j-t|} \widehat{v}_j \widehat{v}_j'$ and $\widehat{\sigma}_{\widehat{u},t}^2 := K_t^{-1} \sum_{j=1}^T b_{H,|j-t|} \widehat{u}_j^2$ based on residuals $\widehat{u}_j = y_j - x_j' \widetilde{\beta}_{1,j}$, $\widehat{v}_j = x_j - \widehat{\Psi}_j' x_j$ have the property $\widehat{\Sigma}_{\widehat{v}\widehat{v},t} \to_p \Sigma_{vv,t}$, $\widehat{\sigma}_{\widehat{u},t}^2 \to_p \sigma_{u,t}^2$.

The correspondingly modified statistic $\mathcal{H}_{T_0,T_1}$ has property

$$\mathcal{H}_{T_0,T_1}' \mathcal{H}_{T_0,T_1} \to_D \chi_p^2. \tag{38}$$

Of course, all the usual statistics related to IV estimation can be generalised to the time-varying case. For example, a time-varying version of the J-test statistic is given by

$$J_t = \frac{K_t}{K_{2,t}} \frac{1}{\widehat{\sigma}_{\widehat{t}}^2} \Big(\sum_{j=1}^T b_{H,|j-t|} z_j' \widehat{u}_j \Big) \Big(\sum_{j=1}^T b_{H,|j-t|} z_j z_j' \Big)^{-1} \Big(\sum_{j=1}^T b_{H,|j-t|} z_j \widehat{u}_j \Big). \tag{39}$$

It can be easily shown that, for each t, J_t follows asymptotically a χ^2_p distribution, under the null hypothesis of valid overidentifying restrictions. Obtaining a test for exogeneity based on $\max_{t\in[T_0,T_1]} |V_{T,t}|$ in line with Chen and Hong (2012) test for parameter stability is an interesting open problem.

3 Monte Carlo study

In this Section we evaluate the finite sample performance of the time-varying IV estimators $\widetilde{\beta}_{1,t}$, $\widetilde{\beta}_{2,t}$, the OLS estimator $\widehat{\beta}_t$ and the time-varying Hausman test.

As data generating process (DGP), we consider model (1) and (2). Our baseline case is the exactly identified model, i.e. $n = p = 1$:

$$y_t = \beta_t x_t + u_t, \quad x_t = \psi_t z_t + v_t, \quad t = 1, ..., T. \tag{40}$$

We introduce correlation between u_t and v_t by specifying them as

$$u_t = se_{1,t} + (1-s)e_{2,t}, \qquad v_t = se_{1,t} + (1-s)e_{3,t}, \tag{41}$$

where $s = 0, 0.2, 0.5$ and $(e_{1,t})$, $(e_{2,t})$ and $(e_{3,t})$ are mutually independent $NIID(0,1)$ sequences.

The parameters $\beta_t = T^{-1/2}\xi_{1,t}$, $\psi_t = T^{-1/2}\xi_{2,t}$, $t = 1, ..., T$ are generated as two independent rescaled random walks, such that $\xi_{\ell,t} - \xi_{\ell,t-1} \sim NIID(0,1)$ for $\ell = 1, 2$. This implies that both the structural and the reduced form regressions have time-varying coefficients. We assume that (z_t) is a sequence of standard normal i.i.d. random variables independent of (ψ_t), (u_t) and (v_t). Exogeneity of x_t is implied by $s = 0$, while for $s = 0.2, 0.5$, x_t is endogenous. The magnitude of s provides a means for controlling the extent of endogeneity. We also consider the same set of experiments for an overidentified case where

$$y_t = \beta_t x_t + u_t, \quad x_t = \psi_{1,t} z_{1,t} + \psi_{2,t} z_{2,t} + v_t, \quad t = 1, ..., T, \tag{42}$$

where $(\psi_{1,t})$ and $(z_{1,t})$ have the same specification as (ψ_t) and (z_t), $\psi_{2,t} = T^{-1/2}\xi_{3,t}$, $t = 1, ..., T$, where $\xi_{3,t} - \xi_{3,t-1} \sim NIID(0,1)$, and $(z_{2,t})$ is a sequence of standard normal i.i.d. random variables.

We consider three estimators of β: time-varying $\widehat{\beta}_t$ (OLS), time-varying $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ (IV). They are computed using the Gaussian kernel $K(x) = \exp(-x^2/2)$ with a variety of bandwidth values H for estimation of β_t and L for ψ_t. Specifically, we set $H = T^{h_1}$ and $L = T^{h_2}$

with $h_1, h_2 = 0.4$ and 0.5. In an online appendix we also report results for 0.7. Lower values for the bandwidth increase robustness of estimates to parameter changes but decrease efficiency, and it is interesting to evaluate the trade-off. Further, we consider four sample sizes: $T = 100, 200, 400, 1000$.

Next, we proceed with a detailed Monte Carlo analysis, based on 1000 Monte Carlo replications. To assess the performance of our estimators, we use a variety of performance indicators that are of relevance for IV regression, where estimator variances may not be finite in small samples. We consider the following measures of performance: average over all the replications of the median deviation and of the absolute median deviation

$$p^{-1} \sum_{r=1}^{p} med_{t=1,\dots,T}(\widetilde{\beta}_{r,t} - \beta_{r,t}), \quad p^{-1} \sum_{r=1}^{p} med_{t=1,\dots,T}|\widetilde{\beta}_{r,t} - \beta_{r,t}|, \quad p = 1000,$$

average over all the periods of the interdecile range and of the 95% coverage rates

$$T^{-1} \textstyle\sum_{t=1}^{T} (\widetilde{\beta}_{t,90\%} - \widetilde{\beta}_{t,10\%}), \quad T^{-1} \textstyle\sum_{t=1}^{T} cover_t,$$

where $\widetilde{\beta}_{t,x\%}$ denotes the x-th quantile of the empirical distribution of $\widetilde{\beta}_t$ obtained via Monte Carlo simulations and $cover_t$ is the estimated probability that β_t lies in the interval $(\widetilde{\beta}_t - 1.96 * std(\widetilde{\beta}_t), \widetilde{\beta}_t + 1.96 * std(\widetilde{\beta}_t)$ where $std(\widetilde{\beta}_t)$ is computed using results from the asymptotic distribution of $\widetilde{\beta}_t$. Tables 1-6 report values of these four measures of performance for all experiments. The first three tables relate to the exactly identified case, while the last three present results for the overidentified case.

For the Hausman test, comparing $\widehat{\beta}_t$ and $\widetilde{\beta}_{1,t}$, Table 7 reports rejection probabilities for the just-identified case for the middle point $t = T/2$ which is representative as in the DGP there are no changes of endogeneity status. Table 8 focuses instead on the global Hausman test (calculated over the interval $(T_0, T_1) = (5, T-5)$) for the just-identified case.

Based on Tables 1-3, a number of comments can be made. First, starting with Table 1 where x_t is exogenous ($s = 0$), the median deviations of all the three estimators are very similar and all close to zero. In terms of median absolute deviations, $\widehat{\beta}_t$ is the best performer (as it is indeed best in this context) while $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ are comparable. In all cases the reported values are rather small and decrease when the sample size T increases. About the bandwidth parameters $H = T^{h_1}$, $L = T^{h_2}$, typically values around 0.5 for both h_1 and h_2 yield the lowest values of criteria. However, the differences of those values are rather small unless T is very large ($T = 1000$). Moving to the interdecile range, for $\widehat{\beta}_t$ the key parameters are $H = T^{h_1}$ and T (as $\widehat{\beta}_t$ does not depend on L). Higher values of h_1 generally shorten the ranges, while

T has little effect. The effects of $H = T^{h_1}$ and T on $\widetilde{\beta}_{2,t}$ are similar to those on $\widehat{\beta}_t$, as $\widetilde{\beta}_{2,t}$ does not dependent on ψ_t when $n = p = 1$. However, the interdecile ranges are systematically larger for $\widetilde{\beta}_{2,t}$ than for $\widehat{\beta}_t$, which is expected with $\widehat{\beta}_t$ being appropriate in this context. For $\widetilde{\beta}_{1,t}$, it turns out that $L = T^{h_2}$ matters more than $H = T^{h_1}$, and larger values for h_2 shorten the interdecile range substantially, for any value of h_1. However, the coverage ranges also diminish substantially when h_2 increases, so that intermediate values 0.5 for h_1 and h_2 appear as a reasonable choice. A similar comment also applies for $\widetilde{\beta}_{2,t}$ and $\widehat{\beta}_t$, as in their case the coverage rates decrease when $H = T^{h_1}$ increases.

Second, the effects of endogeneity on the time-varying OLS estimator $\widehat{\beta}_t$ are substantial, the more so the stronger the correlation of x_t with the error term. From Tables 2 ($s = 0.2$) and 3 ($s = 0.5$), the median deviation of $\widehat{\beta}_t$ increases substantially, while that of $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ is much less affected. A similar comment applies for the absolute median deviation and the coverage rates, while the interdecile ranges are not substantially affected. These results stress the importance of using a time-varying IV estimator in the presence of endogeneity.

Moving on to the overidentified case, presented in Tables 4-6, we see that our main conclusions remain largely unaffected. One clear difference is that coverage rates worsen somewhat, compared to the exactly identified case.

Table 7 reports size and power of the time-varying Hausman test for the just-identified case, constructed using $\widetilde{\beta}_{1,t}$, for various values of $H = T^{h_1}$ and $L = T^{h_2}$. From the panel with $s = 0$, where the null hypothesis of exogeneity is valid and the size of the test is reported, it turns out that size distortions can be sizable, even for very large T. However, intermediate values of h_1 and h_2 reduce the size distortions, with actual size close to the nominal one when $h_1 = 0.4, h_2 = 0.5$. In terms of power, it increases with s and with T, but it remains rather low when $h_1 = 0.4, h_2 = 0.5$. Finally, in Table 8 we present the size and power for the global Hausman test setting $T_0 = 5$ and $T_1 = T - 5$. Clearly, it seems to behave as expected both under the null and the alternative hypothesis[2].

4 Empirical Application

In this section we illustrate the use of time-varying IV with an empirical application. We estimate a version of the traditional Phillips curve that links inflation to unemployment,

[2]We have also computed simulations for higher values of the bandwidth parameters. The results are reported in the Online Supplement, where Tables 9-14 are the counterpart of Tables 1-6, and Tables 15-16 of Tables 7-8. In general, the results are as expected, in the sense that the average median deviation is higher and the coverage rate lower, while the decile range is lower. Hence, we get more efficiency but also more bias. For the time-varying Hausman test, in general the size distortions increase, while for the uniform Hausman test the size gets closer to nominal but the power decreases.

based on the idea that lower unemployment leads to higher wages, costs, prices and hence inflation. The main goal is to understand whether unemployment is indeed significant, it is endogenous, and there were changes in these two features over time.

We use monthly data for the USA over the period 1959-2013 and, due to their high persistence, we use the change in inflation ($\Delta\pi$) as dependent variable and the change in unemployment (Δu) as explanatory variable (together with one lag of the change in inflation). Our instruments are four lags of the change in unemployment and one lag of the change in inflation. The model is:

$$\Delta\pi_t = c + \gamma\Delta\pi_{t-1} + \alpha\Delta u_t + e_t, \tag{43}$$

where e_t is a white noise error term.

We compare the results of time-varying OLS and IV estimation. As time-varying IV estimators, we report results for $\widetilde{\beta}_{1,t}$ with a Gaussian kernel, and $H = L = T^{0.7}$ in order to estimate the parameters in each time period with a large enough number of observations (about 90).

Before proceeding, it is worth noting that an LM test for serial correlation with 4 lags fails to reject the null hypothesis that there is no serial correlation in the residuals of the time-varying IV model. Figure 1 provides the results for this model. The upper panel of Figure 1 graphs $\widehat{\gamma}_t$ and $\widetilde{\gamma}_{1,t}$, with the associated 90% confidence bands. It turns out that the estimators are similar and feature a substantial amount of time-variation. The estimated γ parameter is in the interval $[0, 0.5]$, it increases steadily from the '60s until the mid '80s, then decreases until the late '90s and then increases again. Moreover, the estimated parameter is statistically significant in all the periods after the early '70s. The average value over time for $\widehat{\gamma}_t$ is 0.28, which is comparable to the full sample constant parameter OLS value, about 0.36.

The middle panel of Figure 1 graphs $\widehat{\alpha}_t$ and $\widetilde{\alpha}_{1,t}$, with the associated asymptotic 90% confidence bands. The time-varying IV estimator is now quite different from the time-varying OLS, $\widehat{\alpha}_t$. Specifically, $\widetilde{\alpha}_{1,t}$ returns lower values for α in each time period, in the interval $[-1.2, -0.4]$ versus $[-0.4, 0]$ for $\widehat{\alpha}_t$. The average values over time for $\widehat{\alpha}_t$ and $\widetilde{\alpha}_{1,t}$ are about -0.15 and -0.71 respectively, which are comparable to the full sample constant parameters OLS and TSLS values, about -0.18 and -0.93, respectively. Naturally, the confidence intervals are larger for time-varying IV than for time-varying OLS. Yet, the TV-IV estimator for α is statistically significant over most of the sample period, while TV-OLS only from the late '70s to the late '80s. Interestingly from an economic point of view, both time-varying OLS and IV indicate that α is no longer statistically significant in the most recent period, so that the continued decreases in the unemployment rate should not lead to a pick-up in inflation (which indeed is what happened).

The lower panel of Figure 1 graphs the p-value of the time-varying Hausman test. Taking 0.10 as the significance level, the test rejects the null hypothesis of exogeneity from the mid-60s until the late '70s. Probability values again decline considerably for a few years around 2000. From the upper panel of Figure 1, these are indeed the periods when the time-varying OLS and IV estimators for α deviate most. However, from the Monte Carlo experiment we know that the finite sample power of the Hausman test is rather limited, so that exogeneity is likely absent for longer periods, although we also note that for the choice of bandwidth we consider, the size of the test is also likely to suffer. The global Hausman test rejects exogeneity at the 5% level, the p-value is 0.013.

Given the above, as a robustness check, we repeat the analysis using $H = L = T^{0.5}$, which were the preferred values from the Monte Carlo analysis, even if in our empirical application this value for the bandwidth brings down the number of effective observations to estimate the parameters in each time period from about 90 to 25 and hence we expect a substantial increase in uncertainty. The results, reported in Figure 2, are indeed similar to those discussed above except that, as expected, there is a substantial increase in the uncertainty around the time-varying IV estimator, which makes the Hausman test reject in even fewer time periods than before. Yet, the global test still rejects, the p-value is 0.026.

As an additional robustness check, we now estimate a forward looking (New-Keynesian) Phillips curve, along the lines of Gali and Gertler (2008) . The New-Keynesian Phillips curve is specified as:

$$\Delta\pi_t = c + \rho\Delta\pi^e_{t+1} + \gamma\Delta\pi_{t-1} + \alpha\Delta u_t + v_t, \tag{44}$$

which can be also written as

$$\Delta\pi_t = c + \rho\Delta\pi_{t+1} + \gamma\Delta\pi_{t-1} + \alpha\Delta u_t + \epsilon_t, \tag{45}$$

where $\epsilon_t = \rho(\Delta\pi^e_{t+1} - \Delta\pi_{t+1}) + v_t$, $\Delta\pi^e_{t+1}$ is the optimal one-step ahead forecast of $\Delta\pi_{t+1}$ made in period t, and v_t is an i.i.d. error, uncorrelated at all leads and lags with the forecast error $(\Delta\pi_t - \Delta\pi^e_t)$. It is clear that $\Delta\pi_{t+1}$ is correlated with the error term ϵ_t, the more so the less good is the forecast, so that IV is needed. We experiment with the same time-varying IV estimator as above, using four lags of the change in unemployment and inflation as instruments, a Gaussian kernel, and $H = L = T^{0.7}$. For comparison, we also compute the time-varying OLS estimator.

Figure 3 reports the alternative estimators for, respectively, γ, α and ρ. Looking at the first panel of Figure 3, the time-varying OLS estimator for γ is pretty similar to that reported in Figure 1, both in terms of temporal evolution and of values (note that the scales of the two figures are different). The time-varying IV estimator is also pretty similar, except for a period around the mid 70s. It has also a larger variability than in the backward looking case, such that the time-varying OLS confidence intervals are always within those for the time-varying IV. This finding is also true for ρ, see third panel of Figure 3. The same panel of Figure 3 also shows that the time-varying OLS estimator for ρ is statistically significant, while the time-varying IV estimator does not indicate significance of ρ, while it pointed to statistical significance of γ, at least since the mid '70s (see upper panel of Figure 3).

In terms of the coefficient of the change in unemployment, α, according to the second panel of Figure 3 it is close to zero when estimated by time-varying OLS, and never statistically significant, except for a short period in the early '80s. The time-varying IV estimator instead suggests substantially negative values for α but, due to the increased variability of the estimator, only significantly different from zero until the late '70s. Taking again a 10% significant value, the time-varying Hausman test, reported in the lowest panel of Figure 3, rejects exogeneity until the late '70s and again around the mid 90's. The p-value of the global Hausman test is however larger than for the backward looking specifications, 0.155.

In summary, this simple but economically interesting empirical application highlights the relevance of allowing for parameter time variation in an IV setting. In particular, there is a varying but sizable impact of the change of unemployment on the change of inflation, though the estimated parameter shrinks substantially in the final part of the sample, till it becomes not statistically significant. The exogeneity status of unemployment seems to also vary over time according to the time-varying Hausman test, and exogeneity is clearly rejected at the 10% level in the '70s.

5 Conclusions

Instrumental variable regression is extensively applied in econometric studies but the parameters are typically assumed stable over time (or across units). However, the vast literature on structural change has highlighted that parameter instability is diffuse. Hence, in this paper we introduce time-varying IV estimators, taking a non-parametric approach in order to remain as agnostic as possible on the type of parameter evolution.

We derive asymptotic distributions for two time-varying, kernel based, IV estimators, which turn out to be asymptotically equivalent. We also derive the asymptotic distribution of time-varying and uniform versions of the Hausman exogeneity test, which compares time-varying OLS and IV estimators, possibly also allowing for changes in the endogeneity status of the regressors over time.

Next, we evaluate the finite sample properties of the alternative estimators and size and power of the time-varying and uniform Hausman tests in an extensive Monte Carlo study. The results show that the finite sample bias of the estimators is small, the variability limited and the capacity to recover the temporal evolution of the parameters substantial. However, the time-varying Hausman test has rather low power, while the uniform Hausman test performs better, for the same extent of endogeneity.

