Empirical Economics (2023) 65:3001–3026
https://doi.org/10.1007/s00181-023-02450-6

REPLICATION STUDY

Kernel-based time-varying IV estimation: handle with care

Riccardo Lucchetti[1] · **Francesco Valentini**[1]

Received: 1 July 2022 / Accepted: 1 June 2023 / Published online: 27 June 2023

Abstract
Giraitis et al. (J Econom 224(2):394–415, 2021) proposed a kernel-based time-varying coefficients IV estimator. By using entirely different code, we broadly replicate the simulation results and the empirical application on the Phillips curve, but we note that a possible oversight might have affected some of the reported results. Further, we extend the results by using a different sample and a wider choice of smoothing kernels, including data-based ones; we find that the estimator is remarkably robust across a wide range of smoothing choices, but the effect of outliers may be less obvious than expected.

Keywords Instrumental variables · Time-varying parameters · Hausman test · Phillips curve

JEL Classification C14 · C26

1 Introduction

The issue of estimating linear models when the underlying data generation process may be unstable through time is a long-standing one in the econometric literature. One solution, adopted since the pioneering contribution by Chow (1960), is to assume that abrupt breaks take place at some point in time, and this idea has been extended and generalized in several directions (see, e.g., Bai and Perron 2003).

An increasingly common proposal is to assume instead that the model parameters change continuously through time. The standard solution is to rely on the state-space

Riccardo Lucchetti and Francesco Valentini have contributed equally to this work.

✉ Francesco Valentini
f.valentini@univpm.it

Riccardo Lucchetti
r.lucchetti@univpm.it

[1] DISES, Universitá Politecnica delle Marche, Piazzale R. Martelli 8, 60121 Ancona, AN, Italy

representation and use the Kalman filter[1] apparatus under the Gaussianity assumption (see, e.g., Hamilton 1994, Section 13.8), but other approaches based on moments estimators (Schlicht 2021), flexible least squares (Kalaba and Tesfatsion 1989), penalized least squares (Hastie and Tibshirani 1993) and a pure nonparametric framework (Vogt 2012) have also been put forward.

In recent years, a method has been proposed that deals with time-varying parameter models in a nonparametric way, where the evolution of the parameters through time can be assumed to be "slow."[2] In a series of works by Giraitis, Kapetanios and several co-authors (Giraitis et al. 2014, 2018; Kapetanios et al. 2019), the idea is pursued to generalize the concept of rolling-window regression to kernel-based inference.

One of the advantages of the Giraitis et al. (2014) approach is that it is easily generalizable to many linear models, among which instrumental variable models. IV models with time-varying parameters can, in principle, be handled via a state-space approach, but implementation is far from trivial; for a Bayesian alternative, see Ruisi (2019). None of these approaches, however, match the computational simplicity and flexibility of the kernel-based estimator. On the other hand, in semi- and nonparametric estimation methods one usually has to choose the precise details of the practical implementation of the estimators, and while many of these choices are asymptotically equivalent, in finite samples the numerical results are often affected to quite a large degree.

In this article, we replicate and extend the results presented in Giraitis et al. (2021) (GKM from now on), using a completely independent software implementation[3] and publicly available data. Section 2 briefly describes the model and establishes notation. In Sect. 3, we perform a narrow-sense replication of the original results, while in Sect. 4, we extend the original results in several directions. We find that results are, all in all, rather robust; however, in some cases they can be markedly different across equally legitimate choices. We also find that replicating the original empirical exercise with more recent data, including the COVID-19 pandemic, shows that when large outliers are present, interpretation of the results may be far from obvious. Therefore, it is advisable for practitioners to take great care with the robustness of their results.

2 The model

We consider the model, for $t = 1 \dots T$,

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_t + u_t, \qquad \boldsymbol{x}_t = \boldsymbol{z}_t' \boldsymbol{\psi}_t + \boldsymbol{v}_t, \tag{1}$$

where y_t is the dependent variable, $\boldsymbol{x}_t$ is a $p \times 1$ vector of (possibly endogenous) regressors, $\boldsymbol{z}_t$ denotes a $q \times 1$ vector of instruments, and the parameters $\boldsymbol{\beta}_t$ and $\boldsymbol{\psi}_t$ are allowed to vary over time. The coefficients are assumed to be either smoothly

[1] See Appendix F for a comparison with the technique discussed below.

[2] For a precise definition see, e.g., Giraitis et al. (2018), Section 2.1.

[3] The code we used for this exercise is publicly available as the `ketvals` gretl package.

varying deterministic functions or smoothly varying persistent stochastic processes (see Assumptions 2 and 3 in GKM, respectively).

Let $0 < r \leq p$ the number of elements of $\boldsymbol{x}_t$ that are deemed to be endogenous, so $\boldsymbol{x}_t$ and $\boldsymbol{z}_t$ have $(p - r)$ elements in common. This is a slight generalization to GKM, where $p = r$ is assumed throughout. The disturbance terms u_t and $\boldsymbol{v}_t$ may be correlated, in which case endogeneity arises. We focus on two different estimators proposed in GKM: an OLS estimator $\hat{\boldsymbol{\beta}}_t$ and an IV estimator $\tilde{\boldsymbol{\beta}}_t$. For their definition, see Sect. A.1 in Appendix A.

In addition, GKM define a time-varying equivalent of the classic Hausman test, for the null hypothesis $H_0 : E(u_t \boldsymbol{v}_t) = \mathbf{0}$, and a "Global Hausman Test" in which the hypothesis is tested in a given time span $[T_0, T_1]$. For their definitions, see again Sect. A.1 in Appendix A. Under the null hypothesis, they are both distributed as χ^2 variates. GKM focus on the $p = r$ case, the degrees of freedom of the χ^2 limiting distribution of the Hausman test is given by p, the number of explanatory variables (equation (35) in GKM). However, in the more general case that we analyze ($0 < r \leq p$) the degrees of freedom equal the number of endogenous variables r (see Appendix B for a proof).

It should be noted that the p-values for the time-varying Hausman test are computed for individual (local) tests, and when one computes a test statistic at each time then one runs into a multiple testing problem requiring adjustments of critical values. In light of this, it may make sense to first compute the global Hausman test and then, in case of rejection of the global null, rejections from the time-specific Hausman statistics would point to specific time periods when violation of the null occurs.[4]

3 Narrow-sense replication

This section summarizes the narrow-sense replication of the results in GKM. For reasons of space, most details are given in Appendices.

3.1 Simulations

In the first part of their article, GKM perform several simulation experiments on the univariate data generating process (DGP) $y_t = x_t \beta_t + u_t$ to assess the quality of the time-varying OLS and IV estimators over a range of possible DGPs. For the precise details of the DGP from which data are simulated, we refer the reader to Sect. 3 in GKM.

We briefly describe the DGP and report the annotated results in Appendix C. Suffice it to say here that the simulation results can be replicated almost exactly; in fact, in some cases, our Monte Carlo results are even closer to the asymptotic figures than the original ones.

[4] We thank an anonymous referee for pointing out this aspect.

