Journal of Time Series Analysis

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Automated Bandwidth Selection for Inference in Linear Models With Time-Varying Coefficients

Charisios Grivas[1] | Zacharias Psaradakis[2]

[1]Department of Mathematical Sciences, Aalborg University, Denmark | [2]Birkbeck, University of London, London, UK

Correspondence: Charisios Grivas (charisios.grivas@gmail.com)

Received: 12 October 2024 | **Revised:** 16 March 2025 | **Accepted:** 22 April 2025

Funding: The authors received no specific funding for this work.

Keywords: data-driven bandwidth | instrumental variables | kernel smoothing | time-varying coefficients

ABSTRACT

The problem of selecting the smoothing parameter, or bandwidth, for kernel-based estimators of time-varying coefficients in linear models with possibly endogenous explanatory variables is considered. We examine automated bandwidth selection by means of cross-validation, a nonparametric variant of Akaike's information criterion, and bootstrap procedures based on wild bootstrap and dependent wild bootstrap resampling schemes. Our simulations show that data-driven selectors based on cross-validation and the dependent wild bootstrap are the most successful overall in a variety of settings that are relevant in econometrics. Empirical examples illustrate the practical use of the automated procedures.

1 | Introduction

Structural change and parameter instability are pervasive in relationships among economic and financial variables. To account for such instability in cases where change is considered to be relatively smooth rather than abrupt, various models with smoothly time-varying coefficients have been proposed, along with suitable methods for inference on the coefficient path. These include locally linear models with parameters that vary in a continuous manner according to the values of observable variables (e.g., Teräsvirta 1998), models with deterministic coefficients that are smooth functions of a rescaled time index (e.g., Robinson 1989, 1991; Cai 2007; Zhang and Wu 2012; Chen 2015), and models with stochastic coefficients evolving as multivariate ARIMA processes (e.g., Nicholls and Pagan 1985).

In more recent work, Giraitis et al. (2021) (GKM hereafter) consider linear models in which little structure is imposed on their time-varying coefficients—the latter may be deterministic or stochastic, subject only to certain smoothness and boundedness conditions. In addition, GKM allows the explanatory variables in the model to be potentially endogenous, in the sense of being correlated with the unobservable errors, a setting which, like that of Chen (2015), is often relevant in econometrics. When a set of instrumental variables (IV) is available, inference on the time-varying coefficients may be based on the kernel IV estimators proposed by GKM. The obvious advantage of estimators based on local smoothing is that they do not rely on parametric specifications for the time-dependence of the parameters. However, as is the case with all kernel-based smoothing techniques, the practical use of kernel IV or least-squares (LS) estimators requires the choice of a smoothing parameter, known as the bandwidth, as well as a choice of a suitable kernel function—although it is generally accepted that the former choice has by far the biggest impact on the properties of kernel smoothers in terms of bias–variance trade-off.

In the context of nonparametric regression with deterministic or random (and exogenous) explanatory variables, several automated, data-driven bandwidth selection methods have

The authors are grateful to two anonymous referees for helpful comments and suggestions.

been proposed for popular kernel-type estimators such as local polynomial estimators and estimators of the Nadaraya–Watson, Priestley–Chao and Gasser–Müller type. Those most commonly used are based on cross-validation (CV) methods, undersmoothing-penalized goodness-of-fit criteria such as, for example, Akaike's information criterion (AIC) and Rice's T-criterion, bootstrap resampling methods, and so-called plug-in rules—a useful overview can be found in Köhler et al. (2014). However, as already indicated, the properties of these data-driven bandwidth selection methods have almost exclusively been studied in regression settings where the explanatory variables are uncorrelated with or independent of the unobservable errors (or even deterministic). It is, therefore, of interest to examine whether automated selectors, which are known to provide effective bandwidth choices under exogeneity (or fixed-design) conditions, remain successful in the presence of endogeneity, and whether the performance of such selectors is affected by the strength of the correlation between explanatory variables and errors.

Our objective in this paper is to investigate some of these issues by considering the performance of several automated bandwidth selection methods for kernel IV and LS estimators in a general setting similar to that in GKM, that is, in linear models with time-varying coefficients and explanatory variables which may be endogenous for the parameters of interest. More specifically, we consider automated bandwidth selection by means of four different methods, namely, ordinary (leave-one-out) CV, a nonparametric variant of a bias-corrected version of AIC, and wild bootstrap (WB), and dependent wild bootstrap (DWB) procedures. The models considered are quite general, having stochastically varying coefficients, explanatory variables that may be endogenous, and errors which may be conditionally heteroskedastic and/or serially correlated. We find that DWB and, rather remarkably, ordinary CV provides effective choices of the bandwidth under a variety of conditions that are relevant in econometrics. These data-driven selectors provide a useful and easy-to-implement way to overcome the hurdle of choosing bandwidths in the practical application of kernel IV estimators of time-varying coefficients like those proposed by Chen (2015) and GKM.

The remainder of the paper is organized as follows. Section 2 introduces the model and related nonparametric kernel estimators of interest. Section 3 provides a detailed description of our data-driven procedures for the selection of the bandwidth parameter for IV and LS estimators. Section 4 provides a simulation study of the small-sample performance of automated bandwidth selectors under a variety of data-generating mechanisms. Section 5 illustrates the practical use of the automated selection procedures in the context of two empirical applications. Finally, Section 6 summarizes and concludes.

2 | Model and Estimation

Consider the varying-coefficient linear model given by

$$y_t = \beta_t' x_t + u_t, \quad t = 1, 2, \ldots, T \tag{1}$$

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

where y_t is a scalar variable, x_t is a $p \times 1$ vector of (potentially endogenous) variables, β_t is a $p \times 1$ vector of coefficients, z_t is an $n \times 1$ vector of exogenous variables ($n \geq p$), Ψ_t is an $n \times p$ matrix of coefficients, and u_t and v_t are zero-mean random errors (which may be serially correlated and/or heteroskedastic). As in GKM, x_t is considered to be endogenous for β_t when $E(v_t u_t) \neq 0$ for some t, whereas exogeneity of z_t is taken to mean that $E(z_t u_t) = 0$ and $E(z_t v_t') = 0$ for all t. The parameters β_t and Ψ_t may be deterministic or stochastic, satisfying suitable boundedness and smoothness conditions (see Giraitis et al. 2014 and GKM for details and examples).

For the model (1) and (2), the kernel IV estimator of β_t introduced by GKM is

$$\tilde{\beta}_t = \left(\sum_{j=1}^{T} b_{H,|j-t|} \hat{\Psi}_j' z_j x_j' \right)^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \hat{\Psi}_j' z_j y_j \tag{3}$$

where $b_{H,|j-t|}$ are kernel weights, H is a bandwidth parameter, and $\hat{\Psi}_j$ is a consistent estimator of Ψ_j.[1] A natural choice for the latter is the kernel LS estimator

$$\hat{\Psi}_t = \left(\sum_{j=1}^{T} b_{L,|j-t|} z_j z_j' \right)^{-1} \sum_{j=1}^{T} b_{L,|j-t|} z_j x_j' \tag{4}$$

with bandwidth parameter $L \geq H$. The kernel weights in Equations (3) and (4) are obtained from a nonnegative kernel function $K(\cdot)$ via $b_{M,l} = K(l/M)$, for some $M > 0$ such that $M \to \infty$ and $M/T \to 0$ as $T \to \infty$. Admissible kernel functions are those satisfying $K(w) \leq C/(1 + w^a)$ and $|(\mathrm{d}/\mathrm{d}w)K(w)| \leq C/(1 + w^a)$ for $w > 0$ and some $C > 0$ and $a > 3$; examples include $K(w) \propto \exp(-w^2/2)$, $K(w) \propto \mathbb{I}(0 \leq w < 1)$ and $K(w) \propto (1 - w)\mathbb{I}(0 \leq w < 1)$, where $\mathbb{I}(\cdot)$ is the indicator function.

GKM gives conditions on the dependence, heterogeneity, and moments of z_t, u_t and v_t, and on the variation in β_t and Ψ_t, which guarantee consistency and asymptotic normality of $\tilde{\beta}_t$. In the case where x_t is exogenous, in the sense that $E(v_t u_t) = 0$ for all t, β_t can also be consistently estimated using the kernel LS estimator

$$\hat{\beta}_t = \left(\sum_{j=1}^{T} b_{H,|j-t|} x_j x_j' \right)^{-1} \sum_{j=1}^{T} b_{H,|j-t|} x_j y_j \tag{5}$$

(Throughout the paper, H is used as a generic notation for the bandwidth parameter associated with an estimator of β_t, without implying that $\tilde{\beta}_t$ and $\hat{\beta}_t$ share the same bandwidth.)

The key issue that arises in the use of the estimators (3), (4), and (5) in practice is the selection of reasonable values for the bandwidth parameters H and L for a given sample size T. The choice is important because the finite-sample properties of the estimators can be affected significantly by the bandwidth value. For example, too small a value for H and/or L may yield undersmoothed estimates which have high variance, while too large a value may result in oversmoothing and large bias. The asymptotic results in GKM offer little practical guidance beyond the requirement that $C_1 T^{(4/\vartheta)+\kappa} \leq H \leq L \leq C_2 T^{1-\kappa}$ for some $\kappa, C_1, C_2 > 0$ and $\vartheta > 4$ such that $E(\|\omega_t\|^{4+\vartheta}) \leq \overline{C} < \infty$ uniformly in t, where

$\omega_t' = (u_t, v_t', z_t')$ and $\|\cdot\|$ is the Euclidean norm.[2] For practical use, it is, therefore, desirable to have data-driven rules for choosing the values of the bandwidth parameters.

3 | Data-Driven Bandwidth Selection

In this section, we discuss different methods for selecting the bandwidths L and H that are required for the construction of the kernel IV and LS estimator of β_t. The data-driven selectors considered are based on CV, AIC, WB, and DWB methods.

Throughout the remainder of the paper, we consider bandwidths of the form $L = T^{h_1}$ and $H = T^{h_2}$, with $0 < h_2 \le h_1 < 1$. For any $h \in (0, 1)$, we use $\hat{\Psi}_{t,h}$, $\tilde{\beta}_{t,h}$ and $\hat{\beta}_{t,h}$ to denote, respectively, the LS estimator of Ψ_t defined in Equation (4) with $L = T^h$, the IV estimator of β_t defined in Equation (3) with $H = T^h$, and the LS estimator of β_t defined in Equation (5) with $H = T^h$.

3.1 | Cross-Validation

CV is a widely used method for selecting the smoothing parameter for nonparametric estimators. The basic idea is to use part of the data for fitting and the remaining part to estimate the average squared error of the fitted model under different bandwidths, and select the bandwidth that produces the best performance. Automated CV-based bandwidth selectors for inference in varying-coefficient models have been used by Chen and Hong (2012), Zhang and Wu (2012), and Chen (2015), among others, the latter in the context of nonparametric two-stage LS estimation.

In our IV setting, letting $\hat{\Psi}_{(-t),h}$ be the leave-one-out version of the LS estimator of Ψ_t given by

$$\hat{\Psi}_{(-t),h} = \left(\sum_{1 \le j \le T, j \ne t} b_{T^h, |j-t|} z_j z_j' \right)^{-1} \sum_{1 \le j \le T, j \ne t} b_{T^h, |j-t|} z_j x_j'$$

TABLE 1 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators for model (6) with $T = 100$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 2$)	DWB ($\lambda = 4$)	DWB ($\lambda = 6$)	DWB ($\lambda = 8$)	DWB ($\lambda = 10$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.084	1.147	1.118	1.110	1.094	1.080	1.068	1.068
		$\tilde{\beta}_t$	1.042	0.871	1.110	1.100	1.080	1.066	1.050	1.050
		$\hat{\beta}_t$	1.082	1.143	1.119	1.104	1.086	1.076	1.066	1.064
	Absolute Median Deviation	$\tilde{\beta}_t$	0.245	2.332	0.245	0.257	0.258	0.258	0.258	0.258
		$\hat{\beta}_t$	0.153	0.151	0.153	0.153	0.153	0.154	0.153	0.154
	Coverage	$\tilde{\beta}_t$	84.556	68.566	83.439	78.670	78.995	79.196	79.318	79.432
		$\hat{\beta}_t$	71.340	70.627	70.585	71.036	71.313	71.397	71.684	71.817
	Optimal Coverage	$\tilde{\beta}_t$	86.495							
		$\hat{\beta}_t$	78.039							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.091	1.148	1.123	1.114	1.095	1.084	1.072	1.072
		$\tilde{\beta}_t$	1.051	0.886	1.113	1.101	1.079	1.066	1.052	1.052
		$\hat{\beta}_t$	1.091	1.150	1.136	1.125	1.103	1.087	1.084	1.076
	Absolute Median Deviation	$\tilde{\beta}_t$	0.221	2.099	0.221	0.237	0.238	0.237	0.236	0.237
		$\hat{\beta}_t$	0.148	0.386	0.147	0.148	0.148	0.148	0.149	0.149
	Coverage	$\tilde{\beta}_t$	83.006	66.840	82.251	77.255	75.602	77.493	78.029	78.027
		$\hat{\beta}_t$	70.384	69.982	69.986	69.860	70.192	70.579	70.651	70.795
	Optimal Coverage	$\tilde{\beta}_t$	85.138							
		$\hat{\beta}_t$	77.187							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.114	1.160	1.148	1.137	1.123	1.108	1.093	1.093
		$\tilde{\beta}_t$	1.081	0.881	1.137	1.124	1.105	1.091	1.074	1.074
		$\hat{\beta}_t$	1.027	1.070	1.066	1.053	1.034	1.018	1.016	1.006
	Absolute Median Deviation	$\tilde{\beta}_t$	0.204	2.015	0.204	0.217	0.219	0.218	0.218	0.218
		$\hat{\beta}_t$	0.243	0.242	0.242	0.243	0.243	0.243	0.243	0.243
	Coverage	$\tilde{\beta}_t$	80.993	63.472	80.404	75.568	75.602	76.123	76.334	76.411
		$\hat{\beta}_t$	48.075	47.376	47.302	47.543	47.960	48.283	48.329	48.552
	Optimal Coverage	$\tilde{\beta}_t$	83.631							
		$\hat{\beta}_t$	50.951							

the CV choice of L is $\tilde{L}_{\text{CV}} = T^{\tilde{h}_1}$, where

$$\tilde{h}_1 = \arg\min_h \left\{ \sum_{t=1}^{T} ||x_t - \hat{\Psi}'_{(-t),h} z_t||^2 \right\}$$

In a similar manner, letting $\tilde{\beta}_{(-t),h}$ be the leave-one-out version of the IV estimator of β_t, constructed as

$$\tilde{\beta}_{(-t),h} = \left(\sum_{1\le j\le T, j\ne t} b_{T^h,|j-t|} \hat{\Psi}'_{j,\tilde{h}_1} z_j x_j' \right)^{-1} \sum_{1\le j\le T, j\ne t} b_{T^h,|j-t|} \hat{\Psi}'_{j,\tilde{h}_1} z_j y_j$$

the CV choice of H is obtained as $\tilde{H}_{\text{CV}} = T^{\tilde{h}_2}$, where

$$\tilde{h}_2 = \arg\min_{h\le \tilde{h}_1} \left\{ \sum_{t=1}^{T} |y_t - \tilde{\beta}'_{(-t),h} x_t|^2 \right\}$$

In the case of the LS estimator $\hat{\beta}_t$, the CV choice of H is obtained as $\hat{H}_{\text{CV}} = T^{\hat{h}}$, where

$$\hat{h} = \arg\min_h \left\{ \sum_{t=1}^{T} |y_t - \hat{\beta}'_{(-t),h} x_t|^2 \right\}$$

$\hat{\beta}'_{(-t),h}$ being the leave-one-out version of the LS estimator of β_t given by

$$\hat{\beta}_{(-t),h} = \left(\sum_{1\le j\le T, j\ne t} b_{T^h,|j-t|} x_j x_j' \right)^{-1} \sum_{1\le j\le T, j\ne t} b_{T^h,|j-t|} x_j y_j$$

Note that the estimator $\hat{\Psi}_{j,\tilde{h}_1}$ used to construct $\tilde{\beta}_{(-t),h}$ is based on the bandwidth chosen by CV. It is also worth remarking that, although we focus on the popular leave-one-out CV method, it may be advantageous to construct CV criteria by leaving out more than one observation, or blocks of consecutive observations, especially when the data and/or errors are strongly correlated (see, e.g., Burman et al. 1994; Hall et al. 1995).