Finally, to illustrate in practice the use of time-varying IV, we estimate a simple Phillips Curve for the USA, using unemployment as a forcing variable for inflation. We find substantial fluctuations in the coefficient of unemployment, which remains negative over the entire sample but less so in the most recent period. Moreover, the time-varying Hausman test suggests that unemployment was endogenous until the end of the '70s and for a few years around 2000, while exogeneity cannot be rejected in the most recent period and indeed the time-varying IV estimators get closer to time-varying OLS.

Figure 1: Empirical results for model (43). The first two panels graph OLS and IV estimates of γ and α respectively. The third panel graphs the p-value of the time-varying Hausman test.

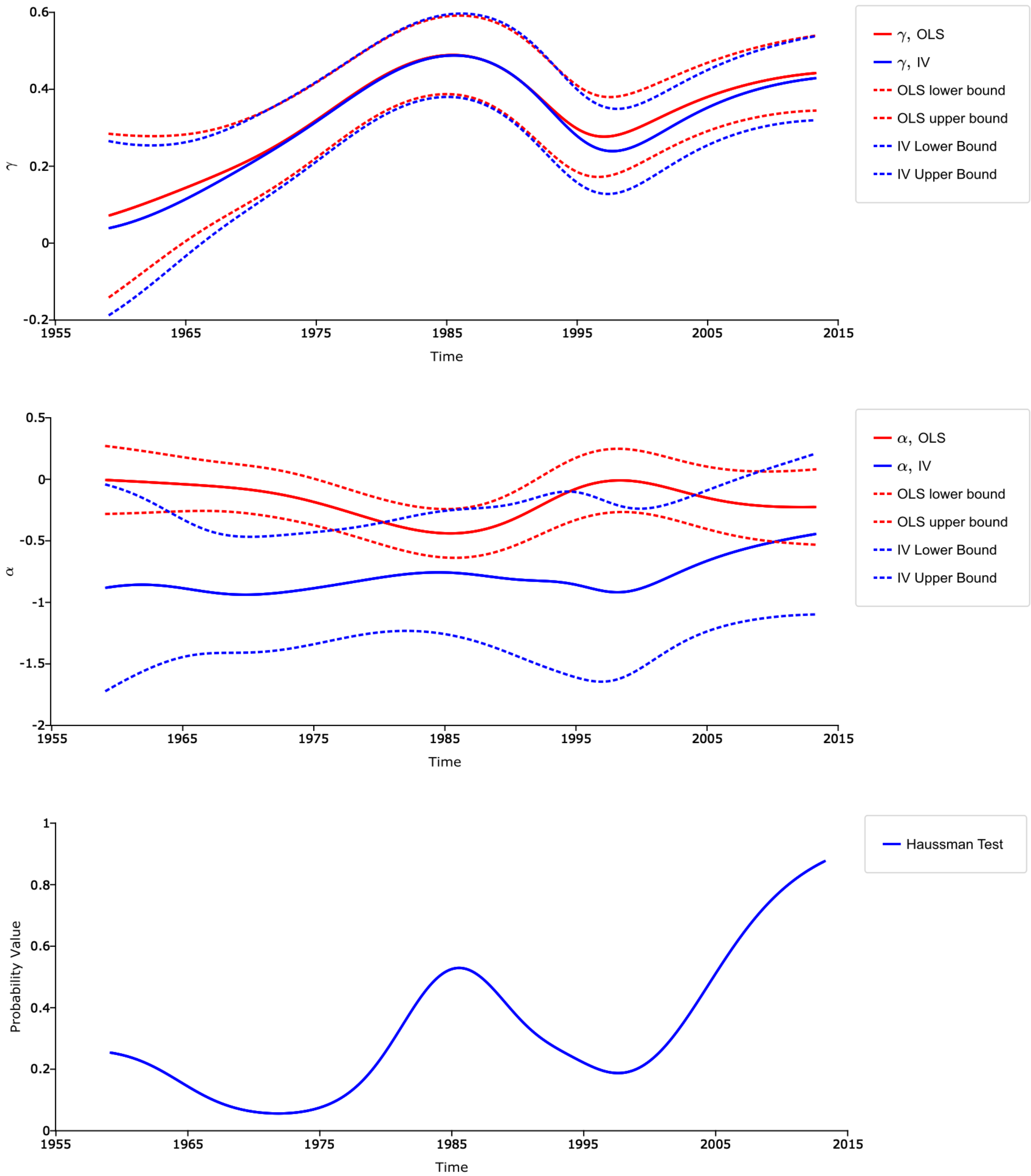

Figure 2: Robustness check for model (43) using $H = L = T^{0.5}$. The first two panels graph OLS and IV estimates of γ and α respectively. The third panel graphs the p-value of the time-varying Hausman test.

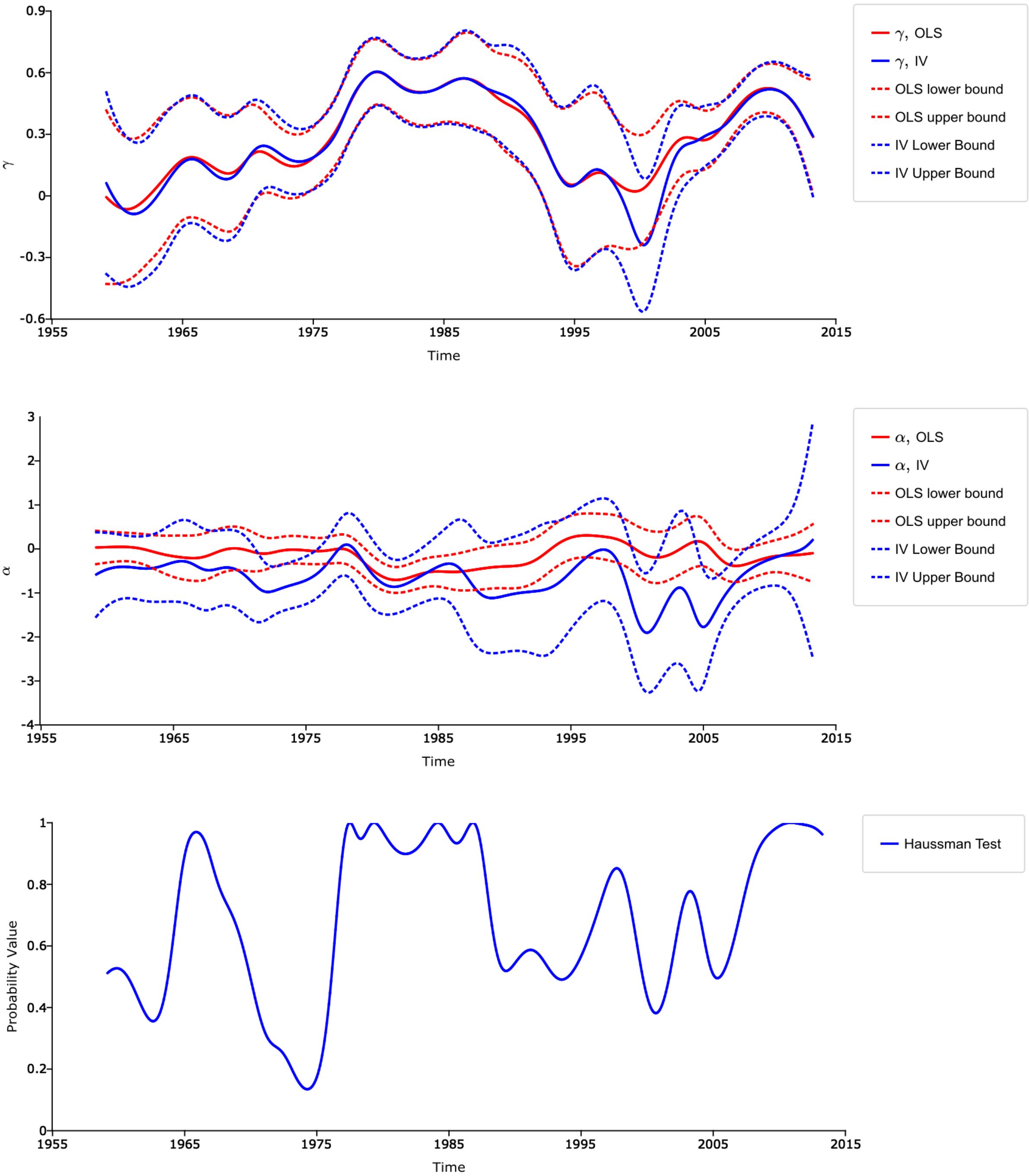

Figure 3: Empirical results for model (45). The first three panels graph OLS and IV estimates of γ, α and ρ respectively. The fourth panel graphs the p-value of the time-varying Hausman test.

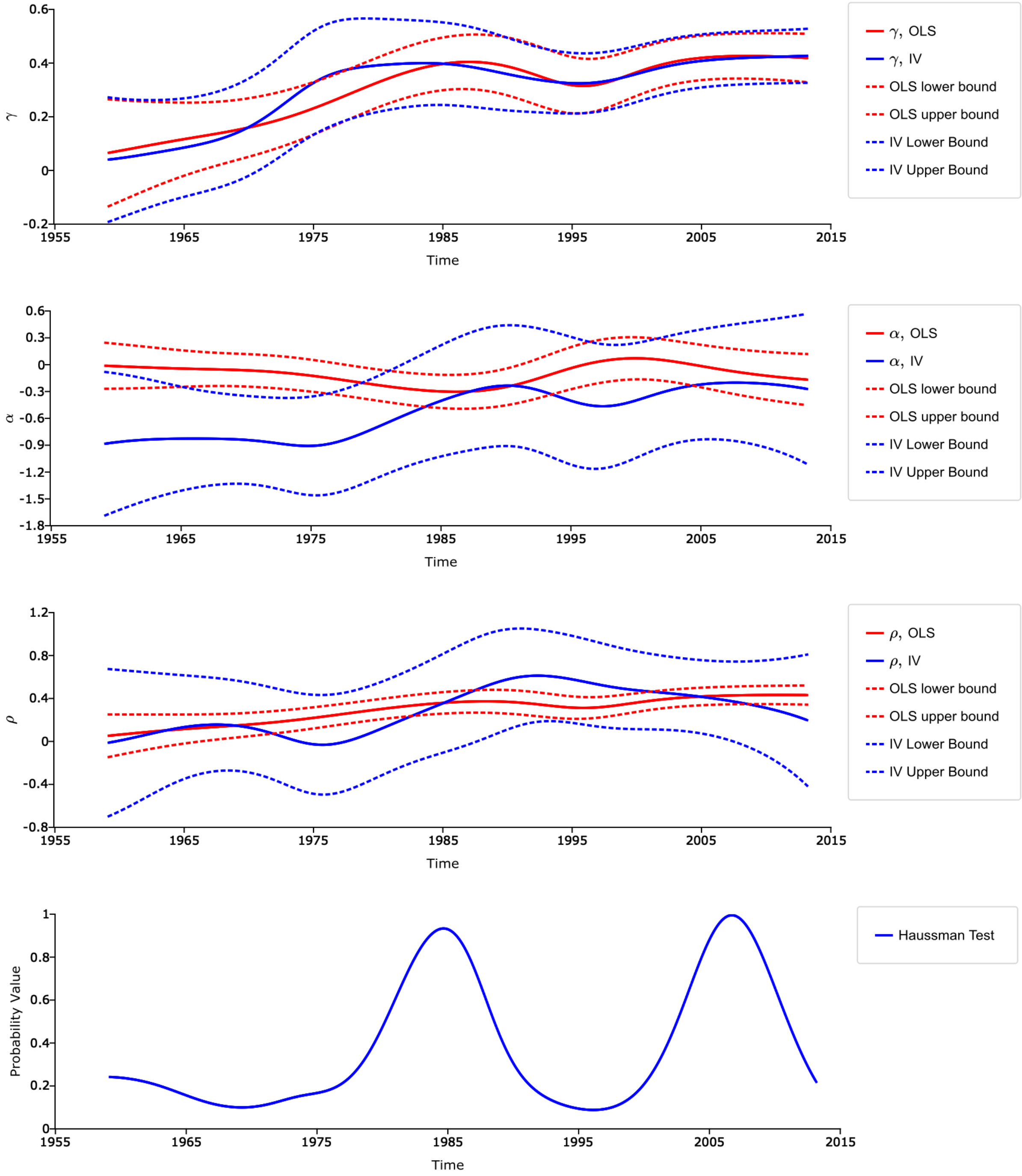

Table 1: Performance of estimators $\widehat{\beta}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ for the model (40)-(41) with exogenous x_t: $s = 0$, $H = T^{h_1}, L = T^{h_2}$.

h_1	h_2	T	Median Deviation			Abs. Median Deviation			Decile Range			Coverage Range		
			$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$
0.4	0.4	100	-0.000	0.000	-0.001	0.200	0.296	0.347	1.353	1.766	2.656	0.815	0.899	0.956
	0.4	200	0.003	-0.004	-0.002	0.168	0.260	0.298	1.382	1.728	2.424	0.830	0.910	0.964
	0.4	400	0.004	0.003	0.003	0.141	0.219	0.244	1.368	1.676	2.166	0.834	0.907	0.966
	0.4	1000	0.001	-0.000	-0.001	0.112	0.177	0.193	1.382	1.633	1.957	0.850	0.913	0.969
	0.5	100	0.001	0.004	0.002	0.198	0.267	0.344	1.361	1.414	2.691	0.821	0.830	0.958
	0.5	200	-0.001	-0.000	-0.001	0.167	0.231	0.294	1.367	1.442	2.429	0.832	0.828	0.963
	0.5	400	0.000	0.002	0.004	0.142	0.197	0.248	1.376	1.441	2.203	0.836	0.826	0.968
	0.5	1000	-0.001	0.002	0.002	0.113	0.159	0.195	1.405	1.455	2.016	0.852	0.818	0.972
0.5	0.4	100	-0.000	0.000	-0.000	0.201	0.322	0.333	1.186	2.202	2.474	0.731	0.937	0.924
	0.4	200	-0.000	-0.005	-0.005	0.168	0.274	0.272	1.246	2.060	2.228	0.734	0.946	0.922
	0.4	400	0.000	-0.001	-0.001	0.143	0.233	0.227	1.293	1.926	1.943	0.730	0.953	0.930
	0.4	1000	-0.001	-0.001	-0.000	0.115	0.192	0.182	1.289	1.830	1.768	0.727	0.959	0.930
	0.5	100	-0.001	-0.001	0.001	0.199	0.287	0.329	1.213	1.555	2.468	0.737	0.855	0.926
	0.5	200	0.000	-0.000	0.000	0.169	0.247	0.281	1.247	1.516	2.216	0.731	0.852	0.926
	0.5	400	0.001	0.001	0.000	0.143	0.205	0.225	1.339	1.531	1.941	0.725	0.839	0.926
	0.5	1000	0.000	-0.000	-0.001	0.115	0.165	0.179	1.296	1.478	1.760	0.722	0.836	0.928

Table 2: Performance of estimators $\widehat{\beta}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ for the model (40)-(41) with $s = 0.2$.

h_1	h_2	T	Median Deviation			Abs. Median Deviation			Decile Range			Coverage Range		
			$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$
0.4	0.4	100	0.018	-0.001	-0.003	0.174	0.199	0.208	1.280	1.354	1.455	0.760	0.806	0.839
	0.4	200	0.019	0.003	0.001	0.145	0.167	0.173	1.322	1.386	1.452	0.775	0.818	0.843
	0.4	400	0.020	0.001	-0.001	0.123	0.142	0.146	1.363	1.416	1.467	0.777	0.820	0.841
	0.4	1000	0.021	0.003	0.002	0.098	0.113	0.115	1.376	1.415	1.444	0.794	0.833	0.849
	0.5	100	0.021	0.004	0.002	0.174	0.203	0.209	1.328	1.224	1.508	0.765	0.715	0.844
	0.5	200	0.020	0.005	0.003	0.144	0.171	0.171	1.313	1.252	1.435	0.773	0.698	0.842
	0.5	400	0.020	0.002	-0.000	0.123	0.147	0.144	1.360	1.328	1.453	0.776	0.690	0.837
	0.5	1000	0.020	0.002	0.001	0.098	0.119	0.115	1.391	1.353	1.460	0.793	0.690	0.849
0.5	0.4	100	0.020	-0.002	-0.002	0.185	0.206	0.216	1.162	1.423	1.319	0.655	0.829	0.765
	0.4	200	0.019	0.002	0.001	0.156	0.174	0.182	1.249	1.458	1.364	0.652	0.836	0.757
	0.4	400	0.020	0.001	0.001	0.134	0.144	0.150	1.253	1.407	1.327	0.642	0.838	0.740
	0.4	1000	0.021	-0.000	-0.001	0.108	0.116	0.121	1.310	1.433	1.361	0.635	0.855	0.729
	0.5	100	0.020	0.002	0.002	0.184	0.209	0.215	1.157	1.214	1.324	0.654	0.723	0.766
	0.5	200	0.021	0.004	0.003	0.158	0.176	0.180	1.245	1.287	1.356	0.648	0.717	0.753
	0.5	400	0.021	0.001	0.000	0.134	0.150	0.152	1.305	1.330	1.368	0.642	0.711	0.744
	0.5	1000	0.020	0.001	0.001	0.108	0.119	0.120	1.314	1.336	1.355	0.626	0.697	0.722

Table 3: Performance of estimators $\widehat{\beta}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ for the model (40)-(41) with $s = 0.5$.