3.2 Empirical application: the Phillips curve

In this section, we replicate the empirical application reported in Sect. 4 of GKM, where a time-varying version of the Phillips curve is estimated. The following two models are considered: a "traditional," backward-looking version

$$\Delta\pi_t = \mu_t + \gamma_t \Delta\pi_{t-1} + \alpha_t \Delta u_t, \tag{2}$$

and a "New-Keynesian" version where a forward-looking term is added:

$$\Delta\pi_t = \mu_t + \gamma_t \Delta\pi_{t-1} + \alpha_t \Delta u_t + \rho_t \Delta\pi_{t+1}. \tag{3}$$

Here, $\Delta\pi_t$ and Δu_t are the first differences of inflation and unemployment, respectively. GKM do not point to a downloadable dataset, but we consider the usage of the St. Louis FED data as a reasonable reconstruction. To be specific, π_t is 100 times the seasonal log difference of the `CPIAUCSL` variable; unemployment u_t is simply a copy of the `UNRATE` variable. We use the sample from 1959:02 to 2013:07.

In Eq. (2), Δu_t is considered endogenous and the instruments are the lags of Δu_t from 1 to 4. In Eq. (3), $\Delta\pi_{t+1}$ is also considered endogenous and the set of instruments is augmented with four lags of $\Delta\pi_t$.

Figure 1 shows the replication results. For Eq. (2), the time paths of estimated coefficients and the associated confidence bands can be replicated almost exactly. We note only very minor discrepancies near the end of the sample, which can be safely attributed to differences in the data vintage used.

On the other hand, the plot for the time-varying Hausman test appears to be rather different, with much lower p-values for the best part of the sample. We believe the discrepancy in Hausman test p-values might be due to the different number of degrees of freedom of the limiting χ^2 distribution considered by GKM. The situation where only Δu_t is considered to be potentially endogenous marks a departure from the standard assumption in GKM ($p = r$) since the number of explanatory variables ($p = 3$) does not coincide with that of endogenous ones ($r = 1$). To substantiate our claim, Fig. 2 compares the time path of the p-value for the Hausman test obtained using the correct number of degrees of freedom ($r = 1$) and the one obtained using the total number of regressors ($r = p = 3$). As can be seen, the time path of the p-value with 3 degrees of freedom is much closer to the one reported in GKM at the end of Section 4.

As for the "New-Keynesian" Phillips curve in Eq. (3), we observe noticeable differences between our results and the original ones. For the three parameters in the model, we are able to almost perfectly replicate the time paths and the confidence intervals of the least squares estimator $\hat{\boldsymbol{\beta}}_t$. On the contrary, the IV estimates appear to be quite different all over the sample, particularly at the beginning and the end of the time span considered. Since the differences in the OLS estimates are relatively minor, we conjecture that the IV estimator, in an over-identified case such as the present one, may be somewhat sensitive to individual data points. This conjecture is consistent with the evidence we present in Sect. 4.3, where we show the effects on the estimates of one

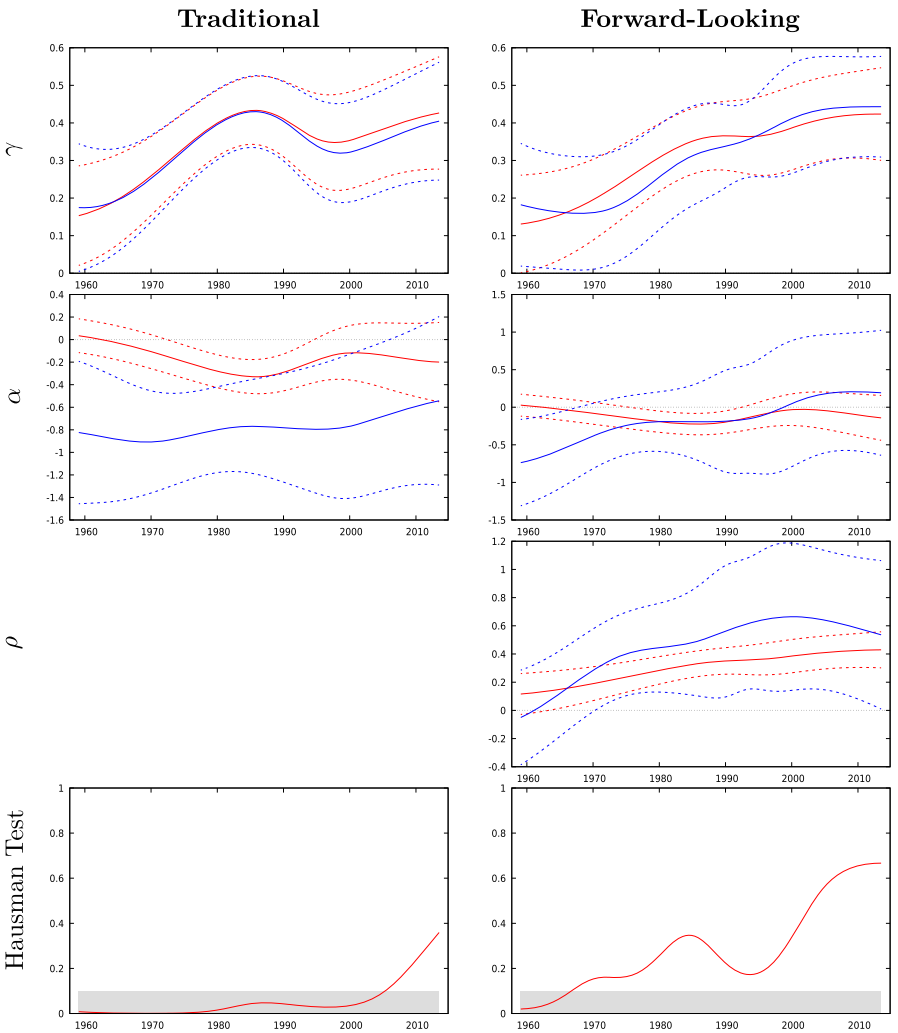

Fig. 1 Replication, $h_1 = h_2 = 0.7$. Note: the two columns report estimated coefficients and the Hausman test (p-value) related to Eqs. (2) and (3), respectively. With reference to coefficients, the red line denotes the OLS point estimates, while the blue one represents the IV estimator. Dotted lines, with the same colors, represent the boundaries of the 90% confidence intervals. Shaded area denotes the region for p-value $\leq 10\%$. (Color figure online)

extreme outlier, given by the unemployment surge of April 2020. Given these premises, it is obvious that the Hausman test results could not be replicated satisfactorily.

In order to provide the complete set of the results shown by GKM, Fig. 8 in Appendix D reports the replication of the robustness check on the "traditional" Phillips curve, where the parameters h_1 and h_2 are set to 0.5. Similarly to our previous results, the time paths of the coefficients are very similar to the reported ones, although the Hausman p-value plot is not.

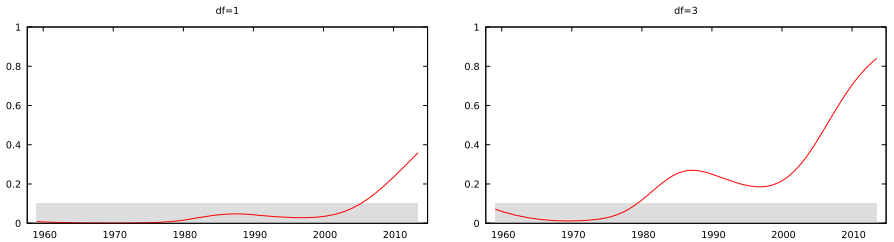

Fig. 2 Hausman test: comparing degrees of freedom. Note: the two panels report the Hausman test p-value related to the model in Eq. (2), assuming a limiting χ^2-distribution with 1 (left) and 3 (right) degrees of freedom. Shaded area denotes the region for p-value $\leq 10\%$

Finally, it is worth recalling that the whole replication exercise is based on data provided by the St. Louis FED, while it is not clear where GKM data come from. As a consequence, small differences between our results and the original ones may be due to the nature of the data.