3.2 | Information Criterion

Hurvich et al. (1998) and Cai (2007), among others, suggested selecting the bandwidth for smoothing regression methods by using a nonparametric version of AIC. In our IV setting, an AIC-based procedure can be used sequentially to obtain data-driven choices of first L and then H.

TABLE 2 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators for model (6) with $T = 200$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 6$)	DWB ($\lambda = 8$)	DWB ($\lambda = 12$)	DWB ($\lambda = 16$)	DWB ($\lambda = 32$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.053	1.081	1.094	1.075	1.065	1.058	1.036	1.036
		$\tilde{\beta}_t$	1.025	0.821	1.085	1.062	1.051	1.043	1.021	1.021
		$\hat{\beta}_t$	1.047	1.068	1.092	1.064	1.058	1.047	1.039	1.028
	Absolute Median Deviation	$\tilde{\beta}_t$	0.219	1.718	0.219	0.229	0.230	0.230	0.230	0.229
		$\hat{\beta}_t$	0.129	0.128	0.128	0.129	0.129	0.129	0.129	0.129
	Coverage	$\tilde{\beta}_t$	87.304	69.345	85.760	81.879	82.322	82.546	82.806	83.424
		$\hat{\beta}_t$	74.016	73.851	72.837	73.524	73.657	74.010	74.400	74.918
	Optimal Coverage	$\tilde{\beta}_t$	88.283							
		$\hat{\beta}_t$	79.475							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.044	1.075	1.087	1.064	1.056	1.043	1.024	1.024
		$\tilde{\beta}_t$	1.021	0.837	1.080	1.053	1.044	1.033	1.014	1.014
		$\hat{\beta}_t$	1.046	1.073	1.091	1.065	1.056	1.044	1.037	1.026
	Absolute Median Deviation	$\tilde{\beta}_t$	0.196	1.641	0.196	0.204	0.204	0.204	0.204	0.204
		$\hat{\beta}_t$	0.128	0.126	0.127	0.128	0.128	0.128	0.128	0.128
	Coverage	$\tilde{\beta}_t$	86.266	65.980	84.904	81.631	81.777	82.167	82.444	82.911
		$\hat{\beta}_t$	72.979	72.538	71.802	72.397	72.648	73.033	73.400	73.805
	Optimal Coverage	$\tilde{\beta}_t$	87.079							
		$\hat{\beta}_t$	77.961							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.051	1.077	1.097	1.070	1.062	1.051	1.031	1.031
		$\tilde{\beta}_t$	1.022	0.808	1.086	1.055	1.046	1.035	1.015	1.015
		$\hat{\beta}_t$	1.010	1.032	1.060	1.026	1.017	1.005	0.999	0.990
	Absolute Median Deviation	$\tilde{\beta}_t$	0.186	1.804	0.186	0.194	0.194	0.195	0.195	0.195
		$\hat{\beta}_t$	0.271	0.270	0.270	0.271	0.271	0.271	0.271	0.271
	Coverage	$\tilde{\beta}_t$	84.441	61.000	83.252	80.091	80.258	80.635	80.848	81.265
		$\hat{\beta}_t$	39.228	38.700	37.767	38.752	39.000	39.422	39.572	39.883
	Optimal Coverage	$\tilde{\beta}_t$	86.012							
		$\hat{\beta}_t$	37.843							

To give a formal description of the procedure, let X and $\hat{X}_h$ be $p \times T$ matrices with t-th column x_t and $\hat{\Psi}'_{t,h} z_t$, respectively, and Q_h be the $pT \times pT$ matrix satisfying $\text{vec}(\hat{X}_h) = Q_h \text{vec}(X)$, where $\text{vec}(\cdot)$ is the vectorization function. The AIC choice of L is obtained as $\tilde{L}_{\text{AIC}} = T^{\overline{h}_1}$, where

$$\overline{h}_1 = \arg\min_h \left\{ \log\left(\sum_{t=1}^{T} \left\| x_t - \hat{\Psi}'_{t,h} z_t \right\|^2 \right) + \frac{2[\text{tr}(Q_h) + 1]}{pT - \text{tr}(Q_h) - 2} \right\}$$

and $\text{tr}(\cdot)$ is the trace function. Next, let R_h be the $T \times T$ matrix satisfying $(\tilde{\beta}'_{1,h} x_1, \ldots, \tilde{\beta}'_{T,h} x_T)' = R_h (y_1, \ldots, y_T)'$, where

$$\tilde{\beta}_{t,h} = \left(\sum_{j=1}^{T} b_{T^h,|j-t|} \hat{\Psi}'_{j,\overline{h}_1} z_j x'_j \right)^{-1} \sum_{j=1}^{T} b_{T^h,|j-t|} \hat{\Psi}'_{j,\overline{h}_1} z_j y_j$$

Then, the AIC choice of H is $\tilde{H}_{\text{AIC}} = T^{\overline{h}_2}$, where

$$\overline{h}_2 = \arg\min_{h \le \overline{h}_1} \left\{ \log\left(\sum_{t=1}^{T} \left| y_t - \tilde{\beta}'_{t,h} x_t \right|^2 \right) + \frac{2[\text{tr}(R_h) + 1]}{T - \text{tr}(R_h) - 2} \right\}$$

Note that, as in the CV selection procedure, the estimator $\hat{\Psi}_{j,\overline{h}_1}$ used to construct $\tilde{\beta}'_{t,h}$ is based on a data-driven bandwidth ($\tilde{L}_{\text{AIC}}$) obtained by the same method. The trace of the smoother matrices Q_h and R_h associated with any given bandwidth h (as well as that of the smoother matrix S_h below) is typically viewed as the effective number of parameters involved in the smoothing procedure.

For the LS estimator $\hat{\beta}_t$, the AIC choice of H is obtained in an analogous manner as $\hat{H}_{\text{AIC}} = T^{\overline{h}}$, with

$$\overline{h} = \arg\min_h \left\{ \log\left(\sum_{t=1}^{T} \left| y_t - \hat{\beta}'_{t,h} x_t \right|^2 \right) + \frac{2[\text{tr}(S_h) + 1]}{T - \text{tr}(S_h) - 2} \right\}$$

where S_h is the $T \times T$ matrix satisfying $(\hat{\beta}'_{1,h} x_1, \ldots, \hat{\beta}'_{T,h} x_T)' = S_h (y_1, \ldots, y_T)'$.

3.3 | Bootstrap

The bootstrap approach to bandwidth selection amounts to choosing a bandwidth which minimizes an appropriate bootstrap

TABLE 3 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators for model (6) with $T = 500$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 15$)	DWB ($\lambda = 22$)	DWB ($\lambda = 32$)	DWB ($\lambda = 45$)	DWB ($\lambda = 62$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.003	1.015	1.054	1.022	1.017	1.004	0.948	0.993
		$\tilde{\beta}_t$	0.994	0.785	1.016	0.989	0.984	0.976	0.971	0.976
		$\hat{\beta}_t$	1.004	1.014	1.056	1.022	1.013	1.001	0.997	0.992
	Absolute Median Deviation	$\tilde{\beta}_t$	0.095	0.705	0.100	0.095	0.096	0.096	0.096	0.096
		$\hat{\beta}_t$	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.051
	Coverage	$\tilde{\beta}_t$	94.735	84.199	93.479	93.390	93.453	93.700	93.849	94.022
		$\hat{\beta}_t$	88.573	88.425	87.354	88.176	88.436	88.700	88.853	88.956
	Optimal Coverage	$\tilde{\beta}_t$	94.880							
		$\hat{\beta}_t$	89.916							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.000	1.015	1.056	1.02	1.01	1.001	0.995	0.990
		$\tilde{\beta}_t$	0.999	0.805	1.016	0.998	0.985	0.985	0.982	0.977
		$\hat{\beta}_t$	1.002	1.011	1.054	1.018	1.006	0.999	0.993	0.989
	Absolute Median Deviation	$\tilde{\beta}_t$	0.174	1.379	0.187	0.172	0.173	0.174	0.173	0.173
		$\hat{\beta}_t$	0.103	0.102	0.103	0.103	0.103	0.103	0.103	0.103
	Coverage	$\tilde{\beta}_t$	89.032	63.708	86.797	86.469	86.788	87.178	87.488	87.565
		$\hat{\beta}_t$	75.608	75.367	73.178	74.899	75.501	75.907	76.262	76.456
	Optimal Coverage	$\tilde{\beta}_t$	89.262							
		$\hat{\beta}_t$	78.042							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.012	1.022	1.068	1.031	1.021	1.012	1.007	1.003
		$\tilde{\beta}_t$	0.996	0.784	1.015	0.993	0.986	0.982	0.982	0.985
		$\hat{\beta}_t$	0.988	0.998	1.043	1.003	0.991	0.982	0.979	0.976
	Absolute Median Deviation	$\tilde{\beta}_t$	0.165	1.513	0.177	0.164	0.165	0.165	0.165	0.166
		$\hat{\beta}_t$	0.297	0.297	0.297	0.297	0.297	0.298	0.298	0.298
	Coverage	$\tilde{\beta}_t$	87.570	58.534	85.780	85.015	85.322	85.592	85.791	85.920
		$\hat{\beta}_t$	24.920	24.506	22.783	24.492	24.825	25.114	25.203	25.282
	Optimal Coverage	$\tilde{\beta}_t$	88.171							
		$\hat{\beta}_t$	21.208							

estimator of the average squared error of the fitted model (e.g., Faraway 1990; Hall 1990; Hall et al. 1995; González Manteiga et al. 2004). In our IV setting, such an approach can be employed to obtain data-driven choices of first L and then H. To allow for the possibility that the errors in the model (1) and (2) may be heteroskedastic or serially correlated, we rely on the WB and DWB resampling schemes, originally proposed by Wu (1986) and Shao (2010), respectively. The idea behind such schemes is to construct of bootstrap errors by perturbing residuals by auxiliary random variables that are independent of the data; these random variables may be chosen to be mutually independent (as in WB) or correlated (as in DWB).

In the case of IV estimation, the selection procedure for L involves the following steps:

i. Using $\tilde{L}_{\text{CV}} = T^{\tilde{h}_1}$ as pilot bandwidth, generate pseudo-data x_t^* according to

$$x_t^* = \hat{\Psi}'_{t,\tilde{h}_1} z_t + \hat{v}_t \eta_{1,t}, \quad t = 1, 2, \ldots, T$$

where $\hat{v}_t = x_t - \hat{\Psi}'_{t,\tilde{h}_1} z_t$ and $\{\eta_{1,t}\}$ are random variables, independent of $\{(y_t, x'_t, z'_t)\}$, having zero mean and unit variance. For any $h \in (0,1)$, let $\hat{\Psi}^*_{t,h}$ be the bootstrap analogue of $\hat{\Psi}_{t,h}$, defined in the same way as the latter but using $(x_t^{*'}, z'_t)$ in place of (x'_t, z'_t).

ii. Repeating the previous step B times (with B sufficiently large), generate copies $\hat{\Psi}^*_{t,h,1}, \ldots, \hat{\Psi}^*_{t,h,B}$ of $\hat{\Psi}^*_{t,h}$ and obtain the bootstrap choice of L as $\tilde{L}_{\text{B}} = T^{h_1^*}$, where

$$h_1^* = \arg\min_h \left\{ \sum_{b=1}^{B} \sum_{t=1}^{T} \left\| \hat{\Psi}^{*\prime}_{t,h,b} z_t - \hat{\Psi}'_{t,\tilde{h}_1} z_t \right\|^2 \right\}$$

Next, given the choice $\tilde{L}_{\text{B}}$, the selection procedure for H is as follows:

i. Using $\hat{\Psi}_{t,h_1^*}$ (the LS estimator of Ψ_t with bandwidth $\tilde{L}_{\text{B}}$) and the pilot bandwidth $\tilde{H}_{\text{CV}} = T^{\tilde{h}_2}$ to construct the estimator