h_1	h_2	T	Median Deviation			Abs. Median Deviation			Decile Range			Coverage Range		
			$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$
0.4	0.4	100	0.143	0.034	0.022	0.203	0.187	0.193	1.286	1.326	1.408	0.631	0.771	0.807
	0.4	200	0.141	0.022	0.012	0.182	0.157	0.161	1.323	1.355	1.407	0.603	0.783	0.813
	0.4	400	0.145	0.019	0.011	0.170	0.133	0.135	1.362	1.377	1.412	0.557	0.792	0.817
	0.4	1000	0.141	0.012	0.007	0.155	0.105	0.106	1.381	1.402	1.422	0.497	0.806	0.824
	0.5	100	0.147	0.043	0.026	0.203	0.199	0.195	1.320	1.200	1.460	0.631	0.659	0.812
	0.5	200	0.148	0.030	0.014	0.186	0.168	0.163	1.314	1.248	1.417	0.602	0.663	0.818
	0.5	400	0.141	0.023	0.012	0.168	0.143	0.134	1.365	1.310	1.427	0.558	0.649	0.813
	0.5	1000	0.139	0.014	0.005	0.154	0.114	0.106	1.408	1.365	1.448	0.497	0.646	0.822
0.5	0.4	100	0.149	0.024	0.017	0.221	0.196	0.208	1.135	1.360	1.267	0.515	0.795	0.724
	0.4	200	0.151	0.017	0.013	0.201	0.163	0.172	1.238	1.380	1.299	0.467	0.808	0.718
	0.4	400	0.147	0.009	0.006	0.185	0.136	0.146	1.324	1.440	1.372	0.417	0.815	0.706
	0.4	1000	0.148	0.006	0.004	0.171	0.107	0.116	1.320	1.407	1.345	0.334	0.827	0.689
	0.5	100	0.147	0.020	0.010	0.220	0.201	0.207	1.144	1.183	1.273	0.514	0.681	0.728
	0.5	200	0.147	0.017	0.011	0.199	0.170	0.173	1.244	1.252	1.299	0.473	0.676	0.716
	0.5	400	0.150	0.011	0.006	0.186	0.145	0.146	1.270	1.280	1.311	0.414	0.674	0.707
	0.5	1000	0.144	0.007	0.004	0.168	0.115	0.115	1.289	1.297	1.314	0.338	0.659	0.684

Table 4: Performance of estimators $\widehat{\beta}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ in the overidentified case for the model (41)-(42) with exogenous x_t, $s = 0$, $H = T^{h_1}, L = T^{h_2}$.

h_1	h_2	T	Median Deviation			Abs. Median Deviation			Decile Range			Coverage Range		
			$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$
0.4	0.4	100	-0.001	0.000	0.001	0.181	0.215	0.228	1.316	1.414	1.538	0.788	0.839	0.872
	0.4	200	-0.001	0.000	0.001	0.154	0.185	0.193	1.359	1.444	1.540	0.802	0.849	0.876
	0.4	400	0.001	0.001	0.001	0.128	0.154	0.160	1.366	1.442	1.505	0.804	0.847	0.869
	0.4	1000	0.001	0.001	0.001	0.101	0.124	0.127	1.378	1.436	1.474	0.820	0.858	0.876
	0.5	100	0.001	0.003	0.003	0.185	0.217	0.233	1.332	1.231	1.568	0.785	0.744	0.871
	0.5	200	-0.001	-0.000	-0.001	0.153	0.184	0.195	1.316	1.265	1.512	0.797	0.745	0.872
	0.5	400	-0.001	0.001	0.000	0.128	0.156	0.163	1.337	1.303	1.478	0.806	0.743	0.875
	0.5	1000	-0.001	0.001	0.000	0.101	0.124	0.127	1.380	1.352	1.479	0.819	0.734	0.875
0.5	0.4	100	-0.007	-0.005	-0.006	0.190	0.229	0.233	1.182	1.506	1.393	0.682	0.857	0.804
	0.4	200	0.000	-0.000	0.000	0.159	0.189	0.190	1.270	1.509	1.408	0.684	0.866	0.794
	0.4	400	-0.001	0.001	-0.000	0.136	0.160	0.162	1.287	1.470	1.383	0.676	0.869	0.784
	0.4	1000	0.001	0.001	0.000	0.110	0.127	0.129	1.297	1.441	1.363	0.669	0.878	0.769
	0.5	100	-0.001	-0.007	-0.006	0.191	0.221	0.232	1.172	1.250	1.386	0.685	0.762	0.806
	0.5	200	0.003	0.004	0.004	0.159	0.184	0.190	1.245	1.312	1.392	0.683	0.758	0.795
	0.5	400	-0.001	-0.000	-0.001	0.136	0.158	0.162	1.293	1.329	1.377	0.677	0.754	0.787
	0.5	1000	-0.001	-0.000	-0.000	0.109	0.127	0.128	1.288	1.325	1.355	0.670	0.745	0.770

Table 5: Performance of estimators $\widehat{\beta}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ in the overidentified case for the model (41) - (42) with $s = 0.2$.

h_1	h_2	T	Median Deviation			Abs. Median Deviation			Decile Range			Coverage Range		
			$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$
0.4	0.4	100	0.032	0.008	0.002	0.192	0.272	0.314	1.356	1.677	2.490	0.808	0.883	0.943
	0.4	200	0.032	0.003	0.000	0.162	0.230	0.257	1.362	1.640	2.180	0.810	0.885	0.948
	0.4	400	0.029	-0.002	-0.003	0.137	0.196	0.216	1.339	1.586	1.985	0.820	0.891	0.953
	0.4	1000	0.033	0.004	0.002	0.111	0.159	0.171	1.366	1.576	1.843	0.828	0.895	0.959
	0.5	100	0.031	0.004	-0.001	0.193	0.252	0.309	1.351	1.335	2.497	0.802	0.808	0.950
	0.5	200	0.032	0.007	0.002	0.163	0.215	0.260	1.332	1.370	2.135	0.814	0.793	0.953
	0.5	400	0.031	0.006	0.003	0.137	0.182	0.214	1.386	1.431	2.018	0.817	0.790	0.952
	0.5	1000	0.033	0.006	0.003	0.112	0.149	0.173	1.388	1.415	1.853	0.830	0.790	0.965
0.5	0.4	100	0.030	0.004	0.004	0.197	0.288	0.302	1.192	2.022	2.156	0.708	0.917	0.898
	0.4	200	0.031	-0.000	-0.002	0.169	0.245	0.249	1.256	1.900	1.981	0.703	0.927	0.905
	0.4	400	0.030	0.000	0.000	0.143	0.209	0.211	1.292	1.817	1.819	0.702	0.935	0.905
	0.4	1000	0.032	0.001	-0.000	0.116	0.168	0.165	1.341	1.750	1.683	0.690	0.945	0.907
	0.5	100	0.029	0.001	0.000	0.198	0.265	0.301	1.163	1.439	2.152	0.711	0.827	0.905
	0.5	200	0.034	0.009	0.007	0.169	0.233	0.258	1.268	1.489	2.021	0.709	0.819	0.909
	0.5	400	0.033	0.004	0.002	0.143	0.190	0.206	1.286	1.451	1.778	0.701	0.809	0.908
	0.5	1000	0.032	0.000	-0.001	0.116	0.152	0.161	1.324	1.453	1.665	0.688	0.799	0.906

Table 6: Performance of estimators $\widehat{\beta}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ in the overidentified case for the model (41)-(42) with $s = 0.5$.

h_1	h_2	T	Median Deviation			Abs. Median Deviation			Decile Range			Coverage Range		
			$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$
0.4	0.4	100	0.249	0.040	0.020	0.279	0.253	0.283	1.331	1.594	2.267	0.581	0.856	0.934
	0.4	200	0.249	0.033	0.017	0.267	0.212	0.234	1.371	1.565	2.043	0.527	0.862	0.940
	0.4	400	0.253	0.024	0.010	0.264	0.181	0.196	1.357	1.523	1.865	0.452	0.866	0.943
	0.4	1000	0.246	0.015	0.005	0.252	0.144	0.153	1.375	1.524	1.734	0.365	0.872	0.951
	0.5	100	0.245	0.052	0.023	0.276	0.245	0.281	1.327	1.298	2.184	0.579	0.750	0.931
	0.5	200	0.249	0.043	0.017	0.267	0.206	0.235	1.380	1.371	2.034	0.525	0.754	0.939
	0.5	400	0.248	0.034	0.014	0.260	0.176	0.195	1.381	1.386	1.870	0.455	0.747	0.942
	0.5	1000	0.254	0.023	0.006	0.260	0.143	0.157	1.388	1.375	1.754	0.356	0.748	0.954
0.5	0.4	100	0.255	0.032	0.027	0.290	0.268	0.284	1.228	1.905	2.108	0.468	0.901	0.885
	0.4	200	0.243	0.013	0.002	0.270	0.223	0.231	1.264	1.790	1.809	0.409	0.909	0.876
	0.4	400	0.252	0.015	0.010	0.270	0.190	0.194	1.305	1.707	1.710	0.324	0.916	0.889
	0.4	1000	0.263	0.009	0.004	0.274	0.157	0.157	1.340	1.692	1.635	0.221	0.935	0.893
	0.5	100	0.248	0.031	0.013	0.287	0.257	0.285	1.172	1.371	2.044	0.472	0.782	0.882
	0.5	200	0.254	0.023	0.012	0.279	0.214	0.234	1.265	1.413	1.838	0.398	0.790	0.889
	0.5	400	0.250	0.011	0.002	0.269	0.180	0.195	1.309	1.419	1.715	0.326	0.778	0.887
	0.5	1000	0.256	0.011	0.004	0.267	0.146	0.154	1.344	1.422	1.603	0.228	0.768	0.892

Table 7: Rejection frequencies for the local Hausman test at $t = T/2$. Model: (40)-(41)

s	h_1	h_2	$T = 100$	$T = 200$	$T = 400$	$T = 1000$
0	0.4	0.4	0.011	0.014	0.013	0.007
	0.4	0.5	0.039	0.031	0.046	0.059
	0.5	0.4	0.113	0.142	0.157	0.167
	0.5	0.5	0.013	0.021	0.014	0.022
0.2	0.4	0.4	0.018	0.017	0.022	0.024
	0.4	0.5	0.071	0.070	0.084	0.075
	0.5	0.4	0.162	0.175	0.199	0.213
	0.5	0.5	0.023	0.023	0.026	0.039
0.5	0.4	0.4	0.126	0.163	0.313	0.415
	0.4	0.5	0.148	0.204	0.247	0.384
	0.5	0.4	0.290	0.358	0.416	0.509
	0.5	0.5	0.197	0.319	0.457	0.618

Table 8: Rejection frequencies for the global Hausman test. Model: (40)-(41)

s	h_1	h_2	$T = 100$	$T = 200$	$T = 400$	$T = 1000$
0	0.4	0.4	0.020	0.020	0.020	0.020
	0.4	0.5	0.022	0.018	0.026	0.022
	0.5	0.4	0.036	0.028	0.024	0.028
	0.5	0.5	0.016	0.020	0.016	0.020
0.2	0.4	0.4	0.028	0.042	0.054	0.150
	0.4	0.5	0.034	0.052	0.078	0.140
	0.5	0.4	0.044	0.052	0.074	0.148
	0.5	0.5	0.028	0.032	0.060	0.144
0.5	0.4	0.4	0.574	0.846	0.966	1.000
	0.4	0.5	0.518	0.832	0.966	1.000
	0.5	0.4	0.458	0.694	0.890	0.966
	0.5	0.5	0.528	0.828	0.942	0.998

Appendix A

The proofs of the Theorem 1 and Lemma 2 exploit a number of additional results, presented after the proofs. The proof of Theorem 4 is reported in the Online Supplement.

Proof of theorems

Proof of Theorem 1. (i) We start with the proof of (19)-(20). By (2), $x_t = \Psi_j' z_t + v_t$. Hence[3]

$$\begin{aligned}
\widehat{\Psi}_t &= \big(\sum_{j=1}^{T} b_{H,|j-t|} z_j z_j'\big)^{-1}\big(\sum_{j=1}^{T} b_{H,|j-t|} z_j x_j'\big) \\
&= \big(\sum_{j=1}^{T} b_{H,|j-t|} z_j z_j'\big)^{-1}\big(\sum_{j=1}^{T} b_{H,|j-t|} z_j z_j' \Psi_j + \sum_{j=1}^{T} b_{H,|j-t|} z_j v_j'\big) \\
&= \Psi_t + S_{zz,t}^{-1}(\Delta_t + S_{zv,t})
\end{aligned}$$

where

$$S_{zz,t} = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} z_j z_j', \quad \Delta_t = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} z_j z_j' (\Psi_j - \Psi_t), \quad S_{zv,t} = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} z_j v_j'.$$

Using the following property of spectral and Frobenius norms, $||AB|| \le ||A||_{sp}||B||$, it follows

$$\max_{t=1,\ldots,T} ||\widehat{\Psi}_t - \Psi_t|| \le \max_{t=1,\ldots,T} ||S_{zz,t}^{-1}||_{sp}\big(\max_{t=1,\ldots,T} ||\Delta_t|| + \max_{t=1,\ldots,T} ||S_{zv,t}||\big). \tag{46}$$

We will show that

$$\max_{t=1,\ldots,T} ||S_{zz,t}^{-1}||_{sp} = O_p(1), \tag{47}$$

$$\max_{t=1,\ldots,T} ||\Delta_t|| = O_p(H/T) \quad \text{if } \Psi_t \text{ satisfies Assumption 2}, \tag{48}$$

$$\max_{t=1,\ldots,T} ||\Delta_t|| = O_p\big((H/T)^{1/2} \log^{2/\alpha} T\big) \quad \text{if } \Psi_t \text{ satisfies Assumption 3}, \tag{49}$$

$$\max_{t=1,\ldots,T} ||S_{zv,t}|| = O_p(H^{-1/2} \log^{1/2} T). \tag{50}$$

These bounds together with (46) prove (19)-(20).

[3] In the proofs we use the generic notation H for the bandwidth for simplicity, only introducing the notation L when the two bandwidths interact.

Proof of (47). Write

$$S_{zz,t} = K_t^{-1}\sum_{j=1}^{T} b_{H,|j-t|}E[z_j z_j'] + K_t^{-1}\sum_{j=1}^{T} b_{H,|j-t|}(z_j z_j' - E[z_j z_j']) =: S_{zz,t}^{(1)} + S_{zz,t}^{(2)}.$$

Rewrite $S_{zz,t}$ as

$$S_{zz,t} = S_{zz,t}^{(1)}(1+\widetilde{\Delta}_t), \quad \widetilde{\Delta}_t = S_{zz,t}^{(1)\,-1}(S_{zz,t} - S_{zz,t}^{(1)}).$$

If $||\widetilde{\Delta}_t^{-1}||_{sp} < 1$, then

$$\begin{aligned} ||S_{zz,t}^{-1}||_{sp} &\leq ||(S_{zz,t}^{(1)})^{-1}||_{sp}||(1+\widetilde{\Delta}_t)^{-1}||_{sp} \leq ||(S_{zz,t}^{(1)})^{-1}||_{sp}(1 - ||\widetilde{\Delta}_t||_{sp})^{-1} \\ &\leq \max_{t=1,...,T} ||(S_{zz,t}^{(1)})^{-1}||_{sp}(1 - \max_{t=1,...,T} ||\widetilde{\Delta}_t||_{sp})^{-1}. \end{aligned} \tag{51}$$

We will show that

$$\max_{t=1,...,T} ||(S_{zz,t}^{(1)})^{-1}||_{sp} = O_p(1), \tag{52}$$

$$\max_{t=1,...,T} ||\widetilde{\Delta}_t||_{sp} = o_p(1) \tag{53}$$

which together with (51) imply (47): $\max_{t=1,...,T} ||(S_{zz,t}^{(1)})^{-1}||_{sp} = O_p(1)$.

To prove (52), recall that by Assumption 1(iii), there exists $\nu > 0$ such that for all $t \geq 1$,

$$a'\Sigma_{zz,t}a \geq 1/\nu > 0.$$

Thus, for any $1 \times n$ vector $a = (a_1, ..., a_p)'$ such that $||a||^2 = 1$,

$$\min_{||a||=1} a' S_{zz,t}^{(1)} a = \min_{||a||=1}\Big(K_t^{-1}\sum_{j=1}^{T} b_{H,|j-t|}a'\Sigma_{zz,j}a\Big) \geq \nu^{-1}(K_t^{-1}\sum_{j=1}^{T} b_{H,|j-t|}) = 1/\nu > 0.$$

Hence, the smallest eigenvalue of $S_{zz,t}^{(1)}$ is not smaller than $1/\nu > 0$ which yields (52).

To show (53), bound

$$||\widetilde{\Delta}_t||_{sp} \leq ||(S_{zz,t}^{(1)})^{-1}||_{sp}||S_{zz,t} - S_{zz,t}^{(1)}||.$$

In view of (52), to verify (53), it suffices to show that

$$\max_{t=1,...,T} ||S_{zz,t} - S_{zz,t}^{(1)}|| = o_P(1).$$

By Assumption 1(ii), the (s,k)-th component $\omega_{\ell k,j} = z_{\ell,j} z_{k,j} - E z_{\ell,j} z_{k,j}$ of $z_j z'$ is α-mixing and by (12), for $\theta' = \theta/2$, $E|z_{\ell,j} z_{k,j}|^{\theta'} \le C < \infty$ uniformly over j. By assumption (17), $c_1 T^{1/(\theta'/2-1)+\delta} \le H$, $L \le c_2 T^{1-\delta}$. Hence, Lemma 5(i) implies (53):

$$\max_{t=1,\dots,T} ||S_{zz,t} - S^{(1)}_{zz,t}|| = O_p(H^{-1/2} \log^{1/2} T) = o_p(1). \tag{54}$$

Proof of (48). A typical element of Δ_t consists of a linear combination of sums

$$s_t := K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \omega_{\ell k,j} (\psi_{km,j} - \psi_{km,t})$$

where $\omega_{\ell k,j} - E\omega_{\ell k,j}$ is an α-mixing sequence and $E|\omega_{\ell k,j}|^{\theta/2} \le C < \infty$ for all j.

Suppose that Ψ_t satisfies Assumption 2. Then $|\psi_{km,j} - \psi_{km,t}| \le C(|j-t|/T)$ and (90) of Lemma 5(ii) implies the bound (48): $\max_{t=1,\dots,T} |s_t| = O_p\big(H/T\big)$.

Suppose that Ψ_t satisfies Assumption 3. Then $|\psi_{km,j} - \psi_{km,t}| \le (|j-t|/T)^{1/2} v_{km,jt}$ and (92) of Lemma 5(iii) implies (49): $\max_{t=1,\dots,T} |s_t| = O_p\big((H/T)^{1/2} (\log T)^{2/\alpha}\big)$.