4 Extended replication

In this section, we extend the empirical example proposed by GKM by investigating robustness of results with respect to two separate issues, namely (a) different choices of the kernel weighting scheme and (b) an extension of the sample used by the authors. As will be argued, both these aspects reveal interesting features of the estimator.

4.1 Kernel functions and bandwidth

In this subsection, we replicate the empirical analysis provided in Sect. 3.2 by using four different kernel functions, namely the Gaussian, Epanechnikov and exponential kernels, as well as the rectangular kernel, which is equivalent to a simple rolling-window estimator.

In GKM, the Gaussian kernel was used, where the kernel weighting function is $K(x) = \exp(-x^2/2)$ and $x = |j - t|/T^h$. The exponential kernel uses the Laplace density instead: $K(x) = \exp(-|x|)$.[5] These two kernels have infinite support, so that all observations contribute to the calculation of each element of the $\hat{\boldsymbol{\beta}}_t$ and $\tilde{\boldsymbol{\beta}}_t$ sequences. The other two, instead, have finite support: for $|x| < 1$, $K(x) = 0.75(1 - x^2)$ for the Epanechnikov kernel and $K(x) = 1$ for the rectangular one; both are 0 for $|x| \geq 1$. This implies that in the calculation of $\hat{\boldsymbol{\beta}}_t$ and $\tilde{\boldsymbol{\beta}}_t$ only "nearby" data matter (although of course with infinite-support kernels small bandwidths achieve the same effect in practice). As will be seen, the distinction between the two kinds has noticeable empirical implications.

Figure 3 reports the estimated coefficients and their confidence intervals, plus the Hausman test p-value for the model in Eq. (2). As a benchmark, we consider the

[5] In fact, more general formulations are possible: GKM, in the original paper, consider using $K(x) = \exp(-c \cdot |x|^\alpha)$.

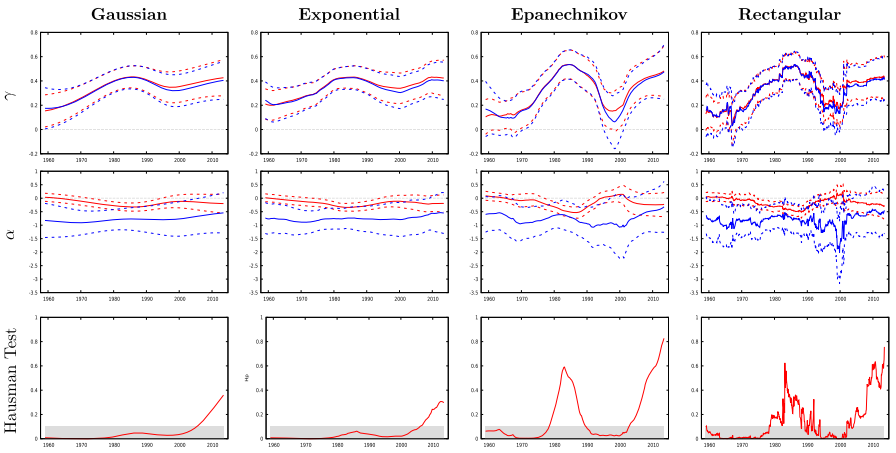

Fig. 3 Traditional PC: 1959:02–2013:07, $h_1 = h_2 = 0.7$. Note: Each column reports estimated coefficients and the Hausman test (p-value) related to Eq. (2). With reference to coefficients, the red line denotes the OLS point estimates, while the blue one represents the IV estimator. Dotted lines, with the same colors, represent the boundaries of the 90% confidence intervals. (Color figure online)

same bandwidth parameters used in GKM, that is $h_1 = h_2 = 0.7$. Using the original sample, the results appear to be qualitatively robust to the choice of kernels. Magnitude of coefficients and confidence intervals are similar across the different smoothing functions, even though estimates relying on finite-support kernels (i.e., rolling window and Epanechnikov) appear more volatile through time and display spiky confidence intervals. As for the Hausman test, results obtained with finite-support kernel functions appear less stable over time. However, this finding has to be considered together with the results in Sect. 4.3 later.

Setting the bandwidth parameters to $h_1 = h_2 = 0.5$ yields the results reported in Fig. 9 in Appendix E. Here too, the time path of all estimators appears to be quite unstable, even though point estimates are comparable to those in Fig. 3. On the contrary, confidence intervals and the Hausman test are rather sensitive to the bandwidth choice: estimates obtained by using finite-support kernels appear to be volatile at the end of the sample and the time-varying p-value for the Hausman test is markedly unstable.

4.2 Data-driven bandwidth selection

As is well known, the finite sample performance of nonparametric estimators and tests depends on bandwidth selection. Moreover, a rigid bandwidth choice, as the one adopted in GKM, may work well for particular signal-to-noise ratios, but not in general.[6] Therefore, we explored the effects of using a data-driven method, namely the corrected AIC proposed by Cai and Tiwari (2000) and Cai (2007). The optimal

[6] We thank an anonymous referee for pointing out this aspect.

Table 1 Phillips curve estimation using data-driven bandwidth

	$h_2 = 0.5$	$h_2 = 0.7$
Traditional PC	0.6758	0.9899
New-Keynesian PC	0.8872	0.8526

bandwidth parameter h_{opt} is the minimizer of $AIC(h)$:

$$AIC(h) = \log(\hat{\sigma}^2) + \frac{2 \cdot (T_h + 1)}{(T - T_h - 2)},$$

where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2$ and where T_h is the trace of the "smoothing matrix" $\boldsymbol{H}_h$, defined in Sect. A.2 in Appendix A.

This method can be applied to the time-varying OLS estimator in a straightforward manner, but its application to IV estimation is more problematic, since it is unclear how to set up the objective function to cover both the first- and second-stage models. The solution we have adopted is to apply the AIC criterion for the selection of the second-stage bandwidth parameter h_1, given the first-stage parameter h_2, which we keep at the values originally used in GKM. Table 1 contains the results for the two models in Eqs. (2) and (3).

As can be seen, the bandwidth parameters selected via AIC are quite large for both models, therefore yielding estimates that are much more stable trough time. Figure 4 compares, for the traditional Phillips curve, the IV estimates obtained with the automatic bandwidth selection and those given by setting $h_1 = h_2$.

Finally, it should be noted that the bandwidth parameter governs the degree of time-variability of the estimators, which in many cases is an aspect on which the analyst may have strong a priori beliefs, dictated by the details of the problem at hand. In particular, economic reasoning can be used to evaluate ex post how credible the estimated time path of a coefficient is. Therefore, committal to a data-driven approach for bandwidth selection should in most cases be supplemented by traditional robustness checks.

4.3 Extending the sample

We now investigate the effects of a sample extension. Figure 5 shows the results for Eq. (2) based on observations ranging from 1948:06 to 2022:09 from the St. Louis FED dataset. This experiment is particularly interesting, because it includes data from the COVID-19 pandemic. As we will show, this provides a very interesting "natural experiment" to assess the performance of the estimator under extreme conditions.[7]

The results we obtain, in terms of the robustness of results, are mixed. On the one hand, the parameter estimates appear to be quite robust to the choice of the time span, apart from small differences around 1960 and 2013, due to the new observations at the

[7] Clearly, the 2020 unemployment spike had nothing to do with changing structural coefficients and everything to do with an exogenous shock that had a dramatic impact on unemployment but a negligible impact on inflation. However, we believe this is a perfect illustration of the numerical consequences on the estimates of one large outlier.