TABLE 4 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators for model (6) with deterministic coefficients and $T = 200$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 6$)	DWB ($\lambda = 8$)	DWB ($\lambda = 12$)	DWB ($\lambda = 16$)	DWB ($\lambda = 32$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	0.974	0.997	0.985	0.957	0.949	0.938	0.933	0.924
		$\tilde{\beta}_t$	0.969	0.741	0.975	0.950	0.942	0.933	0.926	0.917
		$\hat{\beta}_t$	1.151	1.177	1.174	1.158	1.149	1.135	1.127	1.113
	Absolute Median Deviation	$\tilde{\beta}_t$	0.093	1.872	0.092	0.093	0.093	0.094	0.095	0.095
		$\hat{\beta}_t$	0.071	0.070	0.070	0.070	0.071	0.071	0.071	0.071
	Coverage	$\tilde{\beta}_t$	92.417	43.490	92.780	92.660	92.512	92.605	92.693	92.715
		$\hat{\beta}_t$	85.457	85.692	85.715	85.790	86.055	86.382	86.162	86.613
	Optimal Coverage	$\tilde{\beta}_t$	93.047							
		$\hat{\beta}_t$	89.677							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	0.969	1.000	0.974	0.945	0.938	0.929	0.925	0.915
		$\tilde{\beta}_t$	0.963	0.731	0.966	0.942	0.933	0.926	0.918	0.907
		$\hat{\beta}_t$	1.160	1.192	1.186	1.164	1.156	1.141	1.135	1.120
	Absolute Median Deviation	$\tilde{\beta}_t$	0.083	1.996	0.082	0.082	0.083	0.084	0.084	0.085
		$\hat{\beta}_t$	0.068	0.068	0.068	0.069	0.068	0.069	0.069	0.069
	Coverage	$\tilde{\beta}_t$	91.612	41.932	91.945	91.850	91.822	91.923	91.827	91.838
		$\hat{\beta}_t$	83.863	83.448	83.617	83.800	84.627	85.117	85.118	85.537
	Optimal Coverage	$\tilde{\beta}_t$	92.465							
		$\hat{\beta}_t$	88.303							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	0.975	1.012	0.979	0.949	0.940	0.930	0.929	0.918
		$\tilde{\beta}_t$	0.965	0.629	0.977	0.944	0.936	0.923	0.921	0.909
		$\hat{\beta}_t$	1.102	1.130	1.139	1.106	1.097	1.086	1.081	1.072
	Absolute Median Deviation	$\tilde{\beta}_t$	0.074	2.339	0.074	0.074	0.074	0.075	0.075	0.076
		$\hat{\beta}_t$	0.114	0.114	0.114	0.114	0.114	0.114	0.114	0.114
	Coverage	$\tilde{\beta}_t$	91.350	54.165	91.625	91.553	91.505	91.455	91.440	91.362
		$\hat{\beta}_t$	55.648	54.910	54.113	55.728	55.848	56.528	56.537	56.838
	Optimal Coverage	$\tilde{\beta}_t$	92.065							
		$\hat{\beta}_t$	47.106							

$\tilde{\beta}_{t,\tilde{h}_2}$ of β_t, generate pseudo-data $(y_t^*, x_t^{*\prime})$ according to

$$y_t^* = \tilde{\beta}'_{t,\tilde{h}_2} x_t^* + \tilde{u}_t \eta_{2,t}, \quad t = 1, 2, \ldots, T,$$
$$x_t^* = \hat{\Psi}'_{t,h_1^*} z_t + \hat{v}_t \eta_{2,t}$$

where $\tilde{u}_t = y_t - \tilde{\beta}'_{t,\tilde{h}_2} x_t$, $\hat{v}_t = x_t - \hat{\Psi}'_{t,h_1^*} z_t$, and $\{\eta_{2,t}\}$ are random variables, independent of $\{(y_t, x_t', z_t', \eta_{1,t})\}$, having zero mean and unit variance. For any $h \in (0,1)$, let $\tilde{\beta}^*_{t,h}$ be the bootstrap analogue of $\tilde{\beta}_{t,h}$ given by

$$\tilde{\beta}^*_{t,h} = \left(\sum_{j=1}^{T} b_{T^h,|j-t|} \hat{\Psi}^{*\prime}_{j,h_1^*} z_j x_j^{*\prime} \right)^{-1} \sum_{j=1}^{T} b_{T^h,|j-t|} \hat{\Psi}^{*\prime}_{j,h_1^*} z_j y_j^*$$

ii. Repeating the previous step B times, generate copies $\tilde{\beta}^*_{t,h,1}, \ldots, \tilde{\beta}^*_{t,h,B}$ of $\tilde{\beta}^*_{t,h}$ and obtain the bootstrap choice of H as $\tilde{H}_B = T^{h_2^*}$, where

$$h_2^* = \arg\min_{h \le h_1^*} \left\{ \sum_{b=1}^{B} \sum_{t=1}^{T} \left| \tilde{\beta}^{*\prime}_{t,h,b} x_t - \tilde{\beta}'_{t,\tilde{h}_2} x_t \right|^2 \right\}$$

Notice that, following Davidson and MacKinnon (2010) and Chen (2015), $\tilde{u}_t$ and $\hat{v}_t$ are multiplied by the same auxiliary variable $\eta_{2,t}$ to preserve, as much as possible, the correlation between u_t and v_t when generating bootstrap data $(x_t^{*\prime}, y_t^*)$.

In the case of the LS estimator of β_t, the selection procedure for H involves the following steps:

TABLE 5 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators under (9), (10), and (11) with $T = 200$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 6$)	DWB ($\lambda = 8$)	DWB ($\lambda = 12$)	DWB ($\lambda = 16$)	DWB ($\lambda = 32$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.016	1.049	1.062	1.040	1.033	1.022	1.014	0.999
		$\tilde{\beta}_t$	0.996	0.963	1.002	0.981	0.985	0.973	0.972	0.964
		$\hat{\beta}_t$	1.333	1.377	1.372	1.346	1.343	1.337	1.330	1.318
	Absolute Median Deviation	$\tilde{\beta}_t$	0.591	2.770	0.688	0.605	0.583	0.610	0.617	0.624
		$\hat{\beta}_t$	0.256	0.243	0.252	0.253	0.255	0.256	0.256	0.258
	Coverage	$\tilde{\beta}_t$	93.662	79.033	93.015	93.092	93.165	93.255	93.368	93.302
		$\hat{\beta}_t$	78.803	78.833	78.215	78.590	78.778	78.775	78.917	79.140
	Optimal Coverage	$\tilde{\beta}_t$	94.438							
		$\hat{\beta}_t$	87.100							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.041	1.064	1.085	1.061	1.051	1.041	1.034	1.018
		$\tilde{\beta}_t$	1.022	0.959	1.029	1.021	1.006	1.011	0.990	0.991
		$\hat{\beta}_t$	1.347	1.377	1.389	1.371	1.363	1.356	1.349	1.340
	Absolute Median Deviation	$\tilde{\beta}_t$	0.428	2.349	0.467	0.414	0.420	0.423	0.428	0.432
		$\hat{\beta}_t$	0.220	0.216	0.215	0.216	0.217	0.219	0.219	0.219
	Coverage	$\tilde{\beta}_t$	92.093	72.975	91.468	91.555	91.617	91.687	91.631	91.695
		$\hat{\beta}_t$	75.103	74.497	74.067	74.523	74.667	74.535	74.848	75.308
	Optimal Coverage	$\tilde{\beta}_t$	92.985							
		$\hat{\beta}_t$	82.180							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.060	1.090	1.106	1.083	1.068	1.064	1.053	1.039
		$\tilde{\beta}_t$	1.040	0.913	1.055	1.047	1.015	1.017	1.009	1.001
		$\hat{\beta}_t$	1.227	1.266	1.275	1.248	1.238	1.231	1.225	1.209
	Absolute Median Deviation	$\tilde{\beta}_t$	0.346	2.822	0.379	0.339	0.338	0.340	0.346	0.344
		$\hat{\beta}_t$	0.569	0.569	0.570	0.570	0.570	0.570	0.569	0.570
	Coverage	$\tilde{\beta}_t$	90.320	61.815	9.467	89.333	89.263	89.420	89.418	89.722
		$\hat{\beta}_t$	29.240	28.093	27.747	28.457	28.828	28.943	29.335	29.837
	Optimal Coverage	$\tilde{\beta}_t$	91.252							
		$\hat{\beta}_t$	25.215							

i. Using $\hat{H}_{\text{CV}} = T^{\hat{h}}$ as pilot bandwidth, generate pseudo-data y_t^* according to

$$y_t^* = \hat{\beta}_{t,\hat{h}}' x_t + \hat{u}_t \eta_{3,t}, \quad t = 1, 2, \ldots, T$$

where $\hat{u}_t = y_t - \hat{\beta}_{t,\hat{h}}' x_t$ and $\{\eta_{3,t}\}$ are random variables, independent of $\{(y_t, x_t')\}$, having zero mean and unit variance. For any $h \in (0, 1)$, let $\hat{\beta}_{t,h}^*$ be the bootstrap version of $\hat{\beta}_{t,h}$, obtained by replacing (y_t, x_t') in the definition of $\hat{\beta}_{t,h}$ with (y_t^*, x_t').

ii. Repeating the above step B times, generate copies $\hat{\beta}_{t,h,1}^{*'}, \ldots, \hat{\beta}_{t,h,B}^{*'}$ of $\hat{\beta}_{t,h}^*$ and obtain the bootstrap choice of H as $\hat{H}_{\text{B}} = T^{h^*}$, where

$$h^* = \arg\min_h \left\{ \sum_{b=1}^{B} \sum_{t=1}^{T} \left| \hat{\beta}_{t,h,b}^{*\prime} x_t - \hat{\beta}_{t,\hat{h}}' x_t \right|^2 \right\}$$

The bandwidth selection procedures based on WB and DWB differ only in the choice of the correlation structure of the collections of auxiliary random variables $\{\eta_{i,t}\}$ $(i = 1, 2, 3)$. In the WB case, we take $\{\eta_{i,t}\}$ to be independent $\mathcal{N}(0, 1)$ random variables; thus, the bootstrap errors reflect possible heterogeneity in the variance of the original errors. For the DWB, we follow Shao (2010) and Djogbenou et al. (2015) in taking $\{\eta_{i,t}\}$ to be jointly Gaussian with mean zero and covariances $E(\eta_{i,t}\eta_{i,k}) = \Lambda(|t-k|/\lambda)$, where $\Lambda(w) = (1 - |w|)\mathbb{I}(|w| < 1)$ is the triangular kernel function and $\lambda > 0$ is a bandwidth controlling the extent of dependence (with $\lambda \to \infty$ and $\lambda/T \to 0$ as $T \to \infty$); hence, the bootstrap errors reflect possible serial correlation in the original errors.

It is worth noting that, if heteroskedasticity and serial correlation are not a concern, then the bootstrap errors that are required to generate x_t^* and y_t^* may be obtained by resampling from the empirical distribution of the relevant residuals. For instance, when selecting the bandwidth H for $\tilde{\beta}_t$, this amounts to choosing bootstrap errors by sampling independently and uniformly, with replacement, from the residuals $\{(\tilde{u}_t, \hat{v}_t'), t = 1, 2, \ldots, T\}$ after centering them around their arithmetic mean. The use of such

TABLE 6 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators under (9), (10), and (12) with $d = 1.2$ and $T = 200$.

		Estimator	**CV**	**AIC**	**WB**	**DWB ($\lambda = 6$)**	**DWB ($\lambda = 8$)**	**DWB ($\lambda = 12$)**	**DWB ($\lambda = 16$)**	**DWB ($\lambda = 32$)**
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.078	1.086	1.121	1.098	1.097	1.084	1.075	1.066
		$\tilde{\beta}_t$	1.052	1.006	1.084	1.059	1.064	1.044	1.039	1.030
		$\hat{\beta}_t$	1.423	1.455	1.453	1.433	1.430	1.424	1.417	1.409
	Absolute Median Deviation	$\tilde{\beta}_t$	0.852	1.844	0.922	0.826	0.825	0.839	0.836	0.841
		$\hat{\beta}_t$	0.378	0.373	0.371	0.374	0.373	0.378	0.378	0.381
	Coverage	$\tilde{\beta}_t$	93.965	83.048	93.925	94.185	94.095	94.122	94.148	94.190
		$\hat{\beta}_t$	82.813	82.687	82.483	82.848	82.625	82.558	82.783	82.926
	Optimal Coverage	$\tilde{\beta}_t$	94.928							
		$\hat{\beta}_t$	88.465							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.069	1.088	1.121	1.092	1.087	1.073	1.066	1.053
		$\tilde{\beta}_t$	1.040	1.019	1.054	1.038	1.031	1.023	1.021	1.010
		$\hat{\beta}_t$	1.446	1.501	1.487	1.465	1.460	1.448	1.446	1.436
	Absolute Median Deviation	$\tilde{\beta}_t$	0.656	1.785	0.719	0.628	0.632	0.638	0.644	0.642
		$\hat{\beta}_t$	0.323	0.309	0.314	0.321	0.320	0.322	0.322	0.325
	Coverage	$\tilde{\beta}_t$	93.302	75.930	92.887	93.102	93.043	93.088	93.215	93.221
		$\hat{\beta}_t$	9.443	79.412	78.842	78.875	78.945	79.047	79.292	79.397
	Optimal Coverage	$\tilde{\beta}_t$	93.883							
		$\hat{\beta}_t$	84.363							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.062	1.082	1.106	1.079	1.075	1.064	1.054	1.042
		$\tilde{\beta}_t$	1.042	0.941	1.056	1.027	1.027	1.011	1.019	1.008
		$\hat{\beta}_t$	1.305	1.364	1.362	1.332	1.323	1.317	1.307	1.288
	Absolute Median Deviation	$\tilde{\beta}_t$	0.580	3.194	0.618	0.557	0.559	0.570	0.570	0.576
		$\hat{\beta}_t$	0.946	0.941	0.945	0.946	0.945	0.948	0.945	0.947
	Coverage	$\tilde{\beta}_t$	91.383	63.635	91.362	91.192	91.217	91.143	91.128	91.167
		$\hat{\beta}_t$	26.937	25.493	24.735	25.855	26.250	26.423	26.573	27.348
	Optimal Coverage	$\tilde{\beta}_t$	92.295							
		$\hat{\beta}_t$	19.630							

a resampling scheme is, however, inadvisable when the original errors may be serially correlated and/or heteroskedastic (and will not be considered in the sequel).