Proof of (50). The (ℓ,k)-th element of the $n \times p$ matrix $S_{zv,t}$ is

$$\widetilde{s}_t := K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} z_{\ell,j} v_{k,j}$$

where by (3), $E z_{\ell,j} v_{k,j} = 0$. Moreover, by Assumption 1, the sequence $(z_{\ell,j} v_{k,j})$ is α-mixing and $|z_{\ell,j} v_{k,j}|^{\theta/2} \le C < \infty$, uniformly in j. Thus, the same argument as in the proof of (54) implies (50):

$$\max_{t=1,\dots,T} |\widetilde{s}_t| = O(H^{-1/2} \log^{1/2} T).$$

(ii) Next we prove (22)-(23). By definition (1), $y_j = x_j' \beta_j + u_j = x_j' \beta_t + x_j'(\beta_j - \beta_t) + u_j$. Denote $\widehat{x}_j = \widehat{\Psi}_j' z_j$. Then,

$$\begin{aligned}
\widetilde{\beta}_{1,t} &= \big(\sum_{j=1}^{T} b_{H,|j-t|} \widehat{x}_j x_j'\big)^{-1} \big(\sum_{j=1}^{T} b_{H,|j-t|} \widehat{x}_j y_j'\big) \\
&= \big(\sum_{j=1}^{T} b_{H,|j-t|} \widehat{x}_j x_j'\big)^{-1} \big(\sum_{j=1}^{T} b_{H,|j-t|} \widehat{x}_j [x_j' \beta_j + u_j]\big) \\
&= \beta_t + S^{-1}_{\widehat{x}x,t} (\Delta^{(1)}_t + S_{\widehat{x}x,t})
\end{aligned} \tag{55}$$

where $S_{\widehat{x}x,t} = K_t^{-1} \sum_{j=1}^T b_{H,|j-t|} \widehat{x}_j x_j'$,

$$\Delta_t^{(1)} = K_t^{-1} \sum_{j=1}^T b_{H,|j-t|} \widehat{\Psi}_j' z_j x_j' (\beta_j - \beta_t), \quad S_{\widehat{x}x,t} = K_t^{-1} \sum_{j=1}^T b_{H,|j-t|} \widehat{x}_j u_j.$$

Then,

$$||\widetilde{\beta}_{1,t} - \beta_t|| \le ||S_{\widehat{x}x,t}^{-1}||_{sp}^{-1} (||\Delta_t^{(1)}|| + ||S_{\widehat{x}u,t}||), \tag{56}$$

$$||\widetilde{\beta}_{1,t} - \beta_t - S_{\widehat{x}x,t}^{-1} S_{\widehat{x}u,t}|| \le ||S_{\widehat{x}x,t}^{-1}||_{sp}^{-1} ||\Delta_t^{(1)}||. \tag{57}$$

If Assumption 2 holds, set $\delta_{T,H}^* = (H/T)$, $\nu = 0$, $d_{T,H} = r_{T,H}$. If Assumption 3 holds, set $\delta_{T,H}^* = (H/T)^{1/2} \log^{3/\alpha} T$, $\nu = 1$, $d_{T,H} = \bar{r}_{T,H,\alpha}$. We will show that

$$\max_{t=1,\dots,T} ||S_{\widehat{x}x,t}^{-1}||_{sp}^{-1} = O_p(1), \tag{58}$$

$$\max_{t=1,\dots,T} ||\Delta_t^{(1)}|| = O_p(\delta_{T,H}^*), \tag{59}$$

$$\max_{t=1,\dots,T} ||S_{\widehat{x}u,t}^{-1}||_{sp}^{-1} = O_p\big((\log^{1/\alpha} T)^\nu d_{T,H} + d_{T,L}\big). \tag{60}$$

This together with (56) proves (22)-(23).

For notational simplicity we prove (58)-(60) in the scalar case when $p = 1$ and $n = 1$. (The case $p > 1$, $n > 1$ reduces to the analysis of a finite number of similar sums of scalar variables.)

We start with the proof of (58). Denote $M_t = K_t^{-1} \sum_{j=1}^T b_{H,|j-t|} \Psi_j' E[z_j z_j'] \Psi_j$. We will show that

$$\max_{t=1,\dots,T} ||M_t^{-1}||_{sp} = O_p(1), \tag{61}$$

$$\max_{t=1,\dots,T} ||S_{\widehat{x}x,t} - M_t|| = O_P(\log^{-1} T) = o_p(1) \tag{62}$$

which implies (58) using the same argument as in the proof of (47).

Using $x_j' = z_j' \Psi_j + v_j'$, write

$$\begin{aligned} \widehat{\Psi}_j' z_j x_j' &= (\widehat{\Psi}_j' - \Psi_j') z_j x_j' + \Psi_j' z_j x_j' = (\widehat{\Psi}_j' - \Psi_j') z_j x_j' + \Psi_j' z_j z_j' \Psi_j + \Psi_j' z_j v_j' \\ &= \Psi_j' E[z_j z_j'] \Psi_j + \Psi_j' (z_j z_j' - E[z_j z_j']) \Psi_j + \Psi_j' z_j v_j' + (\widehat{\Psi}_j' - \Psi_j') z_j z_j' \Psi_j + (\widehat{\Psi}_j' - \Psi_j') z_j v_j' \\ &=: g_t^{(1)} + \dots + g_t^{(5)}. \end{aligned} \tag{63}$$

Then,

$$\begin{aligned} S_{\widehat{x}x,t} &= K_t^{-1}\sum_{j=1}^{T} b_{H,|j-t|}g_t^{(1)} + ... + K_t^{-1}\sum_{j=1}^{T} b_{H,|j-t|}g_t^{(5)} \\ &=: M_t + S_t^{(2)} + S_t^{(3)} + S_t^{(4)} + S_t^{(5)}. \\ S_{\widehat{x}x,t} - M_t &= S_t^{(2)} + S_t^{(3)} + S_t^{(4)} + S_t^{(5)}. \end{aligned} \tag{64}$$

Observe, $\Sigma_{\Psi zz\Psi,j}$ satisfies assumption (21). Thus, (61) follows using the same argument as in the proof of (52).

To prove (62), recall that z_t and v_t satisfy Assumption 1. Note that

$$||S_{\widehat{x}x,t} - M_t|| \le ||S_t^{(2)}|| + ... + ||S_t^{(5)}||.$$

a) Suppose that Ψ_t satisfies Assumption 2. Then, using (91), we obtain

$$\max_{t=1,...,T}(||S_t^{(2)}|| + ||S_t^{(3)}||) = O_P\big(HT^{-1} + H^{-1/2}\log^{1/2} T\big) = O_p(r_{T,H}).$$

In addition,

$$\max_{t=1,...,T}(||S_t^{(4)}|| + ||S_t^{(5)}||) \le (\max_{j=1,...,T} ||\widehat{\Psi}_j - \Psi_j||)(\max_{j=1,...,T} ||\Psi_j||) r_T, \tag{65}$$

$$r_T = \max_{t=1,...,T} K_t^{-1}\textstyle\sum_{j=1}^{T} b_{H,|j-t|}(|z_j z_j'| + |z_j v_j'|) = O_p(1)$$

where $r_T = O_p(1)$ by (89) and $\max_{j=1,...,T} ||\widehat{\Psi}_j - \Psi_j||$ satisfies (19)-(20) with the bandwidth L. Under Assumption 2, $\max_{j=1,...,T} ||\Psi_j|| = O(1)$, so we obtain

$$\max_{t=1,...,T}(||S_t^{(4)}|| + ||S_t^{(5)}||) = O_p(1) \max_{j=1,...,T} ||\widehat{\Psi}_j' - \Psi_j'|| = O_p(r_{T,L}). \tag{66}$$

This yields

$$\max_{t=1,...,T} ||S_{\widehat{x}x,t} - M_t|| = O_p\big(r_{T,H} + r_{T,L}\big) = O_p(\log^{-1} T) \tag{67}$$

which verifies (62).

b) Suppose that Ψ_t satisfies Assumption 3. Then using (93), we obtain

$$\max_{t=1,...,T}(||S_t^{(2)}|| + ||S_t^{(3)}||) = O_p(\bar{r}_{T,H,\alpha} \log^{1/\alpha} T).$$

while (65), (20), (94) and (89) imply

$$\max_{t=1,\dots,T}(||S_t^{(4)}|| + ||S_t^{(5)}||) = O_P(\bar{r}_{T,L,\alpha}\log^{1/\alpha} T). \tag{68}$$

This yields

$$\max_{t=1,\dots,T} ||S_{\widehat{x}x,t} - M_t|| = O_p\big(\{\bar{r}_{T,H,\alpha} + \bar{r}_{T,L,\alpha}\}\log^{1/\alpha} T\big) = O_p(\log^{-1} T) \tag{69}$$

under (17) which verifies (62) and completes the proof of (58).

Proof of (59). Bound

$$\max_{t=1,\dots,T} ||\Delta_t^{(1)}|| \le (\max_{j=1,\dots,T} ||\widehat{\Psi}_j||)\, s_T, \quad s_T := \max_{t=1,\dots,T} K_t^{-1}\sum_{j=1}^{T} b_{H,|j-t|}||z_j x_j'(\beta_j - \beta_t)||. \tag{70}$$

By Theorem 1(i), $\max_{j=1,\dots,T} ||\widehat{\Psi}_j - \Psi_j|| = o_p(1)$. Thus,

$$\max_{j=1,\dots,T} ||\widehat{\Psi}_j|| = \max_{j=1,\dots,T} ||\widehat{\Psi}_j - \Psi_j|| + \max_{j=1,\dots,T} ||\Psi_j|| = o_p(1) + \max_{j=1,\dots,T} ||\Psi_j||. \tag{71}$$

To bound s_T, we use equation (2) for x_t to obtain $||x_j|| \le ||\Psi_j||\, ||z_j|| + ||v_j|| \le (||\Psi_j|| + 1)(||z_j|| + ||v_j||)$. Thus,

$$s_T := \max_{j=1,\dots,T}(||\Psi_j|| + 1)s_T*, \quad s_T^* = \max_{t=1,\dots,T} K_t^{-1}\sum_{j=1}^{T} b_{H,|j-t|}\big(||z_j|| + ||v_j||\big)||\beta_j - \beta_t)||.$$

To bound s_T^*, we apply Lemma 5 with θ replaced by $\theta/2$.
Let Assumption 2 hold. Then $\max_{j=1,\dots,T} ||\Psi_j|| = O(1)$ and (90) implies $s_T^* = O_p(H/T)$. Hence,

$$\max_{t=1,\dots,T} ||\Delta_t^{(1)}|| = O_p(H/T). \tag{72}$$

Let Assumption 3 hold. Then by (94), $\max_{j=1,\dots,T} ||\Psi_j|| = O_p(\log^{1/\alpha} T)$ and (92) implies $s_T^* = O_p\big((H/T)^{1/2}\log^{2/\alpha} T\big)$. Hence,

$$\max_{t=1,\dots,T} ||\Delta_t^{(1)}|| = O_p\big((H/T)^{1/2}\log^{3/\alpha} T\big). \tag{73}$$

This proves (59).

Proof of (60). Bound

$$||S_{\widehat{x}u,t}|| \le ||S_{\widehat{x}u,t} - S_{x-v,u,t}|| + ||S_{x-v,u,t}||.$$

We will show that

$$\max_{t=1,\dots,T} ||S_{\widehat{x}u,t} - S_{x-v,u,t}|| = O_p(d_{T,L}), \tag{74}$$

$$\max_{t=1,\dots,T} ||S_{x-v,u,t}|| = O_p\big((\log^{1/\alpha} T)^{\nu} d_{T,H}\big). \tag{75}$$

To prove (74), bound

$$\begin{aligned} ||S_{\widehat{x}u,t} - S_{x-v,u,t}|| &\le K_t^{-1} || \sum_{j=1}^{T} b_{H,|j-t|} (\widehat{\Psi}_j' - \Psi_j{}') z_j u_j || \\ &\le (\max_{j=1,\dots,T} ||\widehat{\Psi}_j - \Psi_j||) q_{T,t}, \quad q_{T,t} = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} ||z_j u_j||. \end{aligned}$$

By Theorem 1, $\max_{j=1,\dots,T} ||\widehat{\Psi}_j - \Psi_j|| = O_p(d_{T,L})$. Under assumption of lemma, by (89) of Lemma 5 $\max_{t=1,\dots,T} q_{T,t} = O_p(1)$. This proves (74).

Recall $S_{x-v,u,t} = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \Psi_j{}' z_j u_j$. Thus, (75) follows using (91) and (93) of Lemma 5. This completes the proof of the theorem. □

Proof of Lemma 2. Let M_t be as in (61). Then

$$||\widetilde{\beta}_{1,t} - \beta_t - M_t^{-1} \Psi_t S_{zu,t}|| \le ||\widetilde{\beta}_{1,t} - \beta_t - S_{\widehat{x}x,t}^{-1} S_{\widehat{x}u,t}|| + ||S_{\widehat{x}x,t}^{-1} S_{\widehat{x}u,t} - M_t^{-1} \Psi_t S_{zu,t}||. \tag{76}$$

We will show that

$$||\widetilde{\beta}_{1,t} - \beta_t - S_{\widehat{x}x,t}^{-1} S_{\widehat{x}u,t}|| = O_p\big((H/T)^{\gamma}\big), \tag{77}$$

$$||S_{\widehat{x}x,t}^{-1} S_{\widehat{x}u,t} - M_t^{-1} \Psi_t S_{zu,t}|| = O_p\big((H/T)^{\gamma} + d_{T,L}\big) \tag{78}$$

where $\gamma = 1$, $d_{T,L} = r_{T,L}$ if Assumption 2 holds; $\gamma = 1/2$, $d_{T,L} = \bar{r}_{T,L}$ if Assumption 3 holds. So, (76)-(78) prove (24) since under Assumption 4, $(H/T)^{\gamma} + d_{T,L} = o(H^{-1/2})$.

Proof (77). By (57) and (58),

$$\max_{t=1,\dots,T} ||\widetilde{\beta}_{1,t} - \beta_t - S_{\widehat{x}x,t}^{-1} S_{\widehat{x}u,t}|| \le ||S_{\widehat{x}x,t}^{-1}||_{sp}^{-1} ||\Delta_t^{(1)}|| = O_p(1) ||\Delta_t^{(1)}|| \tag{79}$$

with $\Delta_t^{(1)}$ as in (55). We have $||\Delta_t^{(1)}|| = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}||\widehat{\Psi}_j||\,||z_j x_j'(\beta_j-\beta_t)||$. Bound

$$||\widehat{\Psi}_j|| \le ||\widehat{\Psi}_j - \Psi_j|| + ||\Psi_j|| \le (||\widehat{\Psi}_j - \Psi_j|| + 1)(||\Psi_j|| + 1). \tag{80}$$

Then,

$$||\Delta_t^{(1)}|| \le (H/T)^\gamma \,(\max_{j=1,\dots,T}(||\widehat{\Psi}_j - \Psi_j|| + 1))\, r_{T,t},$$
$$r_{T,t} = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}(||\Psi_j|| + 1)\,||z_j x_j'(\beta_j - \beta_t)(T/H)^{1/2}||.$$

By Theorem 1, $\max_{j=1,\dots,T}||\widehat{\Psi}_j - \Psi_j|| = o_p(1)$. Under assumptions of lemma, $\max_{j=1,\dots,T} E[(||\Psi_j||+1)||z_j x_j'(\beta_j-\beta_t)(T/H)^{1/2}||] = O(1)$. This implies $Er_{T,t} = O_p(1)$ and $r_{T,t} = O_p(1)$ which proves (77).

Proof (78). Bound

$$\begin{aligned}||S^{-1}_{\widehat{x}x,t}S_{\widehat{x}u,t} - M_t^{-1}\Psi_t S_{zu,t}|| &\le ||S^{-1}_{\widehat{x}x,t}(S_{\widehat{x}u,t} - \Psi_t S_{zu,t})|| + ||(S^{-1}_{\widehat{x}x,t} - M_t^{-1})\Psi_t S_{zu,t}|| \\ &\le ||S^{-1}_{\widehat{x}x,t}||_{sp}||S_{\widehat{x}u,t} - \Psi_t S_{zu,t}|| + ||S^{-1}_{\widehat{x}x,t} - M_t^{-1}||\,||\Psi_t||\,||S_{zu,t}||.\end{aligned} \tag{81}$$

By (57), $||S^{-1}_{\widehat{x}x,t}||_{sp} = O_p(1)$. We will show

$$||S_{\widehat{x}u,t} - \Psi_t S_{zu,t}|| = O_p(d_{T,L}), \tag{82}$$
$$||S^{-1}_{\widehat{x}x,t} - M_t^{-1}||\,||\Psi_t||\,||S_{zu,t}|| = o_p(H^{-1/2}). \tag{83}$$

which proves (78).

First we prove (82). Bound

$$||S_{\widehat{x}u,t} - \Psi_t S_{zu,t}|| \le ||S_{\widehat{x}u,t} - S_{x-v,u,t}|| + ||S_{x-v,u,t} - \Psi_t S_{zu,t}||.$$

By (74), $\max_{t=1,\dots,T}||S_{\widehat{x}u,t} - S_{x-v,u,t}|| = O_p(d_{T,L})$.

Next, bound

$$||S_{x-v,u,t} - \Psi_t S_{zu,t}|| \le (H/T)^\gamma\, r_{T,t}, \quad r_{T,t} = K_t^{-1}\sum_{j=1}^T b_{H,|j-t|}(T/H)^\gamma||\Psi_j - \Psi_t||\,||z_j u_j||.$$

Under assumptions of lemma, $\max_{j=1,\dots,T} E\big[(T/H)^\gamma||\Psi_j - \Psi_t||\,||z_j u_j||\big] = O(1)$. This implies $Er_{T,t} = O_p(1)$ and $r_{T,t} = O_p(1)$ which proves $||S_{x-v,u,t} - \Psi_t S_{zu,t}|| = o_p((H/T)^\gamma)$ and completes the proof of (82).

To prove (83), notice that under assumptions of lemma, $E||\Psi_t|| = O(1)$ and thus $||\Psi_t|| = O_p(1)$, while by (88) of Lemma 5, $||S_{zu,t}|| = O_p(H^{-1/2}\log^{1/2} T)$. By (58), (61) and (62),

$$||S^{-1}_{\widehat{x}x,t} - M_t^{-1}|| \le ||S^{-1}_{\widehat{x}x,t}||\,||S_{\widehat{x}x,t} - M_t||\,||M_t^{-1}|| = O_p(\log^{-1} T) = o_p(\log^{-1/2} T)$$

which implies (83): $||S^{-1}_{\widehat{x}x,t} - M_t^{-1}||\,||\Psi_t||\,||S_{zu,t}|| = o_p(H^{-1/2})$. This proves (24).