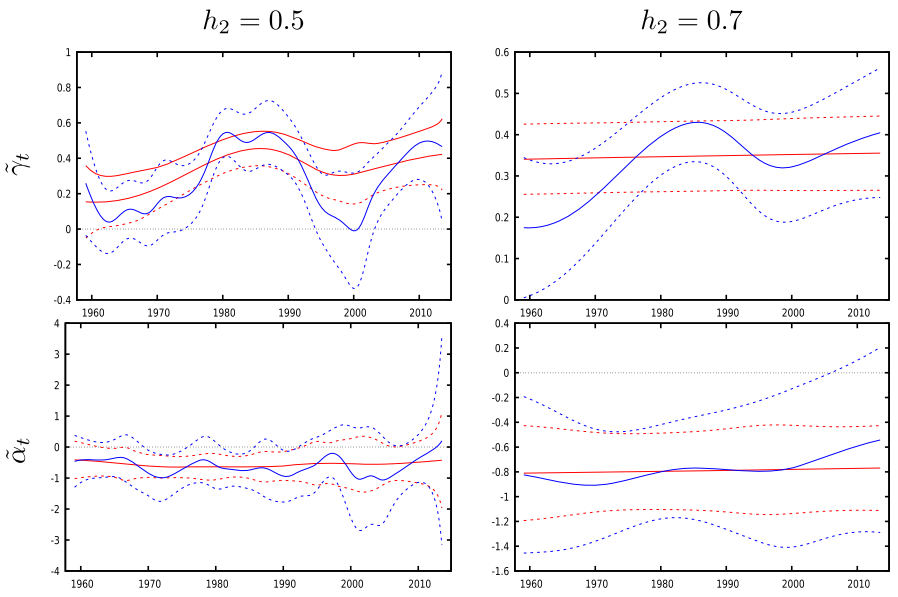

Fig. 4 Traditional PC: automatic bandwidth selection. Note: each plot compares the IV estimation results obtained with a "second-step" bandwidth $h_1 = h_2$ (blue line) and those resulting from the automatic selection procedure (red line, see Table 1). Dotted lines denote 90% confidence intervals. (Color figure online)

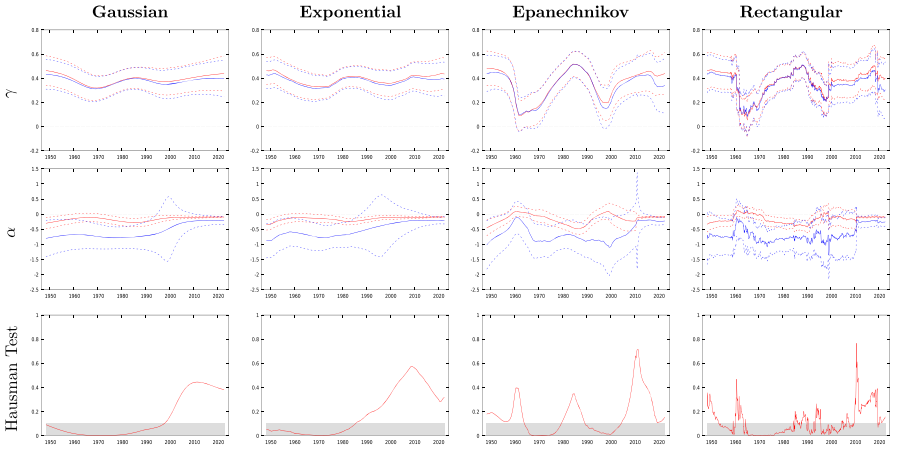

Fig. 5 Traditional PC: 1948:06–2022:09, $h_1 = h_2 = 0.7$. Note: see Fig. 3. (Color figure online)

beginning and at the end of the sample. On the other hand, standard errors are quite different from the original ones throughout the sample span. Using infinite-support kernels (Gaussian and exponential), the confidence intervals for the parameter α_t in the late 1990s appear much larger than the original ones. This is hardly surprising, given that the standard errors depend on the first stage residuals: extending the sample to include the COVID-19 pandemic has a dramatic effect on the estimated standard errors.

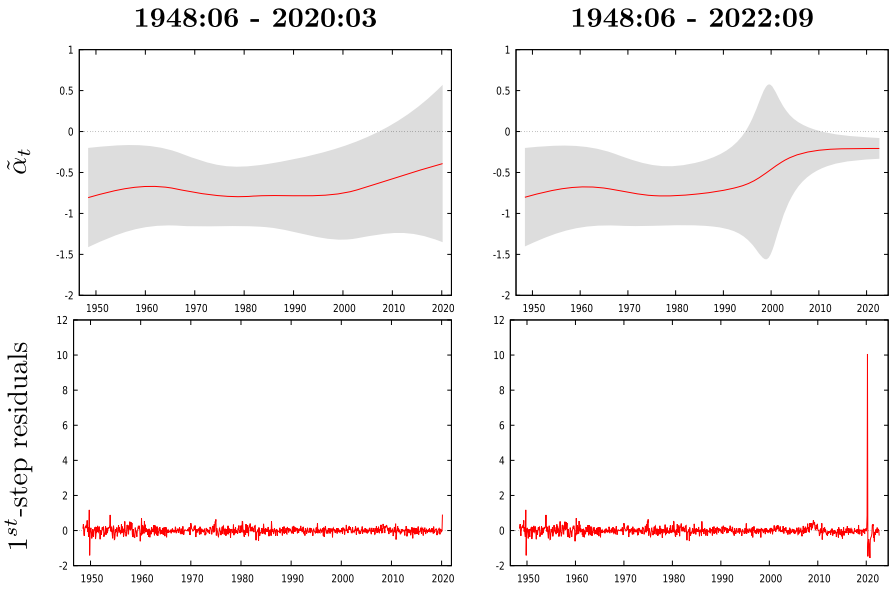

Fig. 6 Excluding Covid-19. Note: the two columns report IV estimates and first-stage residuals of the endogenous variable Δu_t, obtained by excluding (left) the Covid-19 period and with the full sample (right). Gaussian kernel function, bandwidth parameters $h_1 = h_2 = 0.7$. Shaded area denotes 90% confidence interval

In order to show this effect, Fig. 6 reports IV estimates for α_t and first stage residuals for Δu_t obtained by excluding the COVID-19 period (left) and by using the full sample to 2022:09 (right). In the pre-COVID-19 sample, residuals do not contain any outliers and the confidence interval of the coefficient is very similar to the one in GKM. When using the full dataset, instead, the first-stage residuals display a huge spike in April 2020, as an effect of the unemployment surge which took place during the first wave of the pandemic: the most striking consequence is the effect this has on estimated standard errors for a wide range of observations.[8]

With infinite-support kernels, the "COVID-19 spike" provokes an effect on confidence bands that extends for a very long period, with features that are somewhat counter-intuitive. The standard error for α_t reaches its maximum around the year 2000, i.e., nearly 20 years *before* the pandemic, whereas a similar effect cannot be seen for the standard errors for the 2020–2022 period. In other terms, the influence of a single outlier on the estimates is markedly asymmetric with respect to time. On the contrary, the Hausman test performs satisfactorily in terms of robustness and provides stable results whatever the sample, with substantial discrepancies in the results obtained only with a Rolling Window kernel during the 60's and the 80's.