4 | Monte Carlo Simulations

In this section, the finite-sample performance of various data-driven bandwidth selectors for the kernel-based LS and IV estimators $\hat{\beta}_t$ and $\tilde{\beta}_t$ is evaluated by means of simulations. The Monte Carlo's experiments are based on data-generating processes (DGPs) that are variants of those previously used by GKM. We consider exactly identified and overidentified models, with errors that may be independent and identically distributed (i.i.d.), conditionally heteroskedastic, or serially correlated. As it is generally accepted that the choice of kernel (K) is of secondary importance compared to the choice of smoothing parameters (L, H), we use the Gaussian kernel $K(w) = \exp(-w^2/2)$ in all subsequent computations.

4.1 | Homoskedasticity and Independence

The first set of experiments is based on an exact identified version of the model (1) and (2) with $n = p = 1$, that is,

$$y_t = \beta_t x_t + u_t, \quad x_t = \psi_t z_t + v_t, \quad t = 1, 2, \ldots, T \tag{6}$$

where $T \in \{100, 200, 500\}$. As in GKM, $\{z_t\}$ are i.i.d. $\mathcal{N}(0, 1)$ random variables, while $\{u_t\}$ and $\{v_t\}$ are such that

$$u_t = se_{1,t} + (1-s)e_{2,t}, \quad v_t = se_{1,t} + (1-s)e_{3,t} \tag{7}$$

where $\{e_t' = (e_{1,t}, e_{2,t}, e_{3,t})\}$ are i.i.d. Gaussian random vectors, independent of $\{z_t\}$, having zero mean and identity covariance matrix. Hence, the strength of endogeneity, as measured by $\mathrm{Corr}(u_t, v_t) = s^2/[s^2 + (1-s)^2]$, is controlled by s,

TABLE 7 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators under (9), (10), and (12) with $d = 1.4$ and $T = 200$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 6$)	DWB ($\lambda = 8$)	DWB ($\lambda = 12$)	DWB ($\lambda = 16$)	DWB ($\lambda = 32$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.078	1.086	1.121	1.098	1.097	1.084	1.075	1.066
		$\tilde{\beta}_t$	1.052	1.002	1.083	1.056	1.063	1.043	1.040	1.030
		$\hat{\beta}_t$	1.423	1.465	1.459	1.439	1.432	1.428	1.424	1.415
	Absolute Median Deviation	$\tilde{\beta}_t$	0.823	1.819	0.892	0.799	0.798	0.811	0.809	0.812
		$\hat{\beta}_t$	0.366	0.360	0.358	0.360	0.362	0.364	0.366	0.368
	Coverage	$\tilde{\beta}_t$	93.828	82.617	93.722	94.010	93.933	93.950	93.937	94.032
		$\hat{\beta}_t$	82.425	82.003	81.887	82.352	82.252	82.258	82.452	82.597
	Optimal Coverage	$\tilde{\beta}_t$	94.803							
		$\hat{\beta}_t$	88.358							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.069	1.088	1.121	1.092	1.087	1.073	1.066	1.053
		$\tilde{\beta}_t$	1.040	1.014	1.055	1.037	1.031	1.021	1.021	1.010
		$\hat{\beta}_t$	1.432	1.502	1.476	1.456	1.447	1.434	1.432	1.422
	Absolute Median Deviation	$\tilde{\beta}_t$	0.633	1.786	0.690	0.606	0.609	0.615	0.621	0.619
		$\hat{\beta}_t$	0.314	0.299	0.307	0.312	0.310	0.313	0.313	0.315
	Coverage	$\tilde{\beta}_t$	93.113	75.428	92.730	92.903	92.845	92.905	93.047	93.018
		$\hat{\beta}_t$	79.300	79.045	78.577	78.677	78.992	79.213	79.207	79.335
	Optimal Coverage	$\tilde{\beta}_t$	93.782							
		$\hat{\beta}_t$	84.190							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.062	1.082	1.106	1.079	1.075	1.064	1.054	1.042
		$\tilde{\beta}_t$	1.043	0.941	1.056	1.029	1.028	1.013	1.019	1.009
		$\hat{\beta}_t$	1.290	1.364	1.346	1.316	1.307	1.300	1.292	1.274
	Absolute Median Deviation	$\tilde{\beta}_t$	0.558	3.113	0.591	0.537	0.538	0.549	0.550	0.554
		$\hat{\beta}_t$	0.911	0.906	0.910	0.910	0.911	0.912	0.912	0.911
	Coverage	$\tilde{\beta}_t$	91.286	62.773	91.210	91.063	91.100	91.032	91.046	91.078
		$\hat{\beta}_t$	27.308	25.592	25.133	26.062	26.473	26.793	26.923	27.572
	Optimal Coverage	$\tilde{\beta}_t$	92.265							
		$\hat{\beta}_t$	20.165							

with $s \in \{0, 0.2, 0.5\}$. The coefficients $\{\beta_t\}$ and $\{\psi_t\}$ vary stochastically as rescaled random walks,

$$\beta_t = T^{-1/2} \sum_{j=0}^{t-1} \xi_{1,t-j}, \quad \psi_t = T^{-1/2} \sum_{j=0}^{t-1} \xi_{2,t-j} \tag{8}$$

where $\{\xi_{1,t}\}$ and $\{\xi_{2,t}\}$ are collections of i.i.d. $\mathcal{N}(0,1)$ random variables, independent of each other and of $\{(e_t', z_t)\}$.

As noted in the description of the bandwidth selection procedures in Section 3, the same data-driven procedure is used for the selection of both L and H in the case of the IV estimator (the only exception being the use of a CV pilot bandwidth in the construction of the IV estimator of β_t required to generate bootstrap data). For bootstrap-based selection procedures, the number of bootstrap replications is $B = 399$. In the case of DWB, we consider $\lambda \in \{2, 4, 6, 8, 10\}$ when $T = 100$, $\lambda \in \{6, 8, 12, 16, 32\}$ when $T = 200$, and $\lambda \in \{15, 22, 32, 45, 62\}$ when $T = 500$.[3] In all cases, the relevant objective functions are minimized over an equispaced grid of 30 points corresponding to bandwidths ranging in the interval $[T^{0.2}, T^{0.9}]$.

The properties of bandwidth selectors for IV and LS estimators of β_t are evaluated using several performance indicators. Specifically, for a kernel-based estimator of β_t, say $\check{\beta}_{t,\check{h}}$ (IV or LS), with bandwidth $T^{\check{h}}$ selected by one of the methods discussed in Section 3, we consider the following performance measures (based on R Monte Carlo replications):

i. Average ratio of selected bandwidth to optimal bandwidth, computed as

$$R^{-1} \sum_{r=1}^{R} T^{\check{h}_r - h_r^{\text{opt}}}$$

where $\check{h}_r$ is the value of $\check{h}$ in the r-th Monte Carlo replication and h_r^{opt} is the corresponding optimal value; the latter is obtained as the minimizer of $T^{-1} \sum_{t=1}^{T} |\check{\beta}_{t,h} - \beta_t|^m$, with $m = 1$ and $m = 2$ for the IV and LS estimators, respectively.[4]

TABLE 8 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators under (6) and (13) with $T = 200$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 6$)	DWB ($\lambda = 8$)	DWB ($\lambda = 12$)	DWB ($\lambda = 16$)	DWB ($\lambda = 32$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.016	1.049	1.062	1.040	1.033	1.022	1.014	0.999
		$\tilde{\beta}_t$	1.015	0.813	1.046	1.034	1.025	1.022	1.016	1.003
		$\hat{\beta}_t$	1.004	1.027	1.052	1.021	1.013	1.002	0.995	0.986
	Absolute Median Deviation	$\tilde{\beta}_t$	0.222	2.904	0.2447	0.240	0.234	0.239	0.239	0.240
		$\hat{\beta}_t$	0.127	0.125	0.127	0.127	0.127	0.127	0.126	0.126
	Coverage	$\tilde{\beta}_t$	86.848	68.980	83.685	82.448	82.860	83.163	83.530	84.093
		$\hat{\beta}_t$	73.460	73.187	71.837	72.708	73.103	73.538	73.823	74.253
	Optimal Coverage	$\tilde{\beta}_t$	88.068							
		$\hat{\beta}_t$	78.518							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.041	1.064	1.085	1.061	1.051	1.041	1.034	1.018
		$\tilde{\beta}_t$	1.026	0.838	1.069	1.053	1.042	1.046	1.030	1.021
		$\hat{\beta}_t$	1.042	1.079	1.092	1.063	1.051	1.040	1.032	1.023
	Absolute Median Deviation	$\tilde{\beta}_t$	0.198	2.406	0.216	0.203	0.204	0.203	0.203	0.203
		$\hat{\beta}_t$	0.125	0.123	0.125	0.126	0.125	0.125	0.125	0.124
	Coverage	$\tilde{\beta}_t$	85.658	66.040	82.773	81.215	81.515	81.960	81.993	82.510
		$\hat{\beta}_t$	72.427	72.043	70.900	71.645	72.058	72.522	72.903	73.295
	Optimal Coverage	$\tilde{\beta}_t$	86.897							
		$\hat{\beta}_t$	76.903							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.060	1.090	1.106	1.083	1.068	1.064	1.053	1.039
		$\tilde{\beta}_t$	1.050	0.825	1.087	1.075	1.047	1.065	1.047	1.030
		$\hat{\beta}_t$	1.035	1.072	1.079	1.050	1.040	1.028	1.015	1.010
	Absolute Median Deviation	$\tilde{\beta}_t$	0.188	2.377	0.202	0.193	0.191	0.192	0.194	0.193
		$\hat{\beta}_t$	0.269	0.268	0.269	0.269	0.269	0.269	0.269	0.269
	Coverage	$\tilde{\beta}_t$	83.548	60.390	81.018	79.420	79.772	79.770	80.075	80.763
		$\hat{\beta}_t$	37.960	37.076	36.642	37.466	37.675	38.210	38.488	38.657
	Optimal Coverage	$\tilde{\beta}_t$	85.825							
		$\hat{\beta}_t$	37.280							

ii. Average median absolute estimation error of $\check{\beta}_{t,\check{h}}$, computed as

$$R^{-1}\sum_{r=1}^{R}\text{med}\{|\check{\beta}_{t,\check{h}_r}-\beta_t| : t=1,2,\ldots,T\}$$

iii. Average coverage rate of 95% two-sided confidence intervals for β_t, computed as

$$100(TR)^{-1}\sum_{t=1}^{T}\sum_{r=1}^{R}\mathbb{I}(|\check{\beta}_{t,\check{h}_r}-\beta_t|/\text{se}(\check{\beta}_{t,\check{h}_r})\leq 1.96)$$

where $\text{se}(\check{\beta}_{t,\check{h}_r})$ is an estimate of the asymptotic standard deviation of $\check{\beta}_{t,\check{h}_r}$ (obtained as in GKM).

All reported simulation results are obtained from $R = 1{,}000$ Monte Carlo replications.

Tables 1–3 contain results for $T = 100$, $T = 200$ and $T = 500$, respectively. The data-driven methods are similarly behaved when selecting the bandwidth for the LS estimator of ψ_t, with DWB (with large bandwidth λ) being slightly superior for the two smaller sample sizes in terms of the ratio of the selected bandwidth to the optimal value that minimizes the mean squared estimation error of $\hat{\psi}_t$. For the IV estimator of β_t, CV outperforms all other methods in terms of the ratio of the selected bandwidth to the optimal value that minimizes the mean absolute estimation error of $\tilde{\beta}_t$ for $T = 100$, regardless of whether x_t is exogenous ($s = 0$) or endogenous ($s \neq 0$); it is less effective than DWB (with $\lambda \geq 16$), but only by a slight margin, when $T = 200$ and $s \neq 0$, and, together with WB, delivers the best results when $T = 500$. Furthermore, CV bandwidths produce pointwise confidence intervals for β_t the average coverage of which is close to the coverage associated with the optimal bandwidth, outperforming other automatically selected bandwidths in this respect for all values of s and T. It must be pointed out, however, that even the optimal bandwidth (for the given simulated data) yields confidence intervals, the average coverage of which (labeled “optimal coverage” in the tables) falls considerably short of the nominal 95% rate, except when x_t is exogenous and $T = 500$.[5] The AIC-based selector is the

TABLE 9 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators for model (14) with $\sigma_t = \tau_t = 1$ and $T = 100$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 2$)	DWB ($\lambda = 4$)	DWB ($\lambda = 6$)	DWB ($\lambda = 8$)	DWB ($\lambda = 10$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.045	0.662	1.103	1.101	1.099	1.095	1.094	1.094
		$\tilde{\beta}_t$	1.053	0.835	1.117	1.109	1.094	1.082	1.068	1.068
		$\hat{\beta}_t$	1.084	1.125	1.117	1.108	1.084	1.071	1.065	1.055
	Absolute Median Deviation	$\tilde{\beta}_t$	0.163	0.225	0.163	0.208	0.208	0.207	0.208	0.208
		$\hat{\beta}_t$	0.137	0.134	0.137	0.137	0.137	0.137	0.137	0.137
	Coverage	$\tilde{\beta}_t$	88.434	85.768	87.547	74.730	74.908	75.004	74.816	74.853
		$\hat{\beta}_t$	70.072	70.106	69.659	69.621	70.209	70.357	70.487	70.845
	Optimal Coverage	$\tilde{\beta}_t$	90.414							
		$\hat{\beta}_t$	76.358							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.042	0.725	1.111	1.108	1.107	1.101	1.103	1.103
		$\tilde{\beta}_t$	1.062	0.919	1.123	1.113	1.093	1.083	1.069	1.069
		$\hat{\beta}_t$	1.079	1.123	1.111	1.103	1.080	1.064	1.059	1.054
	Absolute Median Deviation	$\tilde{\beta}_t$	0.148	0.202	0.148	0.195	0.196	0.195	0.195	0.195
		$\hat{\beta}_t$	0.137	0.134	0.137	0.137	0.137	0.137	0.137	0.137
	Coverage	$\tilde{\beta}_t$	87.219	83.904	86.562	72.684	72.642	72.877	72.758	72.613
		$\hat{\beta}_t$	69.658	69.457	69.279	69.340	69.703	69.907	69.981	70.151
	Optimal Coverage	$\tilde{\beta}_t$	89.251							
		$\hat{\beta}_t$	75.195							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.038	0.766	1.106	1.107	1.105	1.104	1.104	1.104
		$\tilde{\beta}_t$	1.079	0.963	1.141	1.130	1.110	1.100	1.083	1.083
		$\hat{\beta}_t$	1.038	1.076	1.070	1.059	1.035	1.022	1.017	1.009
	Absolute Median Deviation	$\tilde{\beta}_t$	0.138	0.191	0.138	0.187	0.187	0.187	0.187	0.188
		$\hat{\beta}_t$	0.164	0.163	0.163	0.164	0.164	0.164	0.164	0.164
	Coverage	$\tilde{\beta}_t$	85.886	82.506	85.255	70.819	70.835	70.887	70.881	70.794
		$\hat{\beta}_t$	57.479	57.356	56.935	57.057	57.401	57.670	57.855	57.982
	Optimal Coverage	$\tilde{\beta}_t$	88.073							
		$\hat{\beta}_t$	60.373							

least competitive overall, yielding bandwidths that are lower than the optimal bandwidth and associated confidence intervals for β_t which undercover considerably. There is little to choose among competing methods when considering the average median absolute estimation error of $\tilde{\beta}_t$, DWB having a slight advantage and being more successful, the stronger the correlation between x_t and u_t is. It is perhaps noteworthy that DWB, based on relatively large values of λ perform well (and generally dominate WB) even though the errors (u_t, v_t) are i.i.d. in the simulations.