Proof of (25). By Theorem 1, when L satisfies Assumption 4 the estimate $\widehat{\Psi}_j$ of Ψ_j has property $\max_{j=1,\dots,T} ||\widehat{\Psi}_j - \Psi_j|| = o_p(H^{-1/2})$. This, combined with the arguments used in the proof of (24) implies that besides $\widetilde{\beta}_{1,t}$ the estimates $\widetilde{\beta}_{2,t}$ and $\widetilde{\beta}_{3,t}$ also satisfy (24) which yields (25). This completes the proof of the lemma.

□

Auxiliary results

In this section we obtain uniform bounds for sums $H^{-1}\sum_{k=1}^T b_{H,|t-k|}(\xi_k - E\xi_k)$,

$$\nu_{T,H,t} = H^{-1}\sum_{k=1}^T b_{H,|t-k|}g_k(\xi_k - E\xi_k), \quad \Delta_{T,H,t} = H^{-1}\sum_{k=1}^T b_{H,|t-k|}\big|(g_k - g_t)\xi_k\big|,$$

where $\xi_k - E\xi_k$ is an α-mixing (but not necessarily stationary) univariate sequence satisfying (13) and such that

$$\max_{k\ge 1} E|\xi_k|^\theta \le c < \infty \quad \text{for some } \theta > 4. \tag{84}$$

Univariate variables g_t are specified below. Weights $b_{H,|t-k|}$ are as in (6) and computed with bandwidth H such that for some $c_0 > 0$ and $\delta > 0$,

$$c_0 T^{1/(\theta/2-1)+\delta} \le H = o(T). \tag{85}$$

Condition (84) implies that for some $c > 0$, for all $k \ge 1$,

$$P(|\xi_k| \ge x) \le c|x|^{-\theta}, \quad x > 0. \tag{86}$$

We shall write $(x_{kt}) \in \mathcal{E}(\alpha)$, $\alpha > 0$ to denote that

$$P(|x_{kt}| \ge x) \le c_0 \exp(-c_1|x|^\alpha), \quad x > 0 \tag{87}$$

where $c_0, c_1 > 0$ do not depend on k, t.

Lemma 5 *Let (ξ_j) be an α-mixing sequence satisfying (84) and (85) holds.*

(i) Then,

$$\max_{t=1,...,T} \left|H^{-1} \textstyle\sum_{k=1}^{T} b_{H,|t-k|}(\xi_k - E\xi_k)\right| = O_P\big(H^{-1/2}\sqrt{\log T}\big), \tag{88}$$
$$\max_{t=1,...,T} H^{-1} \textstyle\sum_{k=1}^{T} b_{H,|t-k|}|\xi_k| = O_p(1). \tag{89}$$

(ii) Let $|g_k - g_t| \le C|k-j|/T$ *for* $k, t = 1, ..., T$ *where* C *does not depend on* k, t, T*. Then,*

$$\max_{1\le t\le T} |\Delta_{T,H,t}| = O_P(H/T), \tag{90}$$
$$\max_{1\le t\le T} |\nu_{T,H,t}| = O_P\big((H/T) + H^{-1/2}\sqrt{\log T}\big). \tag{91}$$

(iii) Let $|(g_t - g_k)| \le \big(|t-k|/T\big)^{1/2} v_{kt}$ *for* $t, k = 1, ..., T$ *and* $(v_{kt}) \in \mathcal{E}(\alpha)$*,* $(h_t) \in \mathcal{E}(\alpha)$*,* $\alpha > 0$*. Then,*

$$\max_{1\le t\le T} |\Delta_{T,H,t}| = O_P\big(\sqrt{H/T}\,(\log T)^{2/\alpha}\big), \tag{92}$$
$$\max_{1\le t\le T} |\nu_{T,H,t}| = O_P\big(\{\sqrt{H/T}\,(\log T)^{2/\alpha} + H^{-1/2}\sqrt{\log T}\}(\log T)^{1/\alpha}\big). \tag{93}$$

(iv) Let $(g_t) \in \mathcal{E}(\alpha)$*,* $\alpha > 0$*. Then,*

$$\max_{1\le t\le T} |g_t| = O_P\big(\log^{1/\alpha} T\big). \tag{94}$$

Proof. (i) Corollary 7(b) of Dendramis, Giraitis and Kapetanios (2018) (DGK) implies that for any $\varepsilon > 0$,

$$\max_{1\le t\le T} \left|H^{-1} \sum_{k=1}^{T} b_{H,|t-k|}(\xi_k - E\xi_k)\right| = O_P\big(H^{-1/2}\sqrt{\log T} + (HT)^{1/\theta} H^{\varepsilon-1}\big). \tag{95}$$

Under Assumption (85), for sufficiently small $\varepsilon > 0$, it holds $(HT)^{1/\theta}H^{\varepsilon-1} \le H^{-1/2}$ which together with (95) proves (88).

The bound (89) is shown in Corollary 8(b) in DGK.

(ii)-(iii): (90) and (92) are shown in Corollary 10 (b) in DGK.

To show (91) and (93), write

$$\begin{aligned}\nu_{T,H,t} &= H^{-1} \textstyle\sum_{k=1}^{T} b_{H,|t-k|} g_k(\xi_k - E\xi_k) \\ &= H^{-1} \textstyle\sum_{k=1}^{T} b_{H,|t-k|}(g_k - g_t)(\xi_k - E\xi_k) + g_t H^{-1} \textstyle\sum_{k=1}^{T} b_{H,|t-k|}(\xi_k - E\xi_k) \\ &=: \widetilde{\Delta}_{T,H,t} + g_t v_{T,H,t}.\end{aligned}$$

Hence,

$$\max_{t=1,\dots,T} |\nu_{T,H,t}| \leq \max_{t=1,\dots,T} |\widetilde{\Delta}_{T,H,t}| + \max_{t=1,\dots,T} |g_t| \max_{t=1,\dots,T} |\upsilon_{T,H,t}|.$$

Then (91) follows applying to $\max_{t=1,\dots,T} |\widetilde{\Delta}_{T,H,t}|$ the bound (90) and using $\max_{t\geq 1} |g_t| \leq C$ and (i). In turn, (93) follows using (92), (iv) and (i).

(iv) is shown in Lemma A.3 (v) in DGK. □

Appendix B. Supplementary data

Supplementary material related to this article can be found online at [INSERT DOI]

References

Antoine, B., and O. Boldea (2018): "Efficient inference with time-varying information and the New Keynesian Phillips curve," *Journal of Econometrics*, 204(3), 268–300.

Bai, J., and P. Perron (1998): "Estimating and testing linear models with multiple structural changes," *Econometrica*, 66(1), 47–78.

Boldea, O., and A. R. Hall (2013): "Estimation and Inference in Unstable Nonlinear Least Squares Models," *Journal of Econometrics*, 172, 158–167.

Brown, R. L., J. Durbin, and J. M. Evans (1974): "Techniques for testing the constancy of regression relationships over time," *Journal of the Royal Statistical Association, Series A*, 138, 149–163.

Chen, B. (2015): "Modeling and testing smooth structural changes with endogenous regressors," *Journal of Econometrics*, 185, 196–215.

Chen, B., and Y. Hong (2012): "Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression," *Econometrica*, 80(3), 1157–1183.

Chow, A. (1960): "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, 28, 591–605.

Cogley, T., and T. J. Sargent (2005): "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8, 262–302.

Dahlhaus, R. (1997): "Fitting time series models to nonstationary processes," *Annals of Statistics*, 25, 1–37.

Dendramis, Y., L. Giraitis, and G. Kapetanios (2018): "Estimation of time-varying covaraince matrices for large data sets," *Preprint*.

Gali, J., and M. Gertler (2008): "Inflation Dynamics: A structural econometric approach," *Journal of Monetary Economics*, 44, 195–222.

Giraitis, L., G. Kapetanios, and T. Yates (2014): "Inference on stochastic time-varying coefficient models," *Journal of Econometrics*, 179, 46–65.

——— (2018): "Inference on multivariate heteroscedastic time varying random coefficient models," *Journal of Time Series Analysis*, 39(2), 129–149.

Hall, A. R., S. Han, and O. Boldea (2012): "Inference regarding multiple structural changes in linear models with endogenous regressors," *Journal of Econometrics*, 170, 281–302.

Hamilton, J. D. (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57(2), 357–384.

Kapetanios, G., and E. Tzavalis (2010): "Modeling structural breaks in economic relationships using large shocks," *Journal of Economic Dynamics and Control*, 34, 417–436.

Kapetanios, G., and T. Yates (2008): "An analysis of inflation persistence and the Great Moderation using a model of deterministic structural change," *Mimeo, Bank of England*.

Muller, U. K., and P. M. Petalas (2010): "Efficient estimation of the parameter path in unstable time series models," *Review of Economic Studies*, 77, 1508–1539.

Muller, U. K., and M. W. Watson (2008): "Testing models of low-frequency variability," *Econometrica*, 76, 979–1016.

Nyblom, J. (1989): "Testing for constancy of parameters over time," *Journal of the American Statistical Association*, 72, 223–230.

Ploberger, W., and W. Kramer (1992): "The CUSUM test with OLS residuals," *Econometrica*, 60, 271–285.

Priestley, M. (1965): "Evolutionary spectra and nonstationary processes," *Journal of Royal Statistical Society, Series B*, 27, 204–237.

Primiceri, G. (2005): "Time varying structural vector autoregressions and monetary policy," *Review of Economic Studies*, 72, 821–852.

Robinson, P. M. (1989): "Nonparametric estimation of time varying parameters," in *Economic Structural Change: Analysis and Forecasting*, ed. by P. Hackl, pp. 253–264. Springer Berlin.

——— (1991): "Time varying nonlinear regression," in *Economic Structural Change: Analysis and Forecasting*, ed. by P. Hackl, pp. 179–190. Springer Berlin.

Stock, J., and M. Watson (1989): "Phillips curve inflation forecasts," in *Understanding Inflation and the Implications for Monetary Policy, a Phillips Curve Retrospective*, ed. by F. R. B. of Boston. Federal Reserve Bank of Boston.

Terasvirta, T. (1998): "Modelling economic relationships with smooth transition regressions," in *Handbook of Applied Economic Statistics*, ed. by A. Ullah, and D. E. A. Giles, pp. 507–552. Marcel Dekker.

Tong, H. (1990): *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press.

Online Supplement to
"Time-Varying Instrumental Variable Estimation"

Liudas Giraitis[1], George Kapetanios[2] and Massimiliano Marcellino[3]

[1]Queen Mary University of London [2]King's College, London

[3]Bocconi University, IGIER and CEPR

This Supplement provides the proof of Theorem 4 of the main paper. It uses a number of technical lemmas, presented after the proof of the theorem. Section "Additional Tables" contains simulation results for higher values of the bandwidth parameters.

Formula numbering in this supplement is of the form, e.g., (A.1), and references to lemmas are signified as "Lemma #". Equation, lemma and theorem references to the main paper do not include the prefix "A", and are signified as "equation (#)", "Lemma #", "Theorem #".

Proofs for Hausman test

Proof of Theorem 4. (i) Denote $V_{T,t} = V_t = S^{1/2}_{\widehat{x},\widehat{x},t} S^{1/2}_{xx,t}(\widehat{\beta}_t - \widetilde{\beta}_t)$ where, for ease of notation, $\widetilde{\beta}_t = \widetilde{\beta}_{1,t}$. Write

$$\widehat{\beta}_t - \widetilde{\beta}_t = (\widehat{\beta}_t - \beta_t) - (\widetilde{\beta}_t - \beta_t) = S^{-1}_{xx,t} S_{xu,t} - S^{-1}_{\widehat{x}x,t} S_{\widehat{x}u,t} + R_t, \tag{A.1}$$

$$R_t = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|}\omega_t, \quad \omega_t = S^{-1}_{xx,t} x_j x_j' (\beta_j - \beta_t) - S^{-1}_{\widehat{x}\widehat{x},t} \widehat{x}_j x_j' (\beta_j - \beta_t).$$

Thus,

$$\begin{aligned} V_t &= V_{1,t} + V_{2,t}, \\ V_{1,t} &= S^{1/2}_{\widehat{x},\widehat{x},t} S^{1/2}_{xx,t} \{ S^{-1}_{xx,t} S_{xu,t} - S^{-1}_{\widehat{x}x,t} S_{\widehat{x}u,t} \}, \\ V_{2,t} &= S^{1/2}_{\widehat{x},\widehat{x},t} S^{1/2}_{xx,t} R_t. \end{aligned} \tag{A.2}$$

From (A.40),

$$||V_{2,t}|| = O((H/T)^{\gamma}) = o_p(1).$$

Similar arguments as those used in the proofs of Lemma 8 and 9 imply

$$||V_{1,t} - \Sigma^{1/2}_{x-v,x-v,t} \Sigma^{-1/2}_{xx,t} E[v_t u_t]|| = o_p(1)$$

which yields

$$V_t = \Sigma^{1/2}_{x-v,x-v,t}\Sigma^{-1/2}_{xx,t}E[v_t u_t] + o_p(1) = (1-\Sigma_{vv,t}\Sigma^{-1}_{xx,t})^{1/2}E[v_t u_t] + o_p(1).$$

This proves (34) of (i).

(ii) We split V_t into the main term and the remainder:

$$\sigma_t^{-1}\Sigma^{-1}_{vv,t}K_t K^{-1/2}_{2,t}V_t = U_t + r_t, \tag{A.3}$$
$$U_t = \sigma_t^{-1}\Sigma^{-1}_{vv,t}K_t K^{-1/2}_{2,t}V_{1,t},$$
$$r_t = \sigma_t^{-1}\Sigma^{-1}_{vv,t}K_t K^{-1/2}_{2,t}V_{2,t}.$$

From (A.40) and $K_t K^{-1/2}_{2,t} = O(H^{1/2})$ it follows

$$||r_t|| \le K_t K^{-1/2}_{2,t}||\Sigma^{-1}_{vv,t}||\,||V_{2,t}|| \le CH^{1/2}||V_{2,t}|| = O(H^{1/2}(H/T)^{1/2}) = o_p(1)$$

under assumptions of theorem imposed on H.

It remains to show that

$$U_t \to \mathcal{N}(0, I), \tag{A.4}$$

which proves (35).

Proof of (A.4). Recall notation $\Sigma_{vv,t} = E[v_t v_t']$, $\Sigma_{zz,t} = E[z_t z_t']$, $\Sigma_{x-v,x-v,t} = \Psi_t' E[z_t z_t']\Psi_t$, $\Sigma_{xx,t} = \Psi_t' E[z_t z_t']\Psi_t + \Sigma_{v,v,t}$. Set

$$L_{1,t} = S^{1/2}_{\widehat{x}\widehat{x},t}S^{-1/2}_{xx,t}, \quad L_{2,t} = S^{1/2}_{\widehat{x}\widehat{x},t}S^{1/2}_{xx,t}S^{-1}_{\widehat{x}\widehat{x},t}, \quad L_t = \Sigma^{1/2}_{x-v,x-v,t}\Sigma^{-1/2}_{xx,t}, \quad B_t = \sigma_t^{-1}\Sigma^{-1}_{vv,t}. \tag{A.5}$$

Write

$$U_t = B_t K^{-1/2}_{2,t}\sum_{j=1}^{T} b_{H,|j-t|}d_{t,j}u_j, \quad d_{t,j} = L_{1,t}x_j - L_{2,t}\widehat{x}_j. \tag{A.6}$$

We approximate U_t by

$$U_t^* = B_t K^{-1/2}_{2,t}\sum_{j=1}^{T} b_{H,|j-t|}p_{t,j}u_j, \quad p_{t,j} = L_t x_j - L_t^{-1}\Psi_j' z_j. \tag{A.7}$$

We will show

$$||U_t - U_t^*|| = o_p(1), \tag{A.8}$$

$$U_t^* \to_D \mathcal{N}(0, I) \tag{A.9}$$

which proves (A.4).

We start with the proof of (A.8). By assumption of theorem, $||B_t|| = O_p(1)$. Therefore

$$||U_t - U_t^*|| \le O(1)||K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} v_{t,j} u_j||, \quad v_{t,j} = d_{t,j} - p_{t,j}.$$

Since u_j are independent of $w_{t,j}$, we can estimate $||U_t - U_t^*||$ using Lemma 6. Bound

$$\begin{aligned} ||v_{t,j}|| &\le ||(L_{1,t} - L_t)x_j|| + ||(L_{2,t} - L_t^{-1})\widehat{x}_j|| + ||L_t^{-1}(\widehat{\Psi}_j - \Psi_j)z_j|| \qquad \text{(A.10)} \\ &\le ||L_{1,t} - L_t||\,||x_j|| + ||L_{2,t} - L_t^{-1}||\big(||\widehat{\Psi}_j - \Psi_j|| + ||\Psi_j||\big)||z_j|| + ||L_t^{-1}(\widehat{\Psi}_j - \Psi_j)||\,||z_j||. \end{aligned}$$

This together with (A.50) and (A.51) of Lemma 9 and Theorem 1 allows to bound

$$||v_{t,j}|| \le A_T^0\, r_{t,j} \tag{A.11}$$

where $A_T^0 = o_p(1)$, A_T does not depend on j, and $\max_{j=1,\dots,T} Er_{t,j}^2 = O(1)$. Thus by Lemma 6,

$$||U_t - U_t^*||^2 = O_p(A_T^{0\,2})O\big(K_{2,t}^{-1} \sum_{j=1}^{T} b_{H,|j-t|}^2 Er_{t,j}^2\big) = O_p(A_T^{0\,2}) = o_p(1).$$

This proves (A.8).