[8] As an anonymous referee pointed out, it could be argued that the COVID-19 pandemic is not, in principle, the only example of crisis episode that could exert a similar impact: for example, one could also take the 1973 oil crisis or the 2007–2008 financial crisis. In this context, however, we consider the COVID-19 pandemic to be qualitatively different given the enormous impact it had on the unemployment rate, compared to the past. The right-bottom pane of Fig. 6 is quite telling in this respect.

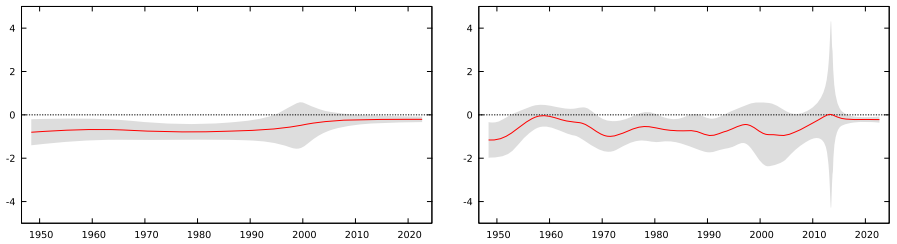

Fig. 7 Bandwidth: local effects on standard errors. Note: the two panels report estimated $\tilde{\alpha}_t$ and 90% confidence intervals for the model in Eq. (2), the kernel is Gaussian. The left panel is obtained with $h_1 = h_2 = 0.7$ and right one with $h_1 = h_2 = 0.5$. Shaded area denotes 90% confidence interval

Unsurprisingly, the adverse effects shown above are somewhat mitigated by kernel choices that put a greater weight on "local" data when calculating the time-varying quantities: for example, a Gaussian kernel with a narrower bandwidth (see Fig. 7) or finite-support kernels (see Fig. 13 in Appendix). Our overall impression is that, in normal times, the estimator is quite robust across a wide range of possible choices, but in the presence of possible abrupt breaks or otherwise exceptional events much care is needed by the investigator when choosing the precise details of the smoothing scheme to use.

5 Conclusions

We have presented a replication of the results provided by Giraitis et al. (2021) for their kernel-based instrumental variable regression estimator, in both a narrow and a wide sense.

In the narrow-sense replication, the Monte Carlo experiment and the real data application on the Phillips curve are replicated almost exactly. As for the empirical example, we came across a possible oversight about the degrees of freedom of the Hausman test which might have affected the results shown in GKM.

The wide sense replication focuses on the robustness of results to different aspects: we use different smoothing strategies (including data-driven ones), as well as a larger sample and show the effects on the results. As for the choice of the smoothing kernel, results seem to be sensitive mainly to the choice of finite-support versus infinite-support kernels. While the latter provide results that are more readily interpretable, the former are more robust to the presence of outliers, which can affect results quite dramatically.

We find that parameter estimates are, in general, strongly robust. On the contrary, the standard errors and the Hausman test appear to be quite sensitive to the bandwidth parameter and to the presence of extreme shocks in the series considered, such as the COVID-19 pandemic period. Although investigating this aspect is outside the scope of the present paper, we believe that further work on the robustness of nonparametric methods with respect to outliers would be very important.

In conclusion, we believe that the time-varying IV estimator put forward by GKM is a very valuable tool that displays remarkable robustness properties under general

conditions. However, it is advisable to carry out a comprehensive set of robustness checks in empirical applications, since the consequences of data outliers and/or kernel weighting choices may be far from obvious.

Acknowledgements We wish to thank the Associate Editor, Allin Cottrell, Giulio Palomba, Ekkehart Schlicht and two anonymous referees for their comments. Francesco Valentini would like to acknowledge the financial support by the European Union - Next Generation EU. Project Code: ECS00000041; Project CUP: C43C22000380007; Project Title: Innovation, digitalization and sustainability for the diffused economy in Central Italy - VITALITY.

Funding Open access funding provided by Università Politecnica delle Marche within the CRUI-CARE Agreement.

Data and computer code availability The datasets of this paper (1. code and programs, 2. Data, 3. detailed readme files) are available in the GitHub repository https://github.com/fravale/replication_gkm.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Appendix A: Definitions and notation

A.1 Estimators and the Hausman test

The two estimators from GKM that we analyze in this paper are

1. The time-varying OLS estimator

$$\hat{\boldsymbol{\beta}}_t = \left(\sum_{j=1}^{T} b_{H,|j-t|} \boldsymbol{x}_j \boldsymbol{x}_j' \right)^{-1} \left(\sum_{j=1}^{T} b_{H,|j-t|} \boldsymbol{x}_j y_j \right) \tag{A1}$$

2. The time-varying IV estimator[9]

$$\tilde{\boldsymbol{\beta}}_t = \left(\sum_{j=1}^{T} b_{H,|j-t|} \hat{\boldsymbol{\psi}}_j' \boldsymbol{z}_j \boldsymbol{x}_j' \right)^{-1} \left(\sum_{j=1}^{T} b_{H,|j-t|} \hat{\boldsymbol{\psi}}_j' \boldsymbol{z}_j y_j \right) \tag{A2}$$

[9] Denoted $\tilde{\beta}_{1,t}$ in GKM.

where $b_{H,|j-t|} = K\left(\frac{|j-t|}{H}\right)$ is a kernel weight with bandwidth parameter $H = T^{h_1}$ and $\hat{\boldsymbol{\psi}}_t$ is used to indicate the time-varying first-stage OLS estimates of $\boldsymbol{\psi}_t$, possibly allowing for a different kernel bandwidth $L = T^{h_2}$ in the first stage regression. In GKM, another IV estimator is also analyzed, namely $\tilde{\boldsymbol{\beta}}_{2,t} = \left(\sum_{j=1}^{T} b_{H,|j-t|}\hat{\boldsymbol{\psi}}_t' \boldsymbol{z}_j \boldsymbol{x}_j'\right)^{-1}\left(\sum_{j=1}^{T} b_{H,|j-t|}\hat{\boldsymbol{\psi}}_t' \boldsymbol{z}_j y_j\right)$, that is asymptotically equivalent to $\tilde{\boldsymbol{\beta}}_t$; however, we do not consider it in our replication since the empirical application focuses on $\tilde{\boldsymbol{\beta}}_t$, which appears to be the preferred form.

Note that both estimators can be written as

$$\bar{\boldsymbol{\beta}}_t = \left[\mathbf{W}_t'\mathbf{X}\right]^{-1}\mathbf{W}_t'\mathbf{y}, \tag{A3}$$

where $\mathbf{W}_t$ and $\mathbf{X}$ are matrices with T rows and p columns and $\mathbf{y}$ is a T-dimensional vector. While the matrices $\mathbf{X}$ and $\mathbf{y}$ simply contain the observations of the explanatory and dependent variables, respectively, the $\mathbf{W}_t$ matrix contains kernel-weighted entries.