Turning to the LS estimator of β_t, the results in Tables 1–3 show that, for all bandwidth selectors, the average median absolute estimation error of $\hat{\beta}_t$ is lower than that of the IV estimator when x_t is exogenous or endogeneity is weak ($s = 0.2$), while the reverse is true under moderate endogeneity ($s = 0.5$). The bandwidths selected by the various methods tend to be somewhat higher than the optimal values, except for DWB when T and λ are large. As in the IV case, CV and DWB (with λ that is not too small) generally provide the most accurate choices relative to the optimal bandwidth for $\hat{\beta}_t$ when $T \leq 200$, the former having a slight advantage when endogeneity is moderately strong, while AIC is the least successful. Selectors based on CV, AIC and DWB (with $\lambda \leq 32$) perform similarly for $T = 500$, CV having a marginal disadvantage when $s = 0.5$. Undercoverage of confidence intervals for β_t is once again a problem, regardless of the bandwidth selector used in the construction of $\hat{\beta}_t$ and of the sample size. Although inaccuracy of LS confidence intervals is not surprising when x_t and u_t are correlated (coverage rates are uniformly lower than 50% when $s = 0.5$), and the use of the LS estimator is clearly not recommended in these circumstances, the problem is also present when x_t is exogenous and $\hat{\beta}_t$ is consistent.

As a robustness check, we also consider a design in which the coefficients in (6) vary deterministically as $\beta_t = \beta(t/T)$ and $\psi_t = \psi(t/T)$, where

$$\beta(w) = 2w + \exp(-16[w - 1/2]^2) - 1,$$
$$\psi(w) = (7/2)\{\exp(-[4w - 1]^2) + \exp(-[4w - 3]^2)\} - 3/2$$

TABLE 10 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators for model (14) with $\sigma_t = \tau_t = 1$ and $T = 200$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 6$)	DWB ($\lambda = 8$)	DWB ($\lambda = 12$)	DWB ($\lambda = 16$)	DWB ($\lambda = 32$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.0331	0.600	1.097	1.093	1.091	1.087	1.086	1.079
		$\tilde{\beta}_t$	1.034	0.701	1.113	1.087	1.082	1.069	1.062	1.047
		$\hat{\beta}_t$	1.037	1.059	1.088	1.055	1.047	1.035	1.029	1.016
	Absolute Median Deviation	$\tilde{\beta}_t$	0.148	0.209	0.148	0.173	0.173	0.173	0.173	0.173
		$\hat{\beta}_t$	0.117	0.115	0.117	0.117	0.117	0.117	0.117	0.117
	Coverage	$\tilde{\beta}_t$	90.418	89.023	88.734	78.551	78.587	78.747	78.796	78.931
		$\hat{\beta}_t$	73.619	73.709	72.256	73.016	73.289	73.691	73.914	74.505
	Optimal Coverage	$\tilde{\beta}_t$	91.534							
		$\hat{\beta}_t$	78.519							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.034	0.669	1.104	1.100	1.100	1.096	1.092	1.092
		$\tilde{\beta}_t$	1.034	0.776	1.106	1.080	1.072	1.060	1.041	1.041
		$\hat{\beta}_t$	1.035	1.057	1.086	1.052	1.043	1.029	1.024	1.015
	Absolute Median Deviation	$\tilde{\beta}_t$	0.133	0.184	0.133	0.161	0.161	0.162	0.161	0.162
		$\hat{\beta}_t$	0.112	0.111	0.111	0.112	0.112	0.112	0.112	0.112
	Coverage	$\tilde{\beta}_t$	89.499	87.661	88.060	76.460	76.375	76.435	76.556	76.489
		$\hat{\beta}_t$	73.135	73.128	71.755	72.531	72.781	73.268	73.469	73.928
	Optimal Coverage	$\tilde{\beta}_t$	90.600							
		$\hat{\beta}_t$	77.465							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.022	0.722	1.098	1.096	1.093	1.094	1.090	1.090
		$\tilde{\beta}_t$	1.045	0.825	1.113	1.086	1.078	1.068	1.050	1.050
		$\hat{\beta}_t$	0.959	0.980	1.007	0.971	0.960	0.951	0.944	0.938
	Absolute Median Deviation	$\tilde{\beta}_t$	0.123	0.170	0.123	0.154	0.153	0.207	0.153	0.154
		$\hat{\beta}_t$	0.175	0.174	0.175	0.175	0.175	0.175	0.175	0.175
	Coverage	$\tilde{\beta}_t$	88.212	85.980	87.147	74.342	74.521	75.004	74.475	74.417
		$\hat{\beta}_t$	52.795	52.380	51.396	52.400	52.817	53.145	53.390	53.537
	Optimal Coverage	$\tilde{\beta}_t$	89.879							
		$\hat{\beta}_t$	52.805							

These functional forms have been previously used by Cai (2007) and Chen (2015). The results recorded in Table 4 generally lead to the same conclusions regarding the relative merits of the data-driven bandwidth selectors as those obtained from designs with stochastically varying coefficients, with a slightly improved performance observed in the case of WB. The most notable difference is the coverage of confidence intervals for β_t based on the IV estimator, which is now much closer to the target nominal rate for all selectors except AIC (LS-based confidence intervals undercover even when $s = 0$). Although it is unwise to draw conclusions from a single Monte Carlo design, it seems that deterministic variation in the regression coefficients is less challenging than stochastic variation in terms of the accuracy of interval estimators of the coefficients.

In sum, although an ordinary CV is sometimes reported to perform poorly in nonparametric regression settings (see, e.g., Härdle et al. 1988), it is found to provide effective choices of the bandwidth for kernel IV and LS estimators of time-varying coefficients—at least when performance measures other than coverage of confidence intervals are considered—both in the presence and absence of endogeneity. DWB is competitive with the CV selector and consistently better than AIC (and often WB).

4.2 | Heteroskedasticity and Serial Correlation

To assess the effect of (conditional) heteroskedasticity on the performance of bandwidth selectors, we consider artificial data from a generalized version of the DGP (6–8) in which the equation for y_t is replaced by

$$y_t = \beta_t x_t + \sigma_t \tau_t u_t,\ t = 1, 2, \ldots, T \tag{9}$$

with

$$\sigma_t = [1 + (0.2u_{t-1}^2 + 0.7)\sigma_{t-1}^2]^{1/2} \tag{10}$$

and $\{\tau_t\}$ satisfying one of the following:

$$\tau_t = 1 \quad \text{for all } t \tag{11}$$

TABLE 11 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators for model (14) with $\sigma_t = \tau_t = 1$ and $T = 500$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 15$)	DWB ($\lambda = 22$)	DWB ($\lambda = 32$)	DWB ($\lambda = 45$)	DWB ($\lambda = 62$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.011	0.519	1.066	1.061	1.054	1.052	1.049	1.049
		$\tilde{\beta}_t$	1.008	0.568	1.102	1.110	1.100	1.104	1.100	1.102
		$\hat{\beta}_t$	0.904	0.912	0.955	0.918	0.908	0.902	0.897	0.893
	Absolute Median Deviation	$\tilde{\beta}_t$	0.126	0.210	0.132	0.138	0.138	0.138	0.138	0.138
		$\hat{\beta}_t$	0.094	0.093	0.093	0.093	0.094	0.093	0.093	0.094
	Coverage	$\tilde{\beta}_t$	92.542	88.379	92.415	88.379	84.563	84.817	84.794	85.023
		$\hat{\beta}_t$	77.919	77.877	75.399	77.263	77.695	78.156	78.527	78.593
	Optimal Coverage	$\tilde{\beta}_t$	93.203							
		$\hat{\beta}_t$	79.710							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.011	0.587	1.069	1.063	1.061	1.059	1.057	1.054
		$\tilde{\beta}_t$	1.009	0.644	1.081	1.099	1.101	1.104	1.104	1.107
		$\hat{\beta}_t$	0.922	0.930	0.976	0.935	0.925	0.916	0.915	0.914
	Absolute Median Deviation	$\tilde{\beta}_t$	0.114	0.180	0.120	0.127	0.128	0.128	0.127	0.128
		$\hat{\beta}_t$	0.091	0.090	0.091	0.091	0.091	0.091	0.091	0.091
	Coverage	$\tilde{\beta}_t$	91.945	91.348	88.62	82.805	82.811	82.847	82.881	83.004
		$\hat{\beta}_t$	76.939	76.676	74.349	76.207	76.809	77.279	77.392	77.473
	Optimal Coverage	$\tilde{\beta}_t$	92.703							
		$\hat{\beta}_t$	78.407							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.007	0.643	1.071	1.065	1.064	1.061	1.061	1.060
		$\tilde{\beta}_t$	1.009	0.696	1.065	1.091	1.097	1.099	1.097	1.101
		$\hat{\beta}_t$	0.894	0.909	0.952	0.903	0.895	0.889	0.889	0.885
	Absolute Median Deviation	$\tilde{\beta}_t$	0.106	0.160	0.110	0.121	0.121	0.121	0.121	0.122
		$\hat{\beta}_t$	0.199	0.198	0.198	0.199	0.199	0.199	0.198	0.198
	Coverage	$\tilde{\beta}_t$	91.017	89.652	88.135	80.710	80.860	80.751	80.823	80.673
		$\hat{\beta}_t$	41.111	40.456	38.500	40.769	41.175	41.355	41.361	41.586
	Optimal Coverage	$\tilde{\beta}_t$	91.667							
		$\hat{\beta}_t$	36.151							

$$\tau_t = 1 + T^{-d+(1/2)}|\tau_t^*|, \quad \tau_t^* = \sum_{j=0}^{t-1} \frac{\Gamma(j+d)}{j!\Gamma(d)} \zeta_{t-j}, \quad d \in \{1.2, 1.4\} \tag{12}$$

where $\{\zeta_t\}$ are i.i.d. $\mathcal{N}(0,1)$ variables independent of $\{(e_t', z_t, \xi_{1,t}, \xi_{2,t})\}$ and $\Gamma(\cdot)$ is the gamma function. Thus, the time-varying conditional standard deviation of the noise in Equation (9) has a stationary GARCH component (σ_t) and, under (12), an additional persistent, nonstationary component (τ_t) which is a positive function of a rescaled fractionally integrated process of order $d > 1$. These volatility specifications have been previously used by Chronopoulos et al. (2022).

Table 5 summarizes simulation results under (9–11) when $T = 200$. The performance of data-driven bandwidth selection methods for the IV estimator $\tilde{\beta}_t$ is generally similar to that documented earlier under homoskedastic designs. CV, WB and DWB provide the best choices in terms of closeness of the automatically selected bandwidths to the optimal value and magnitude of the average median absolute estimation error of $\tilde{\beta}_t$. The AIC selector also performs well, but only when considering deviations of the selected bandwidth from the optimal value and only for $s = 0$. Interestingly, there is improvement in the coverage of IV confidence intervals for β_t compared to the case of i.i.d. errors, with only the AIC selector delivering coverage rates lower than 90%. Undercoverage is much more substantial in the case of the LS estimator $\hat{\beta}_t$, even when x_t is exogenous, and becomes unacceptably large when $s = 0.5$. Under exogeneity ($s = 0$) or weak endogeneity ($s = 0.2$), $\hat{\beta}_t$ has lower average median estimation error than $\tilde{\beta}_t$ regardless of the bandwidth selector used; however, unlike the IV case, no selector delivers values that are close to the optimal bandwidth.