Next we show (A.9). In view of the Cramér-Wold theorem, it suffices to prove that for any $p \times 1$ vector b, it holds

$$b'U_t^* \to \mathcal{N}(0, ||b||^2). \tag{A.12}$$

Set $k_t = b'B_t$. Then $k_t p_{t,j}$ is a 1×1 random variable, and $(k_t p_{t,j})^2 = k_t p_{t,j} p_{t,j}' k_t'$. Write

$$b'U_t^* = K_{2,t}^{-1/2} \sum_{j=1}^{T} b_{H,|j-t|} b_{H,|j-t|} k_t p_{t,j} u_j.$$

Notice that u_t is a martingale difference sequence with respect to the sigma field $\mathcal{F}_t = \sigma(u_s, x_s, x_s - v_s \,:\, s \le t)$. So, by the standard argument of the central limit theorem for martingale differences, the normal approximation (A.12) holds if for some $\delta > 0$,

$$K_{2,t}^{-1} \sum_{j=1}^{T} b_{H,|j-t|}^2 (k_t p_{t,j})^2 E u_j^2 \to_p ||b||^2, \tag{A.13}$$

$$K_{2,t}^{-(2+\delta)/2} \sum_{j=1}^{T} \Big| b_{H,|j-t|} k_t p_{t,j} \Big|^{2+\delta} = o_p(1). \tag{A.14}$$

Denote by $j_{T,t}$ the l.h.s. of (A.13). Using notation $S_{xy,t}^{(2)} = K_{2,t}^{-1} \sum_{j=1}^{T} b_{H,|j-t|}^2 x_j y_j' \sigma_{u,j}^2$, write

$$\begin{aligned} j_{T,t} &= K_{2,t}^{-1} \sum_{j=1}^{T} b_{H,|j-t|}^2 k_t p_{t,j} p_{t,j}' k_t' \sigma_{u,j}^2 \qquad (A.15) \\ &= k_t F_{T,t} k_t', \quad F_{T,t} = L_t S_{xx,t}^{(2)} L_t + L_t^{-1} S_{x-v,x-v,t}^{(2)} L_t^{-1} - L_t S_{x,x-v,t}^{(2)} L_t^{-1} - L_t^{-1} S_{x-v,x,t}^{(2)} L_t. \end{aligned}$$

Denote

$$F_{T,t}^* = L_t \Sigma_{xx,t} L_t + L_t^{-1} \Sigma_{x-v,x-v,t} L_t^{-1} - L_t \Sigma_{x-v,x-v,t} L_t^{-1} - L_t^{-1} \Sigma_{x-v,x-v,t} L_t.$$

Notice that

$$\begin{aligned} F_{T,t}^* = \Sigma_{xx,t} - \Sigma_{x-v,x-v,t} = \Sigma_{vv,t}, \quad k_t F_{T,t}^* k_t' \sigma_{u,t}^2 = ||b||^2, \\ ||k_t (F_{T,t} - F_{T,t}^*) k_t'|| = 3||k_t||^2 \big(||L_t||^2 + ||L_t^{-1}||^2\big) \big(||S_{xx,t}^{(2)} - \Sigma_{xx,t}|| \\ + ||\Sigma_{x-v,x-v,t}|| + ||S_{x-v,x-v,t}^{(2)} - \Sigma_{x-v,x-v,t}|| \\ + ||S_{x,x-v,t}^{(2)} - \Sigma_{x-v,x-v,t}|| + ||S_{x-v,x,t}^{(2)} - \Sigma_{x-v,x-v,t}||\big) = o_p(1) \end{aligned}$$

by (A.48) and (A.49) Lemma 9. This proves (A.13).

Similar arguments can be used to verify (A.14). This proves Theorem 4(ii).

(iii) Next we prove (36). By (A.2), $V_t = V_{1,t} + V_{2,t}$. Write

$$\mathcal{H}_{T_0,T_1} = U_T^* + r_T^*, \tag{A.16}$$

$$U_T^* = \widetilde{T}^{-1/2} \sum_{t=T_0+1}^{T_1} K_t K_T^{-1} B_t V_{1,t}, \tag{A.17}$$

$$r_T^* = \widetilde{T}^{-1/2} \sum_{t=T_0+1}^{T_1} K_t K_T^{-1} B_t V_{2,t}.$$

We will approximate U_t^* by

$$I_T = \widetilde{T}^{-1/2} \sum_{j=T_0+1}^{T_1} w_j u_j, \quad w_j = \zeta_{1,j} x_j - \zeta_{2,j}(x_j - v_j) \tag{A.18}$$

where $\zeta_{1,t} = B_t L_t$ and $\zeta_{2,t} = B_t L_t^{-1}$. We will prove the asymptotic normality for I_T and show that r_T^* is a negligible term.

The following three relations together with (A.16) imply the claim (36) of Theorem 4:

$$||r_T^*|| = o_p(1), \tag{A.19}$$

$$||U_T^* - I_T|| = o_p(1), \tag{A.20}$$

$$I_T \to_D \mathcal{N}(0, I). \tag{A.21}$$

Proof of (A.19). From (A.40),

$$||V_{2,t}|| \le (H/T)^\gamma A_T q_t \tag{A.22}$$

where $A_t = O_p(1)$, $\max_{t=1,\dots,T} E||q_t||^2 = O(1)$ and $\gamma = 1$ if Assumption 2 holds, $\gamma = 1/2$ if Assumption 3 holds. From (A.22), $K_t K_T^{-1} = O(1)$ and assumption $\max_{t=1,\dots,T} ||\Sigma_{vv,t}^{-1}|| = O(1)$ it follows

$$||r_T^*|| \le C\widetilde{T}^{-1/2} \sum_{t=T_0+1}^{T_1} ||\Sigma_{vv,t}^{-1}|| \, ||V_{2,t}|| \le C\widetilde{T}^{-1/2} \sum_{t=T_0+1}^{T_1} ||V_{2,t}|| =$$
$$\le C(H/T)^\gamma A_T \widetilde{T}^{-1/2} \sum_{t=T_0+1}^{T_1} ||q_t|| = O_p\big((H/T)^\gamma \widetilde{T}^{1/2}\big) = o_p(1)$$

because

$$E\big[\widetilde{T}^{-1} \sum_{t=T_0+1}^{T_1} ||q_t||\big] = \widetilde{T}^{-1} \sum_{t=T_0+1}^{T_1} E||q_t|| = O(1)$$

and $(H/T)^\gamma \widetilde{T}^{1/2} = o(1)$ under assumptions on $H, \widetilde{T}$ related to (A.14). This completes the proof of (A.19).

Proof of (A.20). Recall that

$$U_T^* = \widetilde{T}^{-1/2} \sum_{j=1}^{T} w_{T,j} u_j, \quad w_{T,j} = \sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|t-j|} B_t d_{t,j} \tag{A.23}$$

$$I_T = \widetilde{T}^{-1/2} \sum_{j=T_0+1}^{T_1} w_j u_j, \quad w_j = \zeta_{1,j} x_j - \zeta_{2,j}(x_j - v_j), \tag{A.24}$$

where $d_{t,j}$ are defined as in (A.6).

To approximate $w_{T,j}$ by w_j, define

$$w_{T,j}^{(1)} = \sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|t-j|} B_t p_{t,j}$$

where $p_{t,j}$ are the same as in (A.7). Write $[1,...,T] = J_T \cup J_T^c$ where $J_T = (T_0, T_1]$, $J_T^c = [1, T_0] \cup [T_1+1, T]$. Set $K_j^* = \sum_{t=T_0+1}^{T_1} b_{H,|t-j|}$. For $j = 1, ..., T$, write

$$\begin{aligned} w_{T,j} - w_j &= \{w_{T,j} - w_{T,j}^{(1)}\} + \{w_{T,j}^{(1)} I(j \in J_T^c)\} \\ &\quad + \{(w_{T,j}^{(1)} - K_j^* K_T^{-1} w_j) I(j \in J_T)\} + \{(K_j^* K_T^{-1} - 1) w_j I(j \in J_T)\} \\ &=: d_{T,j}^{(1)} + d_{T,j}^{(2)} + d_{T,j}^{(3)} + d_{T,j}^{(4)}. \end{aligned}$$

To prove (A.20), it suffices to show for $i = 1, .., 4$,

$$||\widetilde{T}^{-1/2} \sum_{j=1}^{T} d_{T,j}^{(i)}|| = o_p(1). \tag{A.25}$$

Below we denote by A_T^0, A_T generic random variables that have property $A_T^0 = o_P(1)$, $A_T = O_p(1)$ and do not depend on j. We shall show that

$$||d_{T,j}^{(1)}|| \le A_T^0 r_{T,j}^{(1)}, \quad \widetilde{T}^{-1} \sum_{j=1}^{T} E r_{T,j}^{(1)\,2} = O(1), \tag{A.26}$$

$$||d_{T,j}^{(2)}|| \le A_T r_{T,j}^{(2)}, \quad \widetilde{T}^{-1} \sum_{j=1}^{T} E r_{T,j}^{(2)\,2} = o(1), \tag{A.27}$$

$$||d_{T,j}^{(3)}|| \le A_T^0 r_{T,j}^{(3)}, \quad \widetilde{T}^{-1} \sum_{j=1}^{T} E r_{T,j}^{(3)\,2} = O(1), \tag{A.28}$$

$$||d_{T,j}^{(4)}|| \le A_T r_{T,j}^{(4)}, \quad \widetilde{T}^{-1} \sum_{j=1}^{T} E r_{T,j}^{(4)\,2} = o(1). \tag{A.29}$$

By assumption, (u_j) is a sequence of independent variables which is mutually independent of $d_{T,j}^{(i)}$, $i = 1, ..., 4$. Hence, by Lemma 6 and (A.26),

$$||\widetilde{T}^{-1/2} \sum_{j=1}^{T} d_{T,j}^{(1)}||^2 = O_p\big(A_T^{0\,2}\big) O_p\big(\widetilde{T}^{-1} \sum_{j=1}^{T} E r_{T,j}^{(1)\,2}\big) = o_p(1)$$

which proves (A.25) for $i = 1$. For $i = 2, 3, 4$, (A.25) follows from (A.27)-(A.29) using Lemma 6.

Proof of (A.26). Set $m_T = \max_{t=1,\dots,T} ||B_t||$. Then

$$||d^{(1)}_{T,j}|| = ||w_{T,j} - w^{(1)}_{T,j}|| \le m_T \sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|j-t|} ||d_{t,j} - p_{t,j}||.$$

Under assumptions of theorem, $m_T = O_p(1)$. By (A.11),

$$||d_{t,j} - p_{t,j}|| \le A^0_T \, r_{t,j},$$

where $A^0_T = o_p(1)$, A^0_T does not depend on t, j and $\max_{t,j=1,\dots,T} Er^2_{t,j} = O(1)$. Hence

$$||d^{(1)}_{T,j}|| \le A^0_T r^{(1)}_{T,j}, \quad r^{(1)}_{T,j} = \sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|j-t|} r_{t,j}.$$

Together with (A.36) of Lemma 7, this implies

$$\widetilde{T}^{-1} \sum_{j=1}^{T} Er^{(1)\,2}_{T,j} \le \widetilde{T}^{-1} \sum_{j=1}^{T} \Big(\sum_{t=T_0+1}^{T_1} K_T^{-1} b^2_{H,|j-t|} \Big)^2 \big(\max_{t,j=1,\dots,T} Er^2_{t,j} \big) = O(1)$$

which verifies (A.26).

Proof of (A.27). We have

$$||d^{(2)}_{T,j}|| = ||w^{(1)}_{T,j} I(j \in J^c_T)|| \le I(j \in J^c_T) m_T \sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|j-t|} ||p_{t,j}||.$$

From definition of $p_{t,j}$, using (A.50) Lemma 9, it follows that

$$||p_{t,j}|| \le A_T \, r_{t,j},$$

where $A_T = O_p(1)$ does not depend on t, j and $\max_{t,j=1,\dots,T} Er^2_{t,j} = O(1)$. This yields

$$||d^{(2)}_{T,j}|| \le A_T r^{(2)}_{T,j}, \quad r^{(2)}_{T,j} = I(j \in J^c_T) \sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|j-t|} r_{t,j}.$$

Together with (A.37) of Lemma 7, this implies

$$\widetilde{T}^{-1} \sum_{j=1}^{T} Er^{(2)\,2}_{T,j} \le \widetilde{T}^{-1} \sum_{j \in J^c_T} \Big(\sum_{t=T_0+1}^{T_1} K_T^{-1} b^2_{H,|j-t|} \Big)^2 \big(\max_{t,j=1,\dots,T} Er^2_{t,j} \big) = o(1)$$

which verifies (A.27).

Proof of (A.28). From equality

$$w_{T,j}^{(1)} = (\sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|j-t|}\zeta_{1,t})x_j - (\sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|j-t|}\zeta_{2,t})(x_j - v_j).$$

and definition of K_j^* it follows

$$w_{T,j}^{(1)} - K_T^{-1}K_j^* w_j = c_{1,T,j}x_j - c_{2,T,j}(x_j - v_j), \tag{A.30}$$
$$c_{1,T,j} = \sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|j-t|}(\zeta_{1,t} - \zeta_{1,j}),$$
$$c_{2,T,j} = \textstyle\sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|j-t|}(\zeta_{2,t} - \zeta_{2,j}).$$

By (A.52) of Lemma 9, for $i = 1, 2$, we can bound

$$||c_{1,T,j}x_j|| \le A_T^0 r_{T,j}, \quad ||c_{2,T,j}(x_j - v_j)|| \le A_T^0 r_{T,j}$$

where $A_T^0 = o_P(1)$ does not depend on j, and $\max_{j=T_0+1,\ldots,T_1} Er_{T,j}^2 = O(1)$. Then

$$||d_{T,j}^{(3)}|| = ||(w_{T,j}^{(1)} - K_T^{-1}K_j^* w_j)I(j \in J_T)|| \le A_T r_{T,j}^{(4)}, \quad r_{T,j}^{(4)} = I(j \in J_T)\sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|j-t|} r_{T,j}.$$

This implies

$$\widetilde{T}^{-1}\sum_{j\in J_T} Er_{T,j}^{(4)\,2} \le \widetilde{T}^{-1}\sum_{j=T_0+1}^{T_1}\Big(\sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|j-t|}^2\Big)^2 (\max_{t,j=1,\ldots,T} Er_{T,j}^2) = O(1)$$

which verifies (A.28).

Proof of (A.29). Using (A.50) Lemma 9, we can bound $||w_j|| \le A_T r_j$ where $A_T = O_P(1)$ does not depend on j, and $\max_{j=T_0+1,\ldots,T_1} Er_j^2 = O(1)$. Then

$$||d_{T,j}^{(4)}|| = ||(K_T^{-1}K_j^* - 1)w_j)I(j \in J_T)|| \le A_T^0 r_{T,j}^{(4)}, \quad r_{T,j}^{(4)} = I(j \in J_T)|K_T^{-1}K_j^* - 1| r_j.$$

This implies

$$\widetilde{T}^{-1}\sum_{j\in J_T} Er_{T,j}^{(4)\,2} \le \widetilde{T}^{-1}\sum_{j=T_0+1}^{T_1} |K_j^* K_T^{-1} - 1|^2 (\max_{j=1,\ldots,T} Er_{T,j}^2) = o(1)$$

by (A.38) of Lemma 7. This verifies (A.29).

Proof of (A.19). By the same argument as in the proof of (A.12), it suffices to show that for any $p \times 1$ vector b, it holds

$$b' I_T \to \mathcal{N}(0, ||b||^2). \tag{A.31}$$

By the central limit theorem for martingale differences, it suffices to show that for some $\delta > 0$,

$$j_T = \widetilde{T}^{-1} \textstyle\sum_{j=T_0+1}^{T_1} ||b'(\zeta_{1,j} x_j - \zeta_{2,j}(x_j - v_j))||^2 E u_j^2 \to_p ||b||^2, \tag{A.32}$$

$$j_T^* = \widetilde{T}^{-(2+\delta)/2} \textstyle\sum_{j=1}^{T} ||b'(\zeta_{1,j} x_j - \zeta_{2,j}(x_j - v_j))||^{2+\delta} = o_p(1). \tag{A.33}$$

Write

$$j_T = b' \big[\widetilde{T}^{-1} \sum_{j=T_0+1}^{T_1} \sigma_{u,j}^2 \big(\zeta_{1,j} x_j x_j' \zeta_{1,j}' + \zeta_{2,j}(x_j - v_j)(x_j - v_j)' \zeta_{2,j}' \\ - \zeta_{1,j} x_j (x_j - v_j)' \zeta_{2,j}' - \zeta_{2,j}(x_j - v_j) x_j' \zeta_{2,j}'\big)\big] b.$$

We approximate it by

$$\begin{aligned}\widetilde{j}_T &= b' \big[\widetilde{T}^{-1} \sum_{j=T_0+1}^{T_1} \sigma_{u,j}^2 \big(\zeta_{1,j} \Sigma_{xx,j} \zeta_{1,j}' + \zeta_{2,j} \Sigma_{x-v,x-v,j} \zeta_{2,j}' \\ &\qquad - \zeta_{1,j} \Sigma_{x-v,x-v,j} \zeta_{2,j}' - \zeta_{2,j} \Sigma_{x-v,x-v,j} \zeta_{2,j}'\big)\big] b \\ &= b' \big[\widetilde{T}^{-1} \sum_{j=T_0+1}^{T_1} I\big] = ||b||^2.\end{aligned}$$

To prove (A.32), it suffices to show

$$j_t - \widetilde{j}_T = o_p(1).$$

This can be done by summation by parts, using properties (A.53) of $\zeta_{1,j},\ \zeta_{2,j}$ and the bounds $E||k^{-1/2} \sum_{j=T_0+1}^{T_0+k} (z_j z_j' - E[z_j z_j'])||^2 = O(1)$, $E||k^{-1/2} \sum_{j=T_0+1}^{T_0+k} (v_j v_j' - E[v_j v_j'])||^2 = O(1)$ that hold under Assumption 1.

Condition (A.33) can be verified using similar arguments. This proves (A.19) and completes the proof of (36) and the theorem. □

Technical lemmas

In the next lemma we consider the sum

$$\mathcal{S}_{T,t} = \sum_{j=1}^{T} w_{T,j} u_j$$

where $(w_{T,j})$ are random $p \times 1$ vectors and (u_t) are scalar zero mean random variables such that

$$\max_{k=1,\dots,T} T^{-1} \sum_{j=1}^{T} \big| E[u_k u_j] \big| = O(1).$$

We suppose that $t = t_T \in [1, ...T]$ may vary with T.