For example, in the case of the time-varying OLS estimator $\hat{\boldsymbol{\beta}}_t$ (eq. (11) in GKM), the s-th row of $\mathbf{W}_t$ equals

$$[\mathbf{W}_t]_s' = b_{H,|s-t|}\boldsymbol{x}_s', \tag{A4}$$

or, in matrix notation (see Kapetanios et al. (2019)),

$$\mathbf{W}_t^{OLS} = D_t\mathbf{X}, \tag{A5}$$

where D_t is a diagonal $T \times T$ matrix whose element $[D_t]_{i,i}$ equals

$$[D_t]_{i,i} = b_{H,|i-t|}. \tag{A6}$$

The time-varying Hausman test is given by the following quadratic form:

$$\frac{K_t^2}{K_{2,t}} V_{T,t}'\hat{\Sigma}_{\hat{v}\hat{v},t}^{-1} V_{T,t}\hat{\sigma}_{\hat{u},t}^{-2} \xrightarrow{\mathrm{d}} \chi_r^2, \tag{A7}$$

where

$$\begin{aligned}
&K_t = \textstyle\sum_{j=1}^{T} b_{H,|j-t|}; && K_{2,t} = \textstyle\sum_{j=1}^{T} b_{H,|j-t|}^2;\\
&\hat{\Sigma}_{\hat{v}\hat{v},t} = K_t^{-1}\textstyle\sum_{j=1}^{T} b_{H,|j-t|}\hat{v}_j\hat{v}_j'; && \hat{\sigma}_{\hat{u},t}^2 = K_t^{-1}\textstyle\sum_{j=1}^{T} b_{H,|j-t|}\hat{u}_j^2;\\
&V_{T,t} = \left(\textstyle\sum_{j=1}^{T} b_{H,|j-t|}\hat{\boldsymbol{x}}_j\hat{\boldsymbol{x}}_j'\right)^{1/2}\left(\textstyle\sum_{j=1}^{T} b_{H,|j-t|}\boldsymbol{x}_j\boldsymbol{x}_j'\right)^{1/2}\left(\tilde{\boldsymbol{\beta}}_t - \hat{\boldsymbol{\beta}}_t\right),
\end{aligned}$$

$\hat{v}_j$ and $\hat{u}_j$ denote the of the first- and second-stage residuals, respectively, and $\hat{\boldsymbol{x}}_j$ are the first-stage fitted values.

Conversely, the "Global Hausman Test" proposed by GKM for a given time span $[T_0, T_1]$ is defined as

$$\mathcal{H}_{T_0,T_1} = \frac{1}{\sqrt{T_1 - T_0}} \sum_{t=T_0+1}^{T_1} K_t K_{2,t}^{-1} \hat{\sigma}^{-1} \hat{\Sigma}_{\hat{v}\hat{v},t}^{-1/2} V_{T,t},$$
$$H = \mathcal{H}'_{T_0,T_1} \mathcal{H}_{T_0,T1} \xrightarrow{\text{d}} \chi^2_r. \tag{A8}$$

A.2 Smoothing matrix

Consider the series of the fitted values of the dependent variable $\hat{\mathbf{y}} = (\hat{y}_1, \ldots, \hat{y}_T)'$ obtained from the model in Eq. (1). This T-dimensional vector can be expressed as

$$\hat{\mathbf{y}} = \boldsymbol{H}_h \mathbf{y}, \tag{A9}$$

where the "smoothing matrix" $\boldsymbol{H}_h$ mentioned in Sect. 4.2 is given by

$$\boldsymbol{H}_h = \begin{bmatrix} \boldsymbol{x}'_1 & & & \\ & \boldsymbol{x}'_2 & & \\ & & \ddots & \\ & & & \boldsymbol{x}'_T \end{bmatrix} \begin{bmatrix} \boldsymbol{G}_1 \\ \boldsymbol{G}_2 \\ \vdots \\ \boldsymbol{G}_T \end{bmatrix} \tag{A10}$$

using the same notation of Eq. (A3), with

$$\boldsymbol{G}_t = \left[\mathbf{W}'_t \mathbf{X}\right]^{-1} \mathbf{W}'_t, \tag{A11}$$

depending on h through $\mathbf{W}_t^{OLS} = D_t \mathbf{X}$ or $\mathbf{W}_t^{IV} = D_t \hat{\mathbf{X}}$, with $\hat{\mathbf{X}}$ denoting the fitted values in the "first stage" regression.

Appendix B: The behavior of the time-varying Hausman test in presence of both exogenous and endogenous covariates

In this appendix, we first provide a proof concerning the degrees of freedom of the Hausman test discussed in Sect. 2 and then we perform a simulation exercise on an ad hoc data generating process that includes both an exogenous and an endogenous variable.

B.1 Proof that the degrees of freedom of the time-varying Hausman test is *r*

Using the notation in Sect. 2 and Appendix A, the IV estimator $\tilde{\boldsymbol{\beta}}_t$ can be written as in Eq. (A3), with

$$\mathbf{W}_t^{IV} = D_t \hat{\mathbf{X}} = D_t \mathbf{Z}^* \hat{\Psi}^*, \tag{B12}$$

where $\mathbf{Z}^*$ and $\hat{\Psi}^*$ are matrices such that

$$\mathbf{Z}^* = \begin{bmatrix} z_1' & & & \\ & z_2' & & \\ & & \ddots & \\ & & & z_T' \end{bmatrix}, \qquad \hat{\Psi}^* = \begin{bmatrix} \hat{\Psi}_1 \\ \hat{\Psi}_2 \\ \vdots \\ \hat{\Psi}_T \end{bmatrix}. \tag{B13}$$

Now suppose that $\mathbf{X}$ can be split as $[\mathbf{X}_1|\mathbf{Y}]$ where $\mathbf{X}_1$ contains the exogenous regressors (that is, those columns that are contained in the space spanned by $\mathbf{Z}$) and $\mathbf{Y}$ the endogenous ones, so that they have $p-r$ and r columns, respectively. The parameter vectors can be split accordingly as

$$\bar{\boldsymbol{\beta}}_t' = [\bar{\boldsymbol{\alpha}}_t|\bar{\boldsymbol{\gamma}}_t]'. \tag{B14}$$

Note that in this case the leftmost $p-r$ columns of $\mathbf{W}_t^{IV}$ coincide with those of $\mathbf{W}_t^{OLS}$, and are equal to $D_t\mathbf{X}_1$.

We now prove that the vector of contrasts $\hat{\boldsymbol{\beta}}_t - \tilde{\boldsymbol{\beta}}_t$, which is the essential ingredient of $V_{T,t}$ in Eq. (A7) (Equation (29) in GKM), can be written as a linear combination of $\hat{\boldsymbol{\gamma}}_t - \tilde{\boldsymbol{\gamma}}_t$, which contains r elements. Consider the OLS and IV residuals for the full sample, based on the estimated parameter vector at time t:

$$\bar{\mathbf{e}}(t) = \mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}}_t = \mathbf{y} - \mathbf{X}_1\bar{\boldsymbol{\alpha}}_t - \mathbf{Y}\bar{\boldsymbol{\gamma}}_t. \tag{B15}$$

For the two estimators considered and any t, the following expression holds:

$$\mathbf{y} - \mathbf{Y}\bar{\boldsymbol{\gamma}}_t = \mathbf{X}_1\bar{\boldsymbol{\alpha}}_t + \bar{\mathbf{e}}(t). \tag{B16}$$