Qualitatively similar results are obtained under the volatility specification (12), as can be seen in Tables 6 and 7. Once again, for all values of s and all bandwidth selectors, the presence of conditional heteroskedasticity is beneficial for the coverage rates of confidence intervals for β_t based on the IV estimator, although AIC is not particularly successful in this respect. The improved coverage may be due to the fact that the covariance estimator

TABLE 12 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators under (14), (10), and (11) with $T = 200$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 6$)	DWB ($\lambda = 8$)	DWB ($\lambda = 12$)	DWB ($\lambda = 16$)	DWB ($\lambda = 32$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.029	0.611	1.099	1.094	1.093	1.091	1.089	1.084
		$\tilde{\beta}_t$	1.017	0.626	1.092	1.085	1.085	1.080	1.079	1.068
		$\hat{\beta}_t$	1.203	1.223	1.231	1.213	1.211	1.201	1.194	1.184
	Absolute Median Deviation	$\tilde{\beta}_t$	0.300	0.440	0.297	0.294	0.294	0.299	0.296	0.297
		$\hat{\beta}_t$	0.229	0.223	0.225	0.227	0.227	0.227	0.229	0.229
	Coverage	$\tilde{\beta}_t$	95.713	94.178	94.693	92.940	93.006	92.900	92.825	92.841
		$\hat{\beta}_t$	75.392	75.460	74.810	75.393	75.385	75.778	75.708	76.242
	Optimal Coverage	$\tilde{\beta}_t$	96.408							
		$\hat{\beta}_t$	83.803							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.027	0.677	1.102	1.096	1.095	1.095	1.094	1.092
		$\tilde{\beta}_t$	1.015	0.698	1.078	1.079	1.074	1.075	1.071	1.067
		$\hat{\beta}_t$	1.139	1.192	1.184	1.158	1.153	1.141	1.134	1.120
	Absolute Median Deviation	$\tilde{\beta}_t$	0.230	0.318	0.228	0.231	0.232	0.233	0.232	0.234
		$\hat{\beta}_t$	0.195	0.189	0.191	0.193	0.194	0.194	0.195	0.195
	Coverage	$\tilde{\beta}_t$	94.298	92.738	93.005	89.352	89.280	89.120	89.205	89.048
		$\hat{\beta}_t$	73.478	72.378	72.600	72.960	72.972	73.493	73.603	74.373
	Optimal Coverage	$\tilde{\beta}_t$	94.743							
		$\hat{\beta}_t$	81.298							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.010	0.723	1.090	1.088	1.086	1.084	1.085	1.082
		$\tilde{\beta}_t$	1.014	0.762	1.068	1.071	1.066	1.066	1.065	1.065
		$\hat{\beta}_t$	1.061	1.102	1.108	1.060	1.051	1.044	1.044	1.041
	Absolute Median Deviation	$\tilde{\beta}_t$	0.200	0.272	0.197	0.203	0.204	0.205	0.205	0.207
		$\hat{\beta}_t$	0.349	0.346	0.348	0.349	0.349	0.349	0.349	0.349
	Coverage	$\tilde{\beta}_t$	92.280	90.570	91.450	86.445	86.290	86.342	86.248	86.115
		$\hat{\beta}_t$	44.768	43.755	42.963	44.782	45.177	45.567	45.515	45.612
	Optimal Coverage	$\tilde{\beta}_t$	93.538							
		$\hat{\beta}_t$	42.245							

used in the construction of confidence intervals explicitly allows for heterogeneity in the error variances. As in designs with $\tau_t = 1$, however, heteroskedasticity is found to be challenging for bandwidth selectors for the LS estimator of β_t, even when x_t is exogenous—all data-driven methods deliver bandwidths that are too high relative to the value that minimizes the mean squared estimation error of $\hat{\beta}_t$. This is in contrast to the corresponding results obtained under homoskedasticity, or results obtained for the IV estimator under heteroskedasticity. Also note that findings are not sensitive with respect to the order of integration of the nonstationary component $\{\tau_t^*\}$, the conclusions reached for $d = 1.2$ and $d = 1.4$ being similar.[6]

Next, to investigate the effect of serial correlation in the errors, we consider a variant of the DGP (6–8) in which

$$u_t = se_{1,t} + (1-s)(1-\varphi^2)^{1/2}\epsilon_t,$$
$$\epsilon_t = \varphi\epsilon_{t-1} + e_{2,t},$$
$$v_t = se_{1,t} + (1-s)e_{3,t} \tag{13}$$

Thus, for $|\varphi| \in (0,1)$, the autocorrelation structure of $\{u_t\}$ is that of a causal ARMA(1, 1) process. The results obtained under this DGP, with $\varphi = 0.8$ and $T = 200$, are collected in Table 8.[7]

Although leave-one-out CV is often found to be problematic in nonparametric regression settings with serially correlated errors (e.g., Hart 1991; Opsomer et al. 2001), deviations from the independence assumption do not appear to have an adverse effect on CV in our varying-coefficients setup. For all values of s, CV performs as well as DWB (which explicitly allows for serial correlation) when selecting the bandwidth for the IV or LS estimator of β_t, yielding bandwidths that are close to the optimal values. Its performance is also almost identical to that of DWB in terms of the average median absolute estimation error of the estimators, while AIC is the least successful selector overall. Once again, the coverage of pointwise confidence intervals leaves much to be desired, even when the optimal bandwidth is used. It should be noted, however, that coverage results should be viewed with caution in this case since confidence intervals are based on an asymptotic normal approximation to the distribution of $\tilde{\beta}_t$ that

TABLE 13 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators under (14), (10), and (12) with $d = 1.2$ and $T = 200$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 6$)	DWB ($\lambda = 8$)	DWB ($\lambda = 12$)	DWB ($\lambda = 16$)	DWB ($\lambda = 32$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.021	0.599	1.086	1.077	1.072	1.066	1.068	1.066
		$\tilde{\beta}_t$	1.013	0.602	1.077	1.068	1.062	1.056	1.057	1.055
		$\hat{\beta}_t$	1.296	1.337	1.318	1.293	1.285	1.282	1.276	1.276
	Absolute Median Deviation	$\tilde{\beta}_t$	0.457	0.743	0.452	0.446	0.444	0.447	0.444	0.447
		$\hat{\beta}_t$	0.329	0.320	0.326	0.332	0.332	0.332	0.332	0.331
	Coverage	$\tilde{\beta}_t$	96.983	95.952	96.617	96.170	96.105	96.105	96.068	96.102
		$\hat{\beta}_t$	81.027	80.992	80.415	80.975	81.243	81.343	81.517	81.467
	Optimal Coverage	$\tilde{\beta}_t$	97.060							
		$\hat{\beta}_t$	87.383							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.022	0.661	1.093	1.080	1.081	1.078	1.080	1.073
		$\tilde{\beta}_t$	1.019	0.671	1.086	1.072	1.075	1.066	1.068	1.060
		$\hat{\beta}_t$	1.302	1.333	1.336	1.298	1.295	1.289	1.283	1.277
	Absolute Median Deviation	$\tilde{\beta}_t$	0.338	0.517	0.331	0.328	0.329	0.329	0.328	0.329
		$\hat{\beta}_t$	0.273	0.268	0.270	0.272	0.272	0.273	0.275	0.272
	Coverage	$\tilde{\beta}_t$	95.748	94.875	95.310	93.960	93.792	93.957	93.787	93.752
		$\hat{\beta}_t$	76.213	75.320	74.750	76.063	76.301	76.568	76.410	77.026
	Optimal Coverage	$\tilde{\beta}_t$	96.123							
		$\hat{\beta}_t$	82.751							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.029	0.716	1.106	1.095	1.098	1.096	1.095	1.094
		$\tilde{\beta}_t$	1.018	0.714	1.085	1.073	1.071	1.066	1.067	1.062
		$\hat{\beta}_t$	1.208	1.252	1.246	1.204	1.195	1.192	1.184	1.185
	Absolute Median Deviation	$\tilde{\beta}_t$	0.302	0.430	0.286	0.287	0.287	0.288	0.289	0.288
		$\hat{\beta}_t$	0.594	0.592	0.593	0.595	0.595	0.596	0.595	0.595
	Coverage	$\tilde{\beta}_t$	94.427	93.665	93.833	92.088	91.887	91.860	91.758	91.848
		$\hat{\beta}_t$	37.513	36.082	36.065	37.483	37.826	38.060	38.175	38.195
	Optimal Coverage	$\tilde{\beta}_t$	95.350							
		$\hat{\beta}_t$	34.012							

is obtained under the assumption that $\{z_t u_t\}$ is an uncorrelated process (cf. Theorem 3ii in GKM).

4.3 | Overidentification

In the final set of experiments, we consider an overidentified version of the model (1) and (2) with $n = p + 1 = 2$, that is,

$$y_t = \beta_t x_t + \sigma_t \tau_t u_t, \quad x_t = \psi_{1,t} z_{1,t} + \psi_{2,t} z_{2,t} + v_t, \quad t = 1, 2, \ldots, T \tag{14}$$

with $\sigma_t, \tau_t > 0$ (to be specified later). As before, $\{u_t\}$ and $\{v_t\}$ satisfy (7), $\{\beta_t\}$, $\{\psi_{1,t}\}$ and $\{\psi_{2,t}\}$ are generated as independent Gaussian random walks (rescaled by $T^{-1/2}$), and $\{z_{1,t}\}$ and $\{z_{2,t}\}$ are collections of i.i.d. $\mathcal{N}(0,1)$ random variables independent of each other and of $\{(u_t, \tau_t, v_t, \beta_t, \psi_{1,t}, \psi_{2,t})\}$.

Simulation results under the DGP in Equation (14) with $\sigma_t = \tau_t = 1$ for all t and $T \in \{100, 200, 500\}$ are collected in Tables 9–11. As in exactly identified models with i.i.d. errors, CV and DWB outperform AIC and WB in the majority of cases in terms of the ratio of the selected bandwidth for IV and LS estimators of β_t to the optimal value (although WB becomes more competitive when $T = 500$). CV and WB result in estimates of β_t that generally have the lowest median absolute estimation error, for all values of s, but the former selector has a clear advantage when considering coverage of confidence intervals relative to the coverage associated with the optimal bandwidth value. However, paralleling earlier findings for an exactly identified model, coverage rates are uniformly below the 95% target value, especially so in the case of confidence intervals based on the LS estimator $\hat{\beta}_t$.

Allowing for conditional heteroskedasticity via the GARCH specification for σ_t given in Equation (10) and the specifications for τ_t given in Equations (11) and (12), the results reported in Tables 12–14 are obtained (with $T = 200$). CV has superior performance in terms of the closeness of the selected bandwidths to the optimal value for IV and LS estimators of β_t (followed by DWB); in the LS case, however, the presence of heteroskedasticity

TABLE 14 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators under (14), (10), and (12) with $d = 1.4$ and $T = 200$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 6$)	DWB ($\lambda = 8$)	DWB ($\lambda = 12$)	DWB ($\lambda = 16$)	DWB ($\lambda = 32$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.021	0.599	1.086	1.077	1.072	1.066	1.068	1.066
		$\tilde{\beta}_t$	1.011	0.602	1.077	1.068	1.063	1.055	1.056	1.054
		$\hat{\beta}_t$	1.290	1.322	1.309	1.287	1.283	1.279	1.273	1.268
	Absolute Median Deviation	$\tilde{\beta}_t$	0.447	0.720	0.439	0.432	0.432	0.434	0.432	0.434
		$\hat{\beta}_t$	0.322	0.315	0.318	0.323	0.323	0.324	0.324	0.323
	Coverage	$\tilde{\beta}_t$	96.960	95.877	96.577	96.066	96.030	95.966	95.958	95.997
		$\hat{\beta}_t$	80.565	80.717	80.233	80.553	80.747	80.827	81.032	81.164
	Optimal Coverage	$\tilde{\beta}_t$	97.018							
		$\hat{\beta}_t$	86.970							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.022	0.661	1.093	1.089	1.087	1.083	1.085	1.077
		$\tilde{\beta}_t$	1.016	0.672	1.084	1.079	1.075	1.074	1.074	1.063
		$\hat{\beta}_t$	1.292	1.329	1.329	1.308	1.306	1.293	1.286	1.274
	Absolute Median Deviation	$\tilde{\beta}_t$	0.331	0.501	0.324	0.317	0.318	0.317	0.318	0.321
		$\hat{\beta}_t$	0.266	0.262	0.264	0.265	0.266	0.265	0.266	0.266
	Coverage	$\tilde{\beta}_t$	95.697	94.808	95.178	93.883	93.850	93.902	93.678	93.651
		$\hat{\beta}_t$	76.250	74.862	74.798	75.435	75.483	76.185	76.377	76.945
	Optimal Coverage	$\tilde{\beta}_t$	96.073							
		$\hat{\beta}_t$	82.842							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.030	0.716	1.106	1.104	1.102	1.101	1.099	1.098
		$\tilde{\beta}_t$	1.019	0.716	1.085	1.084	1.082	1.079	1.074	1.068
		$\hat{\beta}_t$	1.211	1.247	1.250	1.230	1.227	1.212	1.207	1.193
	Absolute Median Deviation	$\tilde{\beta}_t$	0.295	0.417	0.279	0.278	0.278	0.277	0.279	0.280
		$\hat{\beta}_t$	0.576	0.573	0.575	0.577	0.577	0.577	0.577	0.577
	Coverage	$\tilde{\beta}_t$	94.358	93.548	93.737	91.832	91.743	91.740	91.677	91.665
		$\hat{\beta}_t$	37.466	36.565	36.247	36.793	36.777	37.378	37.708	38.000
	Optimal Coverage	$\tilde{\beta}_t$	95.278							
		$\hat{\beta}_t$	34.623							

tends to have a deleterious effect on the accuracy of all selectors. The data-driven bandwidths yield similar average median absolute estimation errors, with AIC generally being the least successful selector in this respect (as well as in terms of bandwidth accuracy relative to the optimal value). As in exactly identified models, the coverage of LS confidence intervals leaves much to be desired even when $s = 0$, while the coverage of IV confidence intervals is improved when compared to the case of homoskedastic models; CV and WB are the most successful in this respect, the former delivering coverage that is closest to that associated with the optimal bandwidth.

Lastly, Table 15 summarizes results when (u_t, v_t) are generated according to (13) and $\sigma_t = \tau_t = 1$ for all t. In the presence of serial correlation, CV remains the most effective method for selecting bandwidths for the IV estimator that are close to the optimal values, while DWB has the edge in the case of LS estimation. Even though CV does better than other methods in terms of coverage of IV confidence intervals, the figures are still well below the target nominal value (which is also the case for the optimal bandwidth). As in exactly identified models, LS confidence intervals associated with any of the automated bandwidths undercover substantially even when x_t is exogenous.

5 | Empirical Examples

In this section, we illustrate the practical use of the automated selection procedures in the context of two empirical applications. In these, we consider well-known specifications for the Phillips curve and for a Taylor policy rule, with coefficients subject to time variation.

5.1 | Phillips Curve

We first revisit the time-varying version of the backward-looking Phillips curve analyzed by GKM relating U.S. price inflation to unemployment. The aim is to compare estimates of the parameters of the model obtained using different data-driven bandwidth

TABLE 15 | Ratio of data-driven bandwidth to optimal, absolute median deviation and pointwise coverage for first-stage LS, IV and LS estimators under (14) and (13) with $\sigma_t = \tau_t = 1$ and $T = 200$.