Lemma 6 *Suppose (u_j) is independent of $(w_{T,j})$. Assume we can bound*

$$||w_{T,j}|| \le A_T q_{T,j}, \quad j = 1, ..., T \tag{A.34}$$

where A_T does not depend on j and $Eq^2_{T,j} < \infty$. Then, as $T \to \infty$,

$$||\mathcal{S}_{T,t}||^2 = O_p(A_T^2)\, O_P(\textstyle\sum_{j=1}^{T} Eq^2_{T,j}). \tag{A.35}$$

Proof. Write $\mathcal{S}_{T,t} = A_T(A_T^{-1}\mathcal{S}_{T,t})$. We will show that $E||A_T^{-1}\mathcal{S}_{T,t}||^2 = O(\sum_{j=1}^{T} Eq^2_{T,j})$ which implies $||A_T^{-1}\mathcal{S}_{T,t}||^2 = O_p\Big(\sum_{j=1}^{T} Eq^2_{T,j}\Big)$ and proves (A.35). We have

$$E||A_T^{-1}\mathcal{S}_{T,t}||^2 \le \sum_{k,j=1}^{T} E\Big[A_T^{-2}||w_{T,k}||\,||w_{T,j}||\Big]\big|E[u_j u_s]\big|.$$

By (A.34),

$$E[A_T^{-2}||w_{T,k}||\,||w_{T,j}||] \le E[q_{T,k}q_{T,j}] \le E[q^2_{T,k} + q^2_{T,j}].$$

So,

$$\begin{aligned} E||A_T^{-1}\mathcal{S}_{T,t}||^2 &\le 2\sum_{j,k=1}^{T} Eq^2_{T,j}|E[u_j u_k]| \\ &\le 2(\sum_{j=1}^{T} Eq^2_{T,j})(\max_{j=1,\dots,T}\sum_{k=1}^{T}|E[u_j u_k]|) = O(\sum_{j=1}^{T} Eq^2_{T,j}) \end{aligned}$$

which completes the proof. □

Lemma 7 *Under Assumptions of Theorem 4,*

$$\tau_{1,T} = \widetilde{T}^{-1} \sum_{j=1}^{T} (\sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|j-t|})^2 = O(1), \tag{A.36}$$

$$\tau_{2,T} = \widetilde{T}^{-1} [\sum_{j=1}^{T_0} ... + \sum_{j=T_1+1}^{T}](\sum_{t=T_0+1}^{T_1} K_T^{-1} b_{H,|j-t|})^2 = o(1), \tag{A.37}$$

$$\tau_{3,T} = \widetilde{T}^{-1} \sum_{j=T_0}^{T_1} |K_j^* K_T^{-1} - 1| = o(1). \tag{A.38}$$

Proof. Recall that $K_T^{-1} = O(H^{-1})$. Then,

$$\tau_{1,T} = \widetilde{T}^{-1} \sum_{t=T_0+1}^{T_1} \{K_T^{-2} \sum_{j=1}^{T} b_{H,|j-t|} [\sum_{s=1}^{T} b_{H,|j-s|}]\} \leq \widetilde{T}^{-1} \sum_{t=T_0+1}^{T_1} C = O(1). \tag{A.39}$$

Next we evaluate $\tau_{2,T}$. Let v be a large number. Then

$$\tau_{2,T} \leq \widetilde{T}^{-1} [\sum_{j=1}^{T_0} + \sum_{j=T_1+1}^{T}]\Big[(\sum_{t=T_0+1:\, |t-j| \geq vH}^{T_1} K_T^{-1} b_{H,|j-t|})^2 + (\sum_{t=T_0+1:\, |t-j| < vH}^{T_1} K_T^{-1} b_{H,|j-t|})^2\Big]$$
$$= s_{1,T} + s_{2,T}.$$

Similarly as in (A.39) it follows

$$s_{1,T} \leq C H^{-1} \sum_{s=vH}^{\infty} b_{H,s} \leq \delta_v \to 0, \quad v \to \infty.$$

On the other hand, for any fixed v,

$$s_{2,T} \leq C \widetilde{T}^{-1} [\sum_{j=T_0-cH}^{T_0} ... + \sum_{j=T_1+1}^{T_1+vH}] 1 \leq C v H \widetilde{T}^{-1} \to 0$$

because, by assumption, $H = o(\widetilde{T})$.

Finally, (A.38) follows, noting that $\max_{j=T_0+1,...,T_1} |K_j^* K_T^{-1}| = O(1)$ and

$$\max_{j=T_0+vH,...,T_1} |K_j^* K_T^{-1} - 1| \to 0 \quad \text{as } H, v \to \infty$$

and noting that $H = o(\widetilde{T})$. □

Lemma 8 *Let assumptions of Theorem 4 be satisfied. Denote* $\gamma = 1$ *and* $\gamma = 1/2$ *if* β_t *satisfies Assumption 2 and Assumption 3, respectively. Then* $V_{2,T}$ *in (A.2) has property:*

$$||V_{2,t}|| \leq (H/T)^{\gamma} A_T q_t = O_p((H/T)^{\gamma}) \tag{A.40}$$

where A_T *does not depend on* t*,*

$$A_T = O_P(1), \quad \max_{t=1,\dots,T} Eq_t = O(1). \tag{A.41}$$

Proof. We have

$$||V_{2,t}| \leq ||S^{1/2}_{\widehat{x}\widehat{x},t}||_{sp}||S^{1/2}_{xx,t}||_{sp}||R_t|| = ||S_{\widehat{x}\widehat{x},t}||^{1/2}_{sp}||S_{xx,t}||^{1/2}_{sp}||R_t||.$$

We will show that

$$||R_t|| \leq (H/T)^{\gamma} A_{1,T} q_{1,t}, \quad ||S_{xx,t}||^{1/2}|| \leq A_{2,T} q_{2,t}, \tag{A.42}$$
$$\max_{t=1,\dots,T} E||S_{xx,t}||^2 = O(1),$$

where $A_{i,T} = O_p(1)$, $\max_{t=1,\dots,T} E|q_{i,t}|^2 = O(1)$, for $i = 1, 2$. Then

$$||V_{2,t}| \leq ||S_{\widehat{x}\widehat{x},t}||^{1/2}||S_{xx,t}||^{1/2}||R_t||$$
$$\leq (H/T) A_T q_t, \quad A_T = A_{1,T} A_{2,T}, \quad q_t = q^{1/2}_{2,t}||S_{xx,t}||^{1/2} q_{1,t}.$$

Clearly, $A_T = O_p(1)$, while

$$\max_{t=1,\dots,T} Eq_t^2 \leq \max_{t=1,\dots,T} \left(Eq^2_{1,t} + ||S_{xx,t}||^2 + Eq^2_{2,t}\right) = O(1).$$

This proves (A.40). □

Proof of (A.42). Recall $\widehat{x}_j = \widehat{\Psi}_j z_j$. Using the bound

$$||\widehat{\Psi}_j|| \leq ||\widehat{\Psi}_j - \Psi_j|| + ||\Psi_j|| \leq (||\widehat{\Psi}_j - \Psi_j|| + 1)(1 + ||\Psi_j||), \tag{A.43}$$

we can bound ω_t in (A.1) as

$$|\omega_t| \leq A_{1,t}\nu_t, \quad A_{1,T} = \max_{j=1,\dots,T}(||S^{-1}_{xx,t}|| + ||S^{-1}_{\widehat{x}\widehat{x},t}|| + ||\widehat{\Psi}_j - \Psi_j|| + 1),$$
$$\nu_t = (H/T)^{\gamma}||x_j||^2||(\beta_j - \beta_t)(T/H)^{\gamma}|| + (1 + ||\Psi_j||)||x'_j(\beta_j - \beta_t)(T/H)^{\gamma}||.$$

Then

$$R_t \leq (H/T)^\gamma A_{1,T} q_{1,t}, \quad q_{1,t} = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \nu_t.$$

(A.45) and Theorem 1 imply $A_{1,T} = O_p(1)$. On the other hand,

$$Eq_{1,t}^2 \leq E(K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \nu_j^2)^2 \tag{A.44}$$
$$\leq (\max_{j=1,\dots,T} E\nu_j^4)(K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|}) = \max_{j=1,\dots,T} E\nu_j^4 = O(1)$$

because under assumptions of lemma, uniformly in j, t,

$$E\nu_j^2 \leq E||x_j||^4 + 2E||(\beta_j - \beta_t)(T/H)^\gamma||^2 + E(1 + ||\Psi_j||)^2 + E||x_j||^2 \leq C.$$

This proves (A.42) for R_t.

To verify (A.42) for $||S_{\widehat{x}\widehat{x},t}||$, notice that by (A.43),

$$||\widehat{x}_j \widehat{x}_j'|| \leq ||\widehat{\Psi}_j||^2 ||z_j||^2 \leq (||\widehat{\Psi}_j - \Psi_j|| + 1)^2 (1 + ||\Psi_j||)^2 ||z_j||^2.$$

Therefore,

$$||S_{\widehat{x}\widehat{x},t}|| = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} ||\widehat{x}_j \widehat{x}_j'|| \leq A_{2,T} q_{2,T},$$
$$A_{2,T} = \max_{j=1,\dots,T} (||\widehat{\Psi}_j - \Psi_j|| + 1)^2, \quad q_{2,T} = K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} (1 + ||\Psi_j||)^2 ||z_j||^2.$$

Again, $A_{2,T} = O_p(1)$ by Theorem 1 while $\max_{j=1,\dots,T} Eq_{2,T} = O(1)$ follows using the same argument as in the proof of (A.44).

Finally, as in (A.44), we obtain

$$\max_{t=1,\dots,T} E||S_{xx,t}||^2 \leq \max_{t=1,\dots,T} E(K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} ||x_j||^2)^2 = O(1).$$

This completes the proof of (A.42) and (A.40). □

In Lemma 9 we consider bounds for random variables $L_{1,t}, L_{2,t}, L_t, S_{xx,t}^{(2)}, c_{1,T,j}, c_{2,T,j}$ indexed by $j = 1, ..., T$ defined in (A.5), (A.15) and (A.30). We denote by $A_T^0 = o_P(1)$, $A_T = O_p(1)$

generic random variables that not depend on $t = 1, ..., T$ and by $r_j = r_{T,t}$ random variables such that $\max_{t=1,...,T} Er_t^2 = O(1)$.

Lemma 9 *Let Assumption 1 holds, (β_t, Ψ_t) satisfy either Assumption 2 or Assumption 3. Assume that H satisfies (33) and L satisfies (17). Then, for $t \in \{1, ..., T\}$,*

$$\max_{t=1,...,T} ||S^{-1}_{xx,t}|| = O_p(1), \quad \max_{t=1,...,T} ||S^{-1}_{\widehat{x}\widehat{x},t}|| = O_p(1), \tag{A.45}$$

$$||S_{xx,t}|| \le A_T r_t, \quad ||S_{\widehat{x}\widehat{x},t}|| \le A_T r_t, \tag{A.46}$$

$$||S_{xx,t} - \Sigma_{xx,t}|| \le A^0_T r_t, \quad ||S_{\widehat{x}\widehat{x},t} - \Sigma_{x-v,x-v,t}|| \le A^0_T r_t, \tag{A.47}$$

$$||S^{(2)}_{xx,t} - \Sigma_{xx,t}|| \le A^0_T r_t, \quad ||S^{(2)}_{x-v,x-v,t} - \Sigma_{x-v,x-v,t}|| \le A^0_T r_t, \tag{A.48}$$

$$||S^{(2)}_{x,x-v,t} - \Sigma_{x-v,x-v,t}|| \le A^0_T r_t, \quad ||S^{(2)}_{x-v,x,t} - \Sigma_{x-v,x-v,t}|| \le A^0_T r_t, \tag{A.49}$$

$$||L_{1,t}|| \le A_T r_t, \quad ||L_{2,t}|| \le A_T r_t, \quad ||L_t|| \le A_T r_t, \quad ||L^{-1}_t|| \le A_T r_t, \tag{A.50}$$

$$||L_{1,t} - L_t|| \le A^0_T r_t, \quad ||L_{2,t} - L^{-1}_t|| \le A^0_T r_t, \tag{A.51}$$

$$||c_{1,T,j}|| \le A^0_T r_j, \quad ||c_{2,T,j}|| \le A^0_T r_j, \tag{A.52}$$

$$||(T/H)^\gamma (L_j - L_{j-1})|| \le A_T r_j, \quad ||(T/H)^\gamma (L^{-1}_j - L^{-1}_{j-1})|| \le A^0_T r_j, \tag{A.53}$$

where $\gamma = 1$ if Assumption 2 holds, $\gamma = 1/2$ if Assumption 3 holds.

Proof.

Proof of (A.45). To bound $||S^{-1}_{xx,t}||$, denote $\mathcal{V}_{T,t} = K^{-1}_t \sum_{j=1}^T b_{|j-t|}(\Psi'_j \Sigma_{zz,t} \Psi_j + \Sigma_{vv,j})$ where $\Sigma_{zz,j} = E[z_j z'_j]$, $\Sigma_{vv,j} = E[v_j v'_j]$. Proof of (57) shows that to verify (A.45) it suffices to show that

$$\max_{t=1,...,T} ||\mathcal{V}^{-1}_{T,t}||_{sp} = O_p(1), \tag{A.54}$$

$$\max_{t=1,...,T} ||S_{xx,t} - \mathcal{V}_{T,t}||_{sp} = o_p(1). \tag{A.55}$$

To prove (A.54), notice that $\Sigma_{zz,t}$ is semi positive definite, and $\Sigma_{vv,j}$ is positive definite. By Assumption 1(iii), there exists $\nu > 0$ such that $a'\Sigma_{vv,j} a \ge \nu$ for $j \ge 1$, $||a|| = 1$. Thus, for $||a|| = 1$,

$$a'(\Psi'_j \Sigma_{zz,j} \Psi_j + \Sigma_{vv,j})a = (\Psi_j a)' \Sigma_{zz,j} (\Psi_j a) + a' \Sigma_{vv,j} a \ge a' \Sigma_{vv,j} a \ge \nu.$$

This implies

$$a' \mathcal{V}_{T,L,t} a = \Big(K^{-1}_t \sum_{j=1}^T b_{H,|j-t|} a'(\Psi'_j \Sigma_{zz,j} \Psi_j + \Sigma_{vv,j}) a \Big) \ge \nu K^{-1}_t \sum_{j=1}^T b_{H,|j-t|} = \nu > 0$$

which proves (A.54).

(A.55) can be obtained using similar argument as in the proof of (58). This proves (A.45) for $||S^{-1}_{xx,t}||$. The claim for $||S^{-1}_{\widehat{x}\widehat{x},t}||$ can be shown as in the proof of (58).

Proof of (A.46) follows from (A.47) noting that $||\Sigma_{xx,t}||| \leq A_T$, $||\Sigma_{x-v,x-v,t}|| \leq A_T$.

Proof of (A.47)-(A.49) follows using similar argument as in the proof of (62).

Proof of(A.50) follows using (A.45), (A.46) and Assumption 1.

Proof of (A.51). We will show that $||L_{1,t} - L_t|| \leq A^0_T r_t$. Notice that

$$\begin{aligned} ||S^{1/2}_{\widehat{x}\widehat{x},t} - \Sigma^{1/2}_{x-v,x-v}|| &= ||(S^{1/2}_{\widehat{x}\widehat{x},t} - \Sigma^{1/2}_{x-v,x-v})(S^{1/2}_{\widehat{x}\widehat{x},t} + \Sigma^{1/2}_{x-v,x-v})(S^{1/2}_{\widehat{x}\widehat{x},t} + \Sigma^{1/2}_{x-v,x-v})^{-1/2}|| \\ &\leq ||S_{\widehat{x}\widehat{x},t} - \Sigma_{x-v,x-v}||\,||(S^{1/2}_{\widehat{x}\widehat{x},t} + \Sigma^{1/2}_{x-v,x-v})^{-1/2}|| \leq A_0 r_t \end{aligned}$$

by (A.47), noting that

$$||(S^{1/2}_{\widehat{x}\widehat{x},t} + \Sigma^{1/2}_{x-v,x-v})^{-1/2}||_{sp} = ||(S^{1/2}_{\widehat{x}\widehat{x},t} + \Sigma^{1/2}_{x-v,x-v})||^{-1/2}_{sp} \leq A_T$$

by (A.45). Similarly it follows that $||S^{1/2}_{xx,t} - \Sigma^{1/2}_{xx,t}|| \leq A^0_T r_t$, while (A.45)-(A.46) imply $||S^{-1/2}_{xx,t}|| \leq A_T$, $||S^{1/2}_{\widehat{x}\widehat{x},t}|| \leq A_T r_t$. Combining these bounds implies the first claim in (A.51). The proof of the second claim is similar.

Proof of (A.52)-(A.53) follows using similar argument as in the proof of (59) using properties of L_t. This completes the proof of the lemma. □

Additional tables. Monte Carlo results for higher bandwidth

In this Appendix we provide further Monte Carlo results for higher values of the bandwidth parameters H and L.