Define the matrix

$$S_t = \left(\mathbf{X}_1' D_t \mathbf{X}_1\right)^{-1} \mathbf{X}_1' D_t, \tag{B17}$$

and premultiply Eq. (B16) by S_t:

$$S_t\left(\mathbf{y} - \mathbf{Y}\bar{\boldsymbol{\gamma}}_t\right) = \bar{\boldsymbol{\alpha}}_t, \tag{B18}$$

since the columns of $D_t\mathbf{X}_1$ are a subset of those of $\mathbf{W}_t$, so $\mathbf{X}_1' D_t \bar{\mathbf{e}}(t) = \mathbf{0}$. Therefore, for the time-varying OLS and IV estimators you get, respectively,

$$S_t\left(\mathbf{y} - \mathbf{Y}\hat{\boldsymbol{\gamma}}_t\right) = \hat{\boldsymbol{\alpha}}_t \tag{B19}$$

$$S_t\left(\mathbf{y} - \mathbf{Y}\tilde{\boldsymbol{\gamma}}_t\right) = \tilde{\boldsymbol{\alpha}}_t \tag{B20}$$

and thus

$$\hat{\boldsymbol{\alpha}}_t - \tilde{\boldsymbol{\alpha}}_t = S_t\mathbf{Y}(\tilde{\boldsymbol{\gamma}}_t - \hat{\boldsymbol{\gamma}}_t). \tag{B21}$$

Table 2 Comparing degrees of freedom for the Hausman test. Rejection rate at $t = T/2$

	h_1	h_2	$T = 100$	$T = 200$	$T = 400$	$T = 1000$
Just-Identified case						
$df = 1$	0.4	0.4	0.029	0.028	0.022	0.026
	0.4	0.5	0.039	0.036	0.029	0.048
	0.5	0.4	0.033	0.031	0.032	0.025
	0.5	0.5	0.034	0.036	0.035	0.041
$df = 2$	0.4	0.4	0.004	0.007	0.006	0.007
	0.4	0.5	0.011	0.012	0.009	0.008
	0.5	0.4	0.005	0.011	0.006	0.007
	0.5	0.5	0.005	0.011	0.009	0.012
Over-Identified case						
$df = 1$	0.4	0.4	0.045	0.047	0.041	0.034
	0.4	0.5	0.052	0.049	0.053	0.049
	0.5	0.4	0.043	0.050	0.036	0.051
	0.5	0.5	0.039	0.052	0.042	0.051
$df = 2$	0.4	0.4	0.006	0.007	0.009	0.008
	0.4	0.5	0.016	0.012	0.016	0.021
	0.5	0.4	0.011	0.015	0.011	0.013
	0.5	0.5	0.012	0.011	0.008	0.011

As a consequence, the p-vector of contrasts $\hat{\boldsymbol{\beta}}_t - \tilde{\boldsymbol{\beta}}_t$ can be written as

$$\hat{\boldsymbol{\beta}}_t - \tilde{\boldsymbol{\beta}}_t = \begin{bmatrix} -S_t\mathbf{Y} \\ I \end{bmatrix} \left(\hat{\boldsymbol{\gamma}}_t - \tilde{\boldsymbol{\gamma}}_t\right), \tag{B22}$$

that is a linear combination of a vector with r elements. Therefore, for any t the covariance matrix of $\left(\hat{\boldsymbol{\beta}}_t - \tilde{\boldsymbol{\beta}}_t\right)$ has rank r, and the claim follows.

B.2 Simulation evidence

We perform a simulation exercise to provide finite sample behavior of the time-varying Hausman test when some of the covariates are considered endogenous.

Our simulation design is similar to the one used in Section C, but we add an exogenous variable, x_t^{exo} to the set of regressors. This additional explanatory variable is drawn from a Gaussian distribution and the related regression parameter follows a $\sqrt{T}$-rescaled random walk.

Table 2 reports the rejection rate of the Hausman test under the null hypothesis of no endogeneity for a nominal size of 5%. We consider several sample sizes T and bandwidth parameters (h_1 and h_2). In the table, df denotes the considered degrees of freedom of the limiting χ^2 distribution.

As we can see, the percentage of rejections for the test obtained with $df = 1$ is close to the nominal size in all the scenarios we consider. As expected, the test with $df = 2$ exhibits a large underrejection rate.

Appendix C: Simulation results

GKM propose two simulation designs based on the model in Eq. (1). The first is one is a just-identified model of the type

$$y_t = x_t\beta_t + u_t, \quad x_t = z_t\psi_t + v_t, \quad t = 1, \ldots, T, \tag{C23}$$

while the second one is an over-identified model

$$y_t = x_t\beta_t + u_t, \quad x_t = z_{1,t}\psi_{1,t} + z_{2,t}\psi_{2,t} + v_t, \quad t = 1, \ldots, T, \tag{C24}$$

where y_t is the dependent variable, x_t is an explanatory variable, and $z_{\cdot,t}$ is the instrument; all are drawn from a Gaussian distribution. The error terms, u_t and v_t, are generated according to

$$u_t = se_{1,t} + (1-s)e_{2,t}, \quad v_t = se_{1,t} + (1-s)e_{3,t}, \tag{C25}$$

where $s = \{0, 0.2, 0.5\}$ governs the correlation between the error terms and where $e_{\cdot,t}$ denotes an i.i.d. Gaussian sequence. Finally, the parameters β_t, $\psi_{1,t}$ and $\psi_{2,t}$ follow a $\sqrt{T}$-rescaled random walk.

In the following Tables 3, 4, 5, 6, 7 and 8, we report the same descriptive statistics used in GKM for both OLS and IV estimators: (1) median deviation; (2) absolute median deviation; (3) interdecile range, and (4) 95% coverage rate. 96 basic scenarios are considered, stemming from the possible combinations of (a) exact identification vs over-identification (b) degree of endogeneity of the x_t explanatory variable: none, moderate and strong (c) sample size $T = 100, 200, 400, 1000$ (d) bandwidth choice: $h_1 = 0.4, 0.5$ and $h_2 = 0.4, 0.5$. Each table refers to a different design and includes different scenarios concerning the sample size T and the bandwidth parameters h_1 and h_2. As in GKM, the number of Monte Carlo replications is set to 1000.

Indicators (1) and (2) are broadly consistent with those of GKM, apart from the OLS endogeneity bias being slightly larger in the just-identified case. On the contrary, the decile range is narrower in our simulation for both OLS and IV, which seems to indicate higher efficiency than reported in GKM.[10] As for indicator (4) (the 95% coverage rate), the figures we find are closer to the nominal value for the IV estimator.

The results for the Hausman test are reported in Table 9. For the Global Hausman Test in Table 10, we find a slightly larger rejection rate (closer to the nominal level) under the null compared to the original results, but apart from this we are able to replicate Table 8 in GKM almost perfectly.

[10] We believe there is a little editorial issue in GKM: in Tables 1–6, rows 4–8 seem to be swapped with rows 9–12.