		Estimator	CV	AIC	WB	DWB ($\lambda = 6$)	DWB ($\lambda = 8$)	DWB ($\lambda = 12$)	DWB ($\lambda = 16$)	DWB ($\lambda = 32$)
$s = 0$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.029	0.611	1.099	1.094	1.093	1.091	1.089	1.084
		$\tilde{\beta}_t$	1.048	0.717	1.164	1.171	1.175	1.176	1.188	1.181
		$\hat{\beta}_t$	0.957	0.985	1.000	0.968	0.963	0.954	0.950	0.938
	Absolute Median Deviation	$\tilde{\beta}_t$	0.143	0.194	0.155	0.173	0.173	0.174	0.173	0.173
		$\hat{\beta}_t$	0.115	0.113	0.115	0.115	0.115	0.115	0.115	0.115
	Coverage	$\tilde{\beta}_t$	90.000	87.888	86.180	78.335	78.398	78.358	78.415	78.628
		$\hat{\beta}_t$	72.602	72.333	71.347	72.137	72.433	72.730	72.983	73.343
	Optimal Coverage	$\tilde{\beta}_t$	91.317							
		$\hat{\beta}_t$	77.743							
$s = 0.2$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.028	0.677	1.102	1.096	1.095	1.095	1.094	1.092
		$\tilde{\beta}_t$	1.044	0.803	1.150	1.168	1.174	1.182	1.187	1.193
		$\hat{\beta}_t$	0.968	0.989	1.014	0.982	0.968	0.962	0.957	0.948
	Absolute Median Deviation	$\tilde{\beta}_t$	0.129	0.175	0.139	0.160	0.160	0.161	0.161	0.161
		$\hat{\beta}_t$	0.111	0.110	0.111	0.111	0.111	0.111	0.111	0.111
	Coverage	$\tilde{\beta}_t$	89.097	86.390	85.850	76.137	76.165	76.098	75.942	76.087
		$\hat{\beta}_t$	71.900	72.210	70.630	71.362	71.810	72.058	72.295	72.483
	Optimal Coverage	$\tilde{\beta}_t$	90.430							
		$\hat{\beta}_t$	76.325							
$s = 0.5$	Bandwidth ratio to optimal	$\hat{\Psi}_t$	1.010	0.723	1.090	1.089	1.085	1.087	1.088	1.082
		$\tilde{\beta}_t$	1.036	0.838	1.109	1.146	1.153	1.164	1.174	1.175
		$\hat{\beta}_t$	0.933	0.959	0.979	0.942	0.934	0.923	0.919	0.912
	Absolute Median Deviation	$\tilde{\beta}_t$	0.121	0.165	0.128	0.153	0.152	0.153	0.153	0.154
		$\hat{\beta}_t$	0.174	0.173	0.174	0.174	0.174	0.174	0.174	0.174
	Coverage	$\tilde{\beta}_t$	88.050	84.727	85.648	74.303	74.295	74.256	74.140	74.213
		$\hat{\beta}_t$	52.877	52.305	51.513	52.577	52.753	53.212	53.330	53.580
	Optimal Coverage	$\tilde{\beta}_t$	89.560							
		$\hat{\beta}_t$	52.195							

TABLE 16 | Selected values of h; the associated bandwidth is $H = T^h$.

	h	
Method	**LS**	**IV**
CV	0.643	0.899
AIC	0.642	0.871
WB	0.690	0.823
DWB ($\lambda = 42$)	0.597	0.853
DWB ($\lambda = 59$)	0.643	0.783
DWB ($\lambda = 68$)	0.597	0.807
DWB ($\lambda = 76$)	0.690	0.877
DWB ($\lambda = 85$)	0.620	0.899

selectors. Specifically, we consider the model

$$\Delta\pi_t = c_t + \gamma_t \Delta\pi_{t-1} + \alpha_t \Delta U_t + \epsilon_t, \quad t = 1, 2, \ldots, T \tag{15}$$

where π_t is the rate of price inflation, U_t is the unemployment rate, $(c_t, \gamma_t, \alpha_t)$ are unknown coefficients, ϵ_t is a random error, and Δ is the first-difference operator. The data (obtained from the FRED database) consist of $T = 648$ monthly observations, from 1959:1 to 2013:12, on the CPI inflation rate and the unemployment rate. Following GKM, estimates are obtained using $(1, \Delta\pi_{t-1}, \Delta U_{t-1}, \Delta U_{t-2}, \Delta U_{t-3}, \Delta U_{t-4})$ as the vector of instruments and $K(w) = \exp(-w^2/2)$ as kernel function. As noted in GKM, a Lagrange multiplier test for fourth-order serial correlation reveals no significant signs of serial correlation in IV residuals of the model (GKM set $L = H = T^{0.7}$).

Table 16 reports the bandwidths $H = T^h$ for LS and IV estimators that are selected by means of the data-driven methods discussed in Section 3. In the case of bootstrap-based selectors, results are obtained using $B = 999$ bootstrap replications, with bandwidth $\lambda \in \{42, 59, 68, 76, 85\}$ for the DWB.[8] While the differences between the bandwidth values chosen by the various procedures do not appear to be substantial, there are some noticeable differences in the resulting coefficient estimates.

In the LS case, these differences can be seen in Figure 1, which shows LS estimates of γ_t and α_t, together with corresponding 95% pointwise confidence bands, based on bandwidths obtained by CV, WB, and DWB (with $\lambda = 68$). Estimates of γ_t based on the three automatically selected bandwidths are quite similar for most of the sample, the only exception being a period around 2,000. The same is true for estimates of α_t, with some differences among the three sets of estimates also observed during the 1980s. In fact, it is only during the latter period that the coefficient on ΔU_t appears to be statistically significant (at the 5% level), regardless of the bandwidth selector used. Needless to say, these results should be viewed with caution since LS estimates are inconsistent unless ΔU_t is exogenous in (15). As a matter of fact, this does not appear to be the case: The time-varying Hausman test of GKM rejects exogeneity.[9]

Turning to IV estimation of the parameters of the model (15), Figure 2 shows IV estimates of γ_t and α_t, and associated 95% pointwise confidence bands. From the mid-1970s onwards, there is little difference between the estimates of either parameter obtained using the CV, WB, and DWB bandwidth choices, small differences being evident only early in the sample period. Interestingly, the coefficient on unemployment is statistically significant (at the 5% level) for all points in the sample, suggesting that a traditional unemployment-inflation trade-off is supported by the data, once endogeneity of unemployment is accounted for via the use of IV.

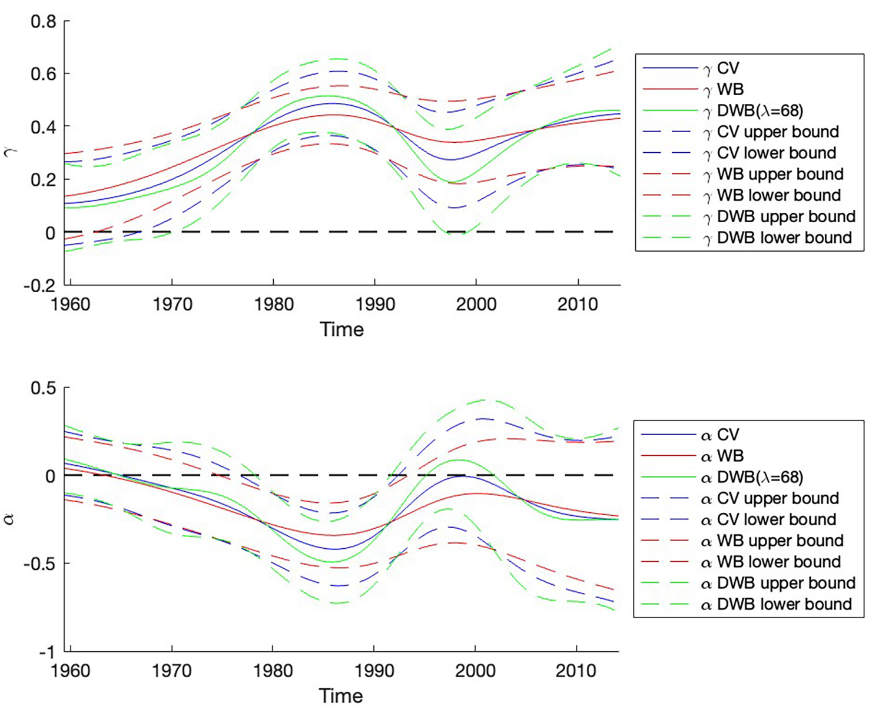

FIGURE 1 | LS estimates of γ_t and α_t based on bandwidths selected by CV, WB and DWB ($\lambda = 68$).

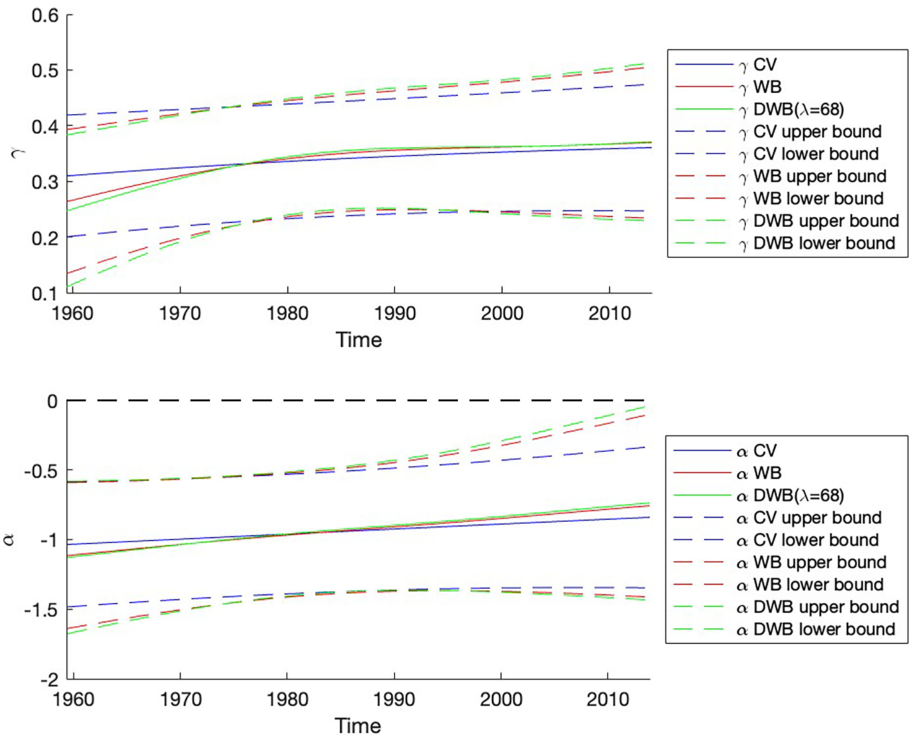

FIGURE 2 | IV estimates of γ_t and α_t based on bandwidths selected by CV, WB and DWB ($\lambda = 68$).

5.2 | Taylor Rule

As a second empirical application, we consider a Taylor rule relating U.S. interest rates to price inflation and the deviation of output from the economy's potential supply (output gap). Econometric formulations of such monetary-policy rules, which have been found to provide a good empirical descriptions of the policy behavior of many central banks, typically involve endogenous covariates and often exhibit structural instability; see, for example, Clarida et al. (2000) and Carvalho et al. (2021) (CNT hereafter).

Our model is a varying-coefficient version of the contemporaneous Taylor-rule specification analyzed in CNT, that is,

$$i_t = c_t + \delta_t \pi_t + \theta_t g_t + \rho_{1,t} i_{t-1} + \rho_{2,t} i_{t-2} + \epsilon_t, \quad t = 1, 2, \ldots, T \tag{16}$$

where i_t is the nominal interest rate (Federal Funds rate), π_t is the rate of price inflation (year-on-year growth rate of the core PCE price index), g_t is the output gap (deviation of real GDP from the potential level estimated by the Congressional Budget Office), $(c_t, \delta_t, \theta_t, \rho_{1,t}, \rho_{2,t})$ are unknown coefficients, and ϵ_t is a random error. The instruments chosen are the same as those used by Clarida et al. (2000) and CNT, namely the first four lags of i_t, π_t and g_t, as well as the same lags of the growth rate of the M2 money stock, the growth rate of the all-commodities producer price index, and the yield spread between 10-year Treasury notes and 3-month Treasury bills. The data, taken from CNT, comprise $T = 192$ real-time quarterly observations from 1960:1 to 2007:4.

Table 17 reports the bandwidths $H = T^h$ for LS and IV estimators of the coefficients in Equation (16) that are selected using the data-driven methods described in Section 3 and the Gaussian kernel function $K(w) = \exp(-w^2/2)$. For selectors based on the bootstrap approach, the results are obtained from $B = 999$ bootstrap replications, using bandwidth $\lambda \in \{2, 4, 6, 8, 10\}$ for the DWB. As in the previous empirical example, the bandwidth values chosen by the various procedures are fairly similar: The CV-selected bandwidth is approximately $T^{0.6}$ for both LS and IV estimators, while the corresponding average DWB-selected values are $T^{0.58}$ and $T^{0.69}$, respectively.

TABLE 17 | Selected values of h; the associated bandwidth is $H = T^h$.

	h	
Method	**LS**	**IV**
CV	0.620	0.597
AIC	0.623	0.610
WB	0.597	0.597
DWB ($\lambda = 2$)	0.527	0.690
DWB ($\lambda = 4$)	0.690	0.713
DWB ($\lambda = 6$)	0.620	0.713
DWB ($\lambda = 8$)	0.433	0.668
DWB ($\lambda = 10$)	0.620	0.667

LS estimates of δ_t and θ_t, together with corresponding 95% pointwise confidence bands, based on bandwidths obtained by CV, WB, and DWB (with $\lambda = 8$), are shown in Figure 3; the periods associated with different Federal Reserve chairs are also indicated.[10] Coefficient estimates based on the three automatically selected bandwidths are quite similar across the sample. In the case of δ_t, estimates vary more in the post-Miller period than

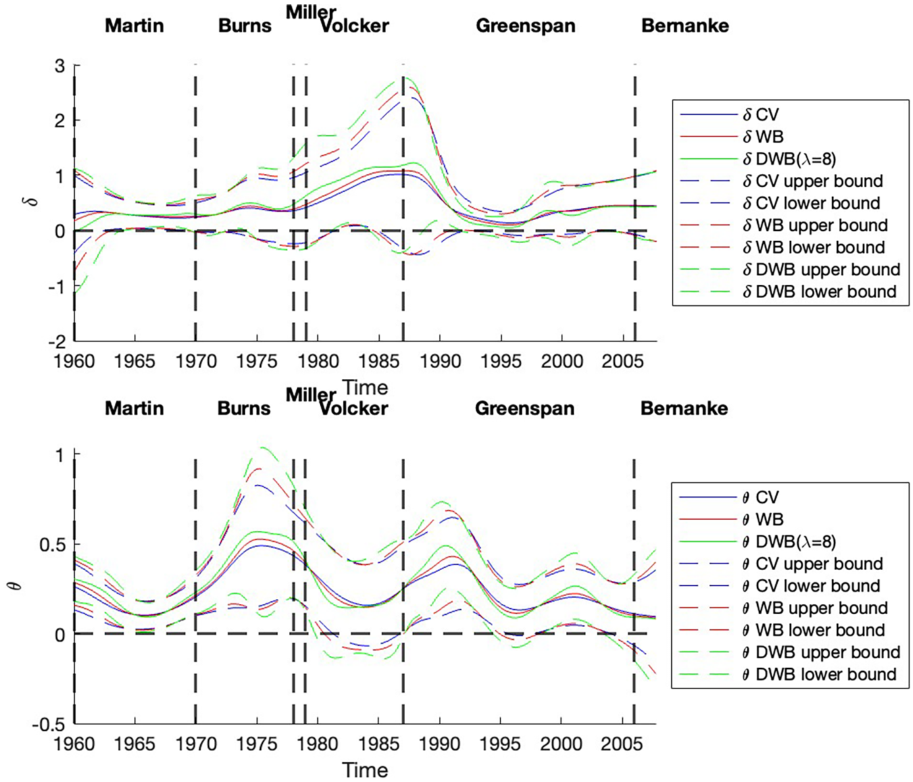

FIGURE 3 | LS estimates of δ_t and θ_t based on bandwidths selected by CV, WB and DWB ($\lambda = 8$).