Table 9: Performance of estimators $\widehat{\beta}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ for the model (40)-(41) with exogenous x_t: $s = 0$, $H = T^{h_1}, L = T^{h_2}$.

h_1	h_2	T	Median Deviation			Abs. Median Deviation			Decile Range			Coverage Range		
			$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$
0.4	0.7	100	0.001	-0.011	-0.001	0.198	0.287	0.347	1.371	0.850	2.701	0.816	0.629	0.960
		200	0.001	-0.004	0.005	0.167	0.256	0.293	1.387	0.926	2.457	0.831	0.571	0.963
		400	-0.002	-0.003	0.001	0.139	0.234	0.238	1.359	0.981	2.136	0.836	0.515	0.964
		1000	-0.000	-0.004	-0.000	0.112	0.202	0.195	1.374	1.077	1.957	0.850	0.469	0.971
0.5	0.7	100	-0.002	-0.008	0.002	0.199	0.297	0.331	1.180	0.834	2.392	0.733	0.648	0.922
		200	0.000	-0.002	0.002	0.168	0.263	0.270	1.269	0.907	2.065	0.731	0.584	0.922
		400	-0.001	0.002	0.001	0.142	0.236	0.228	1.323	1.004	1.998	0.730	0.540	0.924
		1000	-0.000	0.004	0.001	0.114	0.205	0.178	1.293	1.050	1.702	0.726	0.479	0.925
0.7	0.4	100	0.002	-0.001	-0.007	0.245	0.354	0.365	0.772	2.720	2.011	0.502	0.961	0.816
	0.4	200	-0.000	0.001	0.003	0.224	0.295	0.317	0.858	2.420	1.858	0.440	0.965	0.783
	0.4	400	-0.002	-0.001	-0.010	0.204	0.242	0.269	0.944	2.118	1.487	0.381	0.964	0.742
	0.4	1000	-0.002	-0.000	-0.004	0.179	0.195	0.231	1.045	1.981	1.410	0.322	0.971	0.711
	0.5	100	-0.002	0.001	0.001	0.246	0.326	0.357	0.761	2.188	2.000	0.494	0.909	0.804
	0.5	200	0.001	0.000	0.001	0.221	0.266	0.302	0.861	1.983	1.618	0.439	0.906	0.772
	0.5	400	-0.002	0.001	-0.004	0.202	0.224	0.270	0.976	1.870	1.596	0.387	0.914	0.749
	0.5	1000	0.001	-0.001	-0.000	0.178	0.176	0.226	1.084	1.727	1.397	0.323	0.916	0.713
	0.7	100	0.000	0.010	0.008	0.242	0.314	0.355	0.747	0.958	2.036	0.507	0.704	0.822
	0.7	200	-0.003	-0.007	-0.010	0.225	0.281	0.307	0.868	0.991	1.684	0.432	0.626	0.770
	0.7	400	0.001	-0.003	-0.004	0.202	0.252	0.272	0.980	1.050	1.525	0.382	0.572	0.739
	0.7	1000	-0.001	0.002	0.002	0.178	0.215	0.232	1.063	1.120	1.448	0.325	0.518	0.732

Table 10: Performance of estimators $\widehat{\beta}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ for the model (40)-(41) with $s = 0.2$.

h_1	h_2	T	Median Deviation			Abs. Median Deviation			Decile Range			Coverage Range		
			$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$
0.4	0.7	100	0.021	0.003	0.003	0.175	0.256	0.211	1.307	0.754	1.484	0.764	0.470	0.845
		200	0.019	0.005	0.001	0.147	0.236	0.176	1.326	0.847	1.468	0.775	0.413	0.847
		400	0.021	0.005	0.002	0.124	0.211	0.147	1.344	0.961	1.451	0.781	0.369	0.844
		1000	0.021	0.004	0.002	0.098	0.186	0.115	1.356	1.061	1.432	0.794	0.308	0.852
0.5	0.7	100	0.021	0.008	0.004	0.186	0.260	0.214	1.161	0.751	1.317	0.651	0.472	0.760
		200	0.023	0.005	0.003	0.157	0.232	0.182	1.229	0.846	1.331	0.650	0.427	0.753
		400	0.022	0.006	0.003	0.135	0.213	0.153	1.283	0.954	1.354	0.641	0.370	0.738
		1000	0.021	-0.000	0.000	0.108	0.187	0.119	1.328	1.068	1.370	0.631	0.311	0.727
0.7	0.4	100	0.015	-0.003	-0.006	0.238	0.224	0.261	0.756	1.613	0.946	0.422	0.877	0.590
	0.4	200	0.018	0.001	-0.004	0.223	0.183	0.237	0.862	1.571	0.977	0.351	0.881	0.508
	0.4	400	0.018	-0.001	-0.001	0.201	0.155	0.211	0.958	1.523	1.044	0.314	0.889	0.459
	0.4	1000	0.021	0.001	0.002	0.178	0.121	0.185	1.057	1.490	1.110	0.257	0.897	0.387
	0.5	100	0.023	0.004	0.006	0.243	0.225	0.262	0.738	1.341	0.917	0.414	0.772	0.581
	0.5	200	0.021	-0.002	0.001	0.221	0.189	0.237	0.860	1.387	0.985	0.362	0.780	0.521
	0.5	400	0.022	0.002	0.002	0.201	0.157	0.211	0.951	1.363	1.030	0.310	0.764	0.450
	0.5	1000	0.021	0.001	0.002	0.180	0.124	0.185	1.080	1.399	1.126	0.254	0.762	0.381
	0.7	100	0.023	0.003	0.002	0.240	0.259	0.263	0.741	0.762	0.906	0.416	0.523	0.587
	0.7	200	0.021	0.001	0.002	0.222	0.238	0.237	0.854	0.868	0.986	0.357	0.453	0.518
	0.7	400	0.021	-0.004	-0.004	0.198	0.211	0.210	0.940	0.945	1.020	0.316	0.403	0.459
	0.7	1000	0.019	-0.002	-0.001	0.180	0.191	0.187	1.087	1.085	1.142	0.252	0.330	0.380

Table 11: Performance of estimators $\widehat{\beta}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ for the model (40)-(41) with $s = 0.5$.

h_1	h_2	T	Median Deviation			Abs. Median Deviation			Decile Range			Coverage Range		
			$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$
0.4	0.7	100	0.148	0.040	0.024	0.206	0.257	0.196	1.303	0.744	1.419	0.625	0.415	0.808
		200	0.144	0.028	0.015	0.184	0.236	0.162	1.335	0.840	1.422	0.604	0.363	0.817
		400	0.145	0.022	0.012	0.171	0.210	0.135	1.366	0.955	1.418	0.557	0.331	0.817
		1000	0.143	0.015	0.007	0.157	0.186	0.107	1.380	1.055	1.417	0.491	0.276	0.822
0.5	0.7	100	0.147	0.022	0.016	0.218	0.255	0.207	1.177	0.756	1.292	0.518	0.437	0.729
		200	0.141	0.016	0.007	0.195	0.234	0.171	1.232	0.835	1.305	0.477	0.374	0.709
		400	0.148	0.013	0.007	0.184	0.211	0.144	1.283	0.958	1.328	0.415	0.334	0.707
		1000	0.148	0.008	0.004	0.171	0.187	0.116	1.301	1.031	1.325	0.334	0.283	0.691
0.7	0.4	100	0.143	0.011	0.008	0.269	0.206	0.256	0.773	1.530	0.912	0.329	0.847	0.539
	0.4	200	0.138	0.004	0.004	0.250	0.170	0.230	0.865	1.487	0.963	0.275	0.853	0.473
	0.4	400	0.142	0.004	0.005	0.236	0.143	0.209	0.956	1.473	1.009	0.219	0.867	0.416
	0.4	1000	0.144	0.002	0.002	0.220	0.112	0.185	1.060	1.439	1.104	0.164	0.872	0.347
	0.5	100	0.138	0.006	0.004	0.266	0.212	0.255	0.763	1.281	0.911	0.332	0.739	0.541
	0.5	200	0.137	0.002	0.000	0.250	0.179	0.234	0.880	1.339	0.973	0.270	0.737	0.467
	0.5	400	0.145	0.004	0.002	0.237	0.151	0.211	0.964	1.366	1.036	0.221	0.744	0.426
	0.5	1000	0.139	-0.000	-0.003	0.218	0.121	0.185	1.056	1.367	1.097	0.165	0.732	0.347
	0.7	100	0.148	0.014	0.011	0.270	0.259	0.260	0.766	0.771	0.940	0.330	0.477	0.546
	0.7	200	0.147	0.011	0.007	0.255	0.237	0.235	0.828	0.820	0.937	0.264	0.410	0.476
	0.7	400	0.140	0.005	0.003	0.233	0.211	0.208	0.964	0.953	1.024	0.221	0.361	0.412
	0.7	1000	0.143	0.001	0.001	0.217	0.187	0.183	1.085	1.070	1.122	0.165	0.300	0.349

Table 12: Performance of estimators $\widehat{\beta}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ in the overidentified case for the model (41)-(42) with exogenous x_t: $s = 0$, $H = T^{h_1}, L = T^{h_2}$.

h_1	h_2	T	Median Deviation			Abs. Median Deviation			Decile Range			Coverage Range		
			$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$
0.4	0.7	100	0.003	0.006	0.006	0.182	0.259	0.228	1.339	0.775	1.578	0.785	0.510	0.869
		200	-0.001	-0.001	-0.002	0.152	0.234	0.194	1.328	0.850	1.502	0.798	0.457	0.871
		400	-0.000	-0.000	0.001	0.127	0.211	0.159	1.357	0.954	1.484	0.803	0.402	0.869
		1000	-0.000	-0.001	-0.000	0.101	0.187	0.128	1.351	1.043	1.450	0.824	0.352	0.877
0.5	0.7	100	-0.000	0.005	0.005	0.188	0.257	0.229	1.147	0.759	1.353	0.691	0.534	0.811
		200	-0.003	-0.002	-0.001	0.161	0.238	0.195	1.236	0.848	1.377	0.681	0.464	0.795
		400	0.001	0.001	0.001	0.136	0.212	0.163	1.306	0.978	1.404	0.675	0.425	0.783
		1000	-0.000	0.000	0.001	0.109	0.190	0.128	1.339	1.080	1.400	0.673	0.356	0.772
0.7	0.4	100	0.004	0.006	0.008	0.245	0.249	0.275	0.749	1.733	0.946	0.439	0.902	0.632
	0.4	200	0.001	-0.001	0.001	0.222	0.205	0.244	0.863	1.665	1.017	0.385	0.907	0.568
	0.4	400	-0.002	0.000	-0.001	0.201	0.170	0.215	0.926	1.542	1.020	0.337	0.908	0.505
	0.4	1000	0.001	0.000	-0.000	0.178	0.132	0.187	1.034	1.498	1.081	0.275	0.916	0.418
	0.5	100	0.003	0.001	0.004	0.242	0.236	0.274	0.742	1.384	0.954	0.442	0.812	0.625
	0.5	200	0.002	0.003	0.003	0.220	0.198	0.239	0.856	1.413	1.004	0.386	0.820	0.571
	0.5	400	0.000	-0.001	-0.001	0.199	0.165	0.212	0.925	1.380	1.019	0.336	0.804	0.502
	0.5	1000	0.000	-0.000	-0.000	0.178	0.133	0.188	1.045	1.388	1.092	0.279	0.809	0.434
	0.7	100	-0.006	-0.007	-0.007	0.240	0.264	0.267	0.754	0.785	0.947	0.443	0.566	0.630
	0.7	200	0.001	0.002	0.001	0.219	0.240	0.239	0.848	0.863	0.980	0.389	0.509	0.573
	0.7	400	-0.003	-0.002	-0.002	0.201	0.217	0.216	0.966	0.971	1.053	0.336	0.444	0.501
	0.7	1000	-0.001	-0.001	-0.001	0.176	0.189	0.186	1.053	1.053	1.103	0.279	0.379	0.430

Table 13: Performance of estimators $\widehat{\beta}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ in the overidentified case for the model (41)-(42) with $s = 0.2$.

h_1	h_2	T	Median Deviation			Abs. Median Deviation			Decile Range			Coverage Range		
			$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$
0.4	0.7	100	0.033	0.010	0.001	0.196	0.283	0.320	1.335	0.796	2.526	0.801	0.584	0.946
		200	0.029	0.000	0.004	0.162	0.251	0.254	1.344	0.877	2.197	0.812	0.525	0.947
		400	0.035	0.010	0.007	0.138	0.228	0.215	1.370	0.984	1.998	0.813	0.482	0.953
		1000	0.033	0.005	0.002	0.111	0.199	0.174	1.358	1.058	1.841	0.830	0.428	0.962
0.5	0.7	100	0.033	0.002	0.001	0.198	0.288	0.300	1.222	0.812	2.223	0.705	0.597	0.902
		200	0.033	0.011	0.008	0.168	0.256	0.248	1.231	0.862	1.899	0.702	0.538	0.903
		400	0.032	0.006	0.006	0.144	0.232	0.209	1.310	0.992	1.873	0.698	0.491	0.904
		1000	0.031	0.002	0.002	0.115	0.201	0.160	1.297	1.052	1.650	0.689	0.423	0.904
0.7	0.4	100	0.026	-0.001	-0.004	0.246	0.308	0.338	0.768	2.440	1.773	0.467	0.943	0.768
	0.4	200	0.032	0.001	0.003	0.224	0.256	0.295	0.848	2.139	1.569	0.408	0.947	0.737
	0.4	400	0.030	0.002	-0.002	0.204	0.213	0.257	0.936	1.961	1.403	0.359	0.953	0.705
	0.4	1000	0.032	0.001	0.002	0.180	0.171	0.221	1.026	1.816	1.312	0.304	0.960	0.690
	0.5	100	0.034	-0.001	-0.000	0.244	0.296	0.337	0.757	2.009	1.835	0.480	0.894	0.778
	0.5	200	0.034	0.003	0.006	0.224	0.249	0.295	0.854	1.920	1.585	0.416	0.891	0.738
	0.5	400	0.033	0.002	-0.004	0.204	0.209	0.260	0.944	1.766	1.381	0.363	0.891	0.718
	0.5	1000	0.034	0.002	0.005	0.183	0.163	0.223	1.047	1.618	1.350	0.296	0.889	0.679
	0.7	100	0.034	0.006	0.004	0.248	0.302	0.340	0.763	0.881	1.902	0.469	0.650	0.780
	0.7	200	0.030	-0.002	-0.005	0.225	0.269	0.293	0.866	0.947	1.624	0.406	0.575	0.726
	0.7	400	0.035	0.004	0.003	0.204	0.241	0.259	0.958	1.014	1.455	0.359	0.520	0.710
	0.7	1000	0.032	0.001	0.002	0.181	0.210	0.222	1.050	1.074	1.331	0.300	0.461	0.695

Table 14: Performance of estimators $\widehat{\beta}_t$, $\widetilde{\beta}_{1,t}$ and $\widetilde{\beta}_{2,t}$ in the overidentified case for the model (41)-(42) with $s = 0.5$.

h_1	h_2	T	Median Deviation			Abs. Median Deviation			Decile Range			Coverage Range		
			$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$	$\widehat{\beta}_t$	$\widetilde{\beta}_{1,t}$	$\widetilde{\beta}_{2,t}$
0.4	0.7	100	0.246	0.051	0.029	0.278	0.278	0.280	1.328	0.761	2.232	0.579	0.525	0.930
		200	0.242	0.039	0.012	0.262	0.251	0.231	1.348	0.861	2.015	0.533	0.469	0.935
		400	0.245	0.032	0.011	0.257	0.228	0.193	1.372	0.967	1.868	0.459	0.427	0.940
		1000	0.247	0.028	0.008	0.253	0.199	0.152	1.383	1.073	1.731	0.363	0.368	0.949
0.5	0.7	100	0.248	0.031	0.014	0.288	0.286	0.284	1.195	0.797	1.983	0.474	0.550	0.881
		200	0.250	0.033	0.012	0.274	0.254	0.234	1.280	0.887	1.854	0.405	0.491	0.886
		400	0.250	0.018	0.012	0.268	0.226	0.193	1.309	0.943	1.706	0.324	0.445	0.884
		1000	0.251	0.011	0.004	0.263	0.201	0.152	1.327	1.051	1.568	0.231	0.383	0.882
0.7	0.4	100	0.238	0.027	0.013	0.318	0.283	0.329	0.764	2.192	1.595	0.313	0.927	0.737
	0.4	200	0.240	0.020	0.006	0.304	0.236	0.288	0.863	2.013	1.482	0.250	0.934	0.707
	0.4	400	0.238	0.009	0.004	0.292	0.193	0.255	0.978	1.859	1.387	0.193	0.940	0.686
	0.4	1000	0.241	0.008	0.001	0.284	0.155	0.219	1.091	1.774	1.335	0.132	0.951	0.652
	0.5	100	0.247	0.019	0.013	0.321	0.280	0.325	0.772	1.921	1.699	0.312	0.863	0.746
	0.5	200	0.235	0.012	0.005	0.301	0.231	0.283	0.856	1.768	1.462	0.252	0.870	0.708
	0.5	400	0.235	0.005	-0.003	0.290	0.193	0.253	0.965	1.649	1.355	0.195	0.871	0.680
	0.5	1000	0.246	0.005	0.003	0.285	0.155	0.217	1.077	1.551	1.300	0.132	0.878	0.669
	0.7	100	0.235	0.000	-0.009	0.315	0.296	0.326	0.781	0.885	1.645	0.322	0.609	0.742
	0.7	200	0.245	0.013	0.010	0.306	0.263	0.285	0.835	0.905	1.465	0.243	0.539	0.706
	0.7	400	0.233	0.003	-0.000	0.289	0.235	0.254	0.960	0.988	1.373	0.193	0.479	0.669
	0.7	1000	0.245	0.003	0.002	0.286	0.205	0.218	1.080	1.077	1.363	0.131	0.421	0.654

Table 15: Rejection frequencies for the local Hausman test at $t = T/2$. Model: (40)-(41)

s	h_1	h_2	$T = 100$	$T = 200$	$T = 400$	$T = 1000$
0	0.4	0.7	0.129	0.159	0.178	0.202
	0.5	0.7	0.107	0.177	0.181	0.205
	0.7	0.4	0.316	0.333	0.376	0.434
	0.7	0.5	0.267	0.305	0.316	0.428
	0.7	0.7	0.022	0.039	0.059	0.067
0.2	0.4	0.7	0.181	0.225	0.257	0.269
	0.5	0.7	0.147	0.222	0.238	0.287
	0.7	0.4	0.363	0.428	0.431	0.528
	0.7	0.5	0.279	0.347	0.435	0.505
	0.7	0.7	0.044	0.065	0.079	0.125
0.5	0.4	0.7	0.280	0.320	0.346	0.401
	0.5	0.7	0.258	0.343	0.428	0.510
	0.7	0.4	0.443	0.470	0.497	0.551
	0.7	0.5	0.430	0.519	0.568	0.627
	0.7	0.7	0.336	0.562	0.725	0.841

Table 16: Rejection frequencies for the global Hausman test. Model: (40)-(41)

s	h_1	h_2	$T = 100$	$T = 200$	$T = 400$	$T = 1000$
0	0.4	0.7	0.054	0.066	0.064	0.126
	0.5	0.7	0.026	0.052	0.058	0.092
	0.7	0.4	0.060	0.100	0.134	0.122
	0.7	0.5	0.056	0.066	0.084	0.088
	0.7	0.7	0.010	0.026	0.026	0.058
0.2	0.4	0.7	0.078	0.080	0.116	0.196
	0.5	0.7	0.038	0.072	0.112	0.178
	0.7	0.4	0.142	0.156	0.204	0.234
	0.7	0.5	0.082	0.094	0.166	0.200
	0.7	0.7	0.026	0.058	0.084	0.160
0.5	0.4	0.7	0.460	0.730	0.912	0.996
	0.5	0.7	0.426	0.762	0.936	0.990
	0.7	0.4	0.380	0.532	0.680	0.820
	0.7	0.5	0.338	0.550	0.726	0.840
	0.7	0.7	0.426	0.700	0.878	0.984