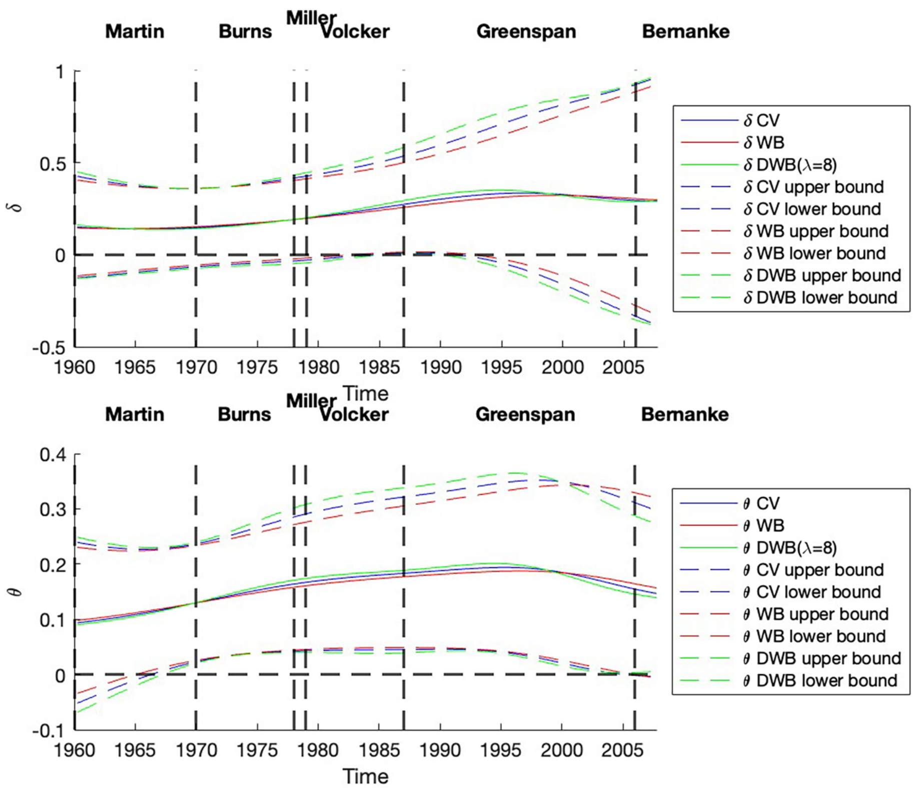

FIGURE 4 | IV estimates of δ_t and θ_t based on bandwidths selected by CV, WB and DWB ($\lambda = 8$).

in earlier periods, the higher values being associated with the late-Volcker and early-Greenspan periods. Estimates of θ_t exhibit more variation, by comparison, with higher values being associated with the Burns and early-Greenspan periods. Estimates of $\rho_{1,t}$ and $\rho_{2,t}$ (not shown) also vary considerably over the sample period, suggesting varying degrees of smoothing of interest-rate changes—CV-based estimates of the sum $\rho_{1,t} + \rho_{2,t}$ range from 0.25 to 0.88, with higher values associated with the Burns–Miller period and the mid-to-late Greenspan and Bernanke periods.[11] Point estimates of the multipliers $\delta_t/(1 - \rho_{1,t} - \rho_{2,t})$ vary between 0.38 and 3.74, and, except for the period 1992:4–1994:4, exceed unity from 1969:1 onwards, indicating a more complex policy response to inflation than that suggested by the split-sample results of CNT.

IV estimates of δ_t and θ_t, and associated 95% pointwise confidence bands, are shown in Figure 4. There is little difference between estimates of either parameter obtained using the CV, WB, and DWB bandwidth choices. Estimates are more stable than those obtained by LS, higher values of the estimated parameters being generally associated with the Greenspan era. Estimates of $\rho_{1,t}$ and $\rho_{2,t}$ (not shown) also exhibit less variation than their LS counterparts, CV-based IV estimates of $\rho_{1,t} + \rho_{2,t}$ ranging from 0.81 to 0.89 (those of $\delta_t/(1 - \rho_{1,t} - \rho_{2,t})$ range from 0.93 to 1.84 and exceed unity from 1963:3 onwards). Note, however, that the estimated coefficients on π_t (both IV and LS) are not significantly different from zero (at the 5% level) in many cases. A possible explanation for this finding may lie with the undercoverage that pointwise confidence intervals for time-varying coefficients tend to have (and their reliance on the assumption that the product of the instruments and the errors is serially uncorrelated). Notwithstanding these observations, the findings suggest gradual changes in the parameters of (16), at least judging by LS point estimates, which do not always coincide with a change in the chairmanship of the Federal Reserve. It is important to bear in mind, however, that the time-varying Hausman test of GKM and the bootstrap-based variant of Grivas (2023) reject exogeneity for the entire sample period, and so, LS results should be viewed cautiously.[12]

6 | Conclusion

We have considered data-driven methods for selecting the smoothing parameter for kernel IV and LS estimators of stochastically time-varying coefficients in linear models with explanatory variables that may be endogenous. Our simulation findings have revealed that CV and DWB are effective automated methods, selecting bandwidths which are close to the optimal values and yielding coefficient estimators with minimal average estimation errors. What is more, DWB and, perhaps surprisingly, ordinary CV work equally well in models with heteroskedastic or serially correlated errors as they do in models with i.i.d. errors. Our results provide valuable insights into the effectiveness of different data-driven methods for bandwidth selection, and can be used to address an obvious hurdle in the practical application of kernel estimators of time-varying coefficients in a rich class of models.

A finding which deserves further attention is that, regardless of the data-driven bandwidth selector used, pointwise confidence intervals for time-varying coefficients appear to have coverage rates that are generally lower than the nominal target value. This difficulty also arises when bandwidth values that are optimal (in the sense of minimizing the average absolute or quadratic estimation error) are used and become more challenging in the presence of homoskedasticity. It would be useful, therefore, to consider data-driven selectors that produce bandwidth choices that control effectively, the error in coverage rates of pointwise confidence intervals, or of simultaneous confidence regions, for time-varying coefficients. The possibility of constructing such confidence intervals/regions using appropriate bootstrap approximations to the sampling distributions of the kernel IV and LS estimators, instead of the asymptotic normal approximations, would also be worth exploring. These problems will be considered in detail elsewhere.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

[1] This is the estimator denoted $\tilde{\beta}_{1,t}$ in GKM. Under certain conditions (see Lemma 2 in GKM), it is asymptotically equivalent to the two-stage local linear estimator of Chen (2015).

[2] Based on results from simulation experiments, GKM recommend setting $H = L = T^{1/2}$.

[3] Recall that our choice of auxiliary random variables $\{\eta_{i,t}\}$ for DWB implies that they form a $\lceil \lambda - 1 \rceil$-dependent sequence, where $\lceil \cdot \rceil$ is the ceiling function.

[4] The absolute estimation error ($m = 1$) is considered in the IV case because the finite-sample distributions of IV-type estimators tend to be heavy-tailed due to lack of finite moments.

[5] These findings are consistent with those of GKM (for bandwidths H and L taking the values $T^{0.4}$ or $T^{0.5}$), who also report undercoverage that becomes more substantial as the strength of endogeneity increases.

[6] We also considered designs in which the generating mechanism of σ_t in Equation (10) is replaced by $\sigma_t = (1 + 0.4\sigma_{t-1}^2 u_{t-1}^2)^{1/2}$, as well as designs in which τ_t is a deterministic trigonometric function of a rescaled time index (and $\sigma_t = 1$). Since the results are not materially different from those obtained with the GARCH specification, we do not report them here.

[7] Similar results are obtained for $\varphi = -0.8$.

[8] These choices of λ are of the order $T^{1/3}$, which is known to be optimal in certain respects (see Shao 2010).

[9] The bootstrap-based version of the test proposed in Grivas (2023) leads to the same conclusion.

[10] Only estimates of the coefficients on inflation and the output gap are shown in the plot to conserve space, but more detailed results are available upon request.

[11] This is consistent with the finding of CNT that interest-rate policy smoothing is particularly prominent in the pre-Volcker and Greenspan–Bernanke periods.

[12] For constant-coefficient variants of a Taylor-rule specification like (16), CNT argue in favor of using LS instead of IV when the contribution of monetary-policy shocks to the variance of the covariates is not substantial (as tends to be the case in practice).

References

Burman, P., E. Chow, and D. Nolan. 1994. "A Cross-Validatory Method for Dependent Data." *Biometrika* 81, no. 2: 351–358.

Cai, Z. 2007. "Trending Time-Varying Coefficient Time Series Models With Serially Correlated Errors." *Journal of Econometrics* 136, no. 1: 163–188.

Carvalho, C., F. Nechio, and T. Tristão. 2021. "Taylor Rule Estimation by OLS." *Journal of Monetary Economics* 124: 140–154.

Chen, B. 2015. "Modeling and Testing Smooth Structural Changes With Endogenous Regressors." *Journal of Econometrics* 185, no. 1: 196–215.

Chen, B., and Y. Hong. 2012. "Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression." *Econometrica* 80, no. 3: 1157–1183.

Chronopoulos, I., L. Giraitis, and G. Kapetanios. 2022. "Choosing Between Persistent and Stationary Volatility." *Annals of Statistics* 50, no. 6: 3466–3483.

Clarida, R., J. Galí, and M. Gertler. 2000. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *Quarterly Journal of Economics* 115, no. 1: 147–180.

Davidson, R., and J. G. MacKinnon. 2010. "Wild Bootstrap Tests for IV Regression." *Journal of Business & Economic Statistics* 28, no. 1: 128–144.

Djogbenou, A., S. Gonçalves, and B. Perron. 2015. "Bootstrap Inference in Regressions With Estimated Factors and Serial Correlation." *Journal of Time Series Analysis* 36, no. 3: 481–502.

Faraway, J. J. 1990. "Bootstrap Selection of Bandwidth and Confidence Bands for Nonparametric Regression." *Journal of Statistical Computation and Simulation* 37, no. 1–2: 37–44.

Giraitis, L., G. Kapetanios, and M. Marcellino. 2021. "Time-Varying Instrumental Variable Estimation." *Journal of Econometrics* 224, no. 2: 394–415.

Giraitis, L., G. Kapetanios, and T. Yates. 2014. "Inference on Stochastic Time-Varying Coefficient Models." *Journal of Econometrics* 179, no. 1: 46–65.

González Manteiga, W., M. D. Martínez Miranda, and A. Pérez González. 2004. "The Choice of Smoothing Parameter in Nonparametric Regression Through Wild Bootstrap." *Computational Statistics & Data Analysis* 47, no. 3: 487–515.

Grivas, C. 2023. "Testing for Time-Varying Exogeneity: A Bootstrap Approach," Available at SSRN 4595905.

Hall, P. 1990. "Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems." *Journal of Multivariate Analysis* 32, no. 2: 177–203.

Hall, P., S. N. Lahiri, and J. Polzehl. 1995. "On Bandwidth Choice in Nonparametric Regression With Both Short- and Long-Range Dependent Errors." *Annals of Statistics* 23, no. 6: 1921–1936.

Härdle, W., P. Hall, and J. S. Marron. 1988. "How Far Are Automatically Chosen Regression Smoothing Parameters From Their Optimum?" *Journal of the American Statistical Association* 83, no. 401: 86–95.

Hart, J. D. 1991. "Kernel Regression Estimation With Time Series Errors." *Journal of the Royal Statistical Society B* 53, no. 1: 173–187. https://doi.org/10.1111/j.2517-6161.1991.tb01816.x.

Hurvich, C. M., J. S. Simonoff, and C.-L. Tsai. 1998. "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion." *Journal of the Royal Statistical Society B* 60, no. 2: 271–293.

Köhler, M., A. Schindler, and S. Sperlich. 2014. "A Review and Comparison of Bandwidth Selection Methods for Kernel Regression." *International Statistical Review* 82, no. 2: 243–274.

Nicholls, D. F., and A. R. Pagan. 1985. "Varying Coeefficient Regressions." In *Handbook of Statistics*, edited by E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, vol. 5, 413–449. Elsevier Science Publishers.

Opsomer, J., Y. Wang, and Y. Yang. 2001. "Nonparametric Regression With Correlated Errors." *Statistical Science* 16, no. 2: 134–153.

Robinson, P. M. 1989. "Nonparametric Estimation of Time-Varying Parameters." In *Statistical Analysis and Forecasting of Economic Structural Change*, edited by P. Hackl, 253–264. Springer.

Robinson, P. M. 1991. "Time-Varying Nonlinear Regression." In *Economic Structural Change*, edited by P. Hackl and A. H. Westlund, 179–190. Springer.

Shao, X. 2010. "The Dependent Wild Bootstrap." *Journal of the American Statistical Association* 105, no. 489: 218–235.

Teräsvirta, T. 1998. "Modelling Economic Relationships With Smooth Transition Regressions." In *Handbook of Applied Economic Statistics*, edited by A. Ullah and D. E. A. Giles, 507–552. Marcel Dekker.

Wu, C. F. J. 1986. "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis." *Annals of Statistics* 14, no. 4: 1261–1295. https://doi.org/10.1214/aos/1176350142.

Zhang, T., and W. B. Wu. 2012. "Inference of Time-Varying Regression Models." *Annals of Statistics* 40, no. 3: 1376–1402.