Table 3 Just-Identified Case, $s = 0$

h_1	h_2	T	Median Deviation		Abs. Med. Dev.		Decile Range		95% Coverage	
			$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$
0.4	0.4	100	0.001	−0.005	0.162	0.338	1.018	1.698	0.826	0.938
	0.4	200	−0.001	−0.005	0.134	0.292	1.034	1.596	0.848	0.944
	0.4	400	0.002	0.001	0.114	0.255	1.021	1.520	0.863	0.949
	0.4	1000	−0.000	−0.002	0.091	0.219	1.018	1.451	0.882	0.954
0.4	0.5	100	−0.006	−0.014	0.163	0.403	1.034	2.377	0.825	0.940
	0.5	200	−0.002	0.001	0.136	0.343	1.032	2.114	0.847	0.944
	0.5	400	−0.001	−0.009	0.112	0.298	1.039	1.976	0.868	0.949
	0.5	1000	−0.001	−0.003	0.090	0.247	1.020	1.785	0.887	0.949
0.5	0.4	100	−0.003	−0.007	0.157	0.289	0.879	1.208	0.775	0.912
	0.4	200	−0.003	−0.007	0.129	0.240	0.921	1.220	0.790	0.922
	0.4	400	−0.000	−0.000	0.106	0.201	1.002	1.243	0.797	0.920
	0.4	1000	0.000	0.003	0.083	0.168	0.949	1.152	0.803	0.918
0.5	0.5	100	−0.003	−0.006	0.158	0.321	0.877	1.415	0.778	0.924
	0.5	200	0.002	0.003	0.130	0.267	0.930	1.375	0.787	0.925
	0.5	400	0.001	−0.002	0.107	0.224	0.974	1.337	0.793	0.922
	0.5	1000	−0.000	−0.001	0.082	0.186	0.958	1.285	0.808	0.928

Table 4 Just-Identified Case, $s = 0.2$

h_1	h_2	T	Median Deviation		Abs. Med. Dev.		Decile Range		95% Coverage	
			$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$
0.4	0.4	100	0.042	0.021	0.161	0.297	1.019	1.564	0.805	0.921
	0.4	200	0.043	0.011	0.134	0.259	1.027	1.503	0.833	0.930
	0.4	400	0.041	0.004	0.114	0.228	1.019	1.421	0.848	0.939
	0.4	1000	0.042	0.006	0.094	0.197	1.032	1.380	0.862	0.945
0.4	0.5	100	0.041	−0.009	0.162	0.349	1.028	2.089	0.808	0.920
	0.5	200	0.043	−0.000	0.135	0.296	1.025	1.929	0.829	0.930
	0.5	400	0.040	0.002	0.112	0.260	1.010	1.784	0.851	0.936
	0.5	1000	0.042	0.005	0.092	0.215	1.017	1.614	0.862	0.938
0.5	0.4	100	0.043	0.005	0.160	0.258	0.879	1.109	0.752	0.893
	0.4	200	0.041	0.015	0.130	0.216	0.924	1.155	0.767	0.903
	0.4	400	0.042	0.010	0.110	0.184	0.949	1.141	0.763	0.896
	0.4	1000	0.041	0.013	0.087	0.149	0.981	1.127	0.762	0.899
0.5	0.5	100	0.039	0.007	0.157	0.289	0.877	1.342	0.760	0.910
	0.5	200	0.043	0.002	0.129	0.237	0.923	1.282	0.772	0.909
	0.5	400	0.042	0.005	0.109	0.201	0.967	1.286	0.760	0.901
	0.5	1000	0.042	0.004	0.087	0.165	0.959	1.209	0.764	0.908

Table 5 Just-Identified Case, $s = 0.5$

h_1	h_2	T	Median Deviation		Abs. Med. Dev.		Decile Range		95% Coverage	
			$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$
0.4	0.4	100	0.332	0.081	0.337	0.282	1.017	1.414	0.419	0.896
	0.4	200	0.330	0.070	0.332	0.243	1.039	1.413	0.346	0.904
	0.4	400	0.330	0.051	0.330	0.211	1.022	1.357	0.271	0.913
	0.4	1000	0.326	0.042	0.327	0.173	1.074	1.323	0.178	0.921
0.4	0.5	100	0.335	0.056	0.341	0.330	0.987	1.976	0.416	0.901
	0.5	200	0.327	0.046	0.329	0.270	1.028	1.761	0.349	0.906
	0.5	400	0.330	0.032	0.330	0.231	1.042	1.646	0.267	0.916
	0.5	1000	0.330	0.023	0.330	0.189	1.052	1.508	0.176	0.923
0.5	0.4	100	0.321	0.078	0.329	0.248	0.869	1.067	0.359	0.861
	0.4	200	0.329	0.066	0.332	0.211	0.948	1.110	0.263	0.866
	0.4	400	0.334	0.061	0.335	0.180	0.968	1.116	0.175	0.865
	0.4	1000	0.326	0.044	0.327	0.140	1.011	1.103	0.109	0.863
0.5	0.5	100	0.329	0.060	0.336	0.271	0.857	1.194	0.355	0.879
	0.5	200	0.331	0.044	0.333	0.227	0.955	1.255	0.255	0.884
	0.5	400	0.327	0.034	0.328	0.188	0.991	1.210	0.183	0.878
	0.5	1000	0.330	0.023	0.330	0.144	1.017	1.165	0.103	0.878

Table 6 Over-Identified Case, $s = 0$

h_1	h_2	T	Median Deviation		Abs. Med. Dev.		Decile Range		95% Coverage	
			$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$
0.4	0.4	100	0.003	0.003	0.149	0.219	1.001	1.204	0.801	0.889
	0.4	200	−0.002	−0.002	0.126	0.189	1.025	1.209	0.829	0.909
	0.4	400	−0.002	−0.004	0.105	0.167	1.023	1.178	0.853	0.919
	0.4	1000	0.001	−0.000	0.083	0.135	1.028	1.164	0.872	0.925
0.4	0.5	100	−0.005	−0.005	0.151	0.236	0.993	1.332	0.805	0.896
	0.5	200	−0.001	−0.001	0.125	0.202	1.013	1.270	0.829	0.908
	0.5	400	−0.002	−0.005	0.104	0.179	1.039	1.293	0.857	0.921
	0.5	1000	−0.001	−0.002	0.084	0.145	1.011	1.200	0.869	0.920
0.5	0.4	100	0.000	0.002	0.148	0.201	0.839	0.951	0.754	0.869
	0.4	200	0.002	0.004	0.122	0.169	0.903	1.009	0.761	0.868
	0.4	400	0.001	0.000	0.101	0.143	0.954	1.035	0.765	0.867
	0.4	1000	−0.000	−0.001	0.079	0.112	0.974	1.038	0.775	0.874
0.5	0.5	100	0.004	0.002	0.150	0.216	0.852	1.011	0.749	0.868
	0.5	200	−0.001	0.000	0.122	0.179	0.935	1.072	0.759	0.874
	0.5	400	0.003	0.000	0.102	0.150	0.958	1.073	0.765	0.877
	0.5	1000	−0.001	−0.001	0.079	0.117	0.970	1.044	0.772	0.879

Table 7 Over-Identified Case, $s = 0.2$

h_1	h_2	T	Median Deviation		Abs. Med. Dev.		Decile Range		95% Coverage	
			$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$
0.4	0.4	100	0.030	0.012	0.148	0.194	1.015	1.156	0.782	0.876
	0.4	200	0.034	0.011	0.121	0.167	1.009	1.145	0.812	0.891
	0.4	400	0.030	0.004	0.101	0.143	1.016	1.130	0.829	0.898
	0.4	1000	0.032	0.005	0.082	0.117	1.003	1.096	0.848	0.910
0.4	0.5	100	0.029	0.005	0.144	0.212	0.978	1.236	0.791	0.879
	0.5	200	0.030	0.007	0.122	0.179	0.997	1.226	0.811	0.891
	0.5	400	0.032	0.004	0.101	0.153	0.996	1.190	0.831	0.900
	0.5	1000	0.033	0.004	0.082	0.124	1.029	1.180	0.849	0.908
0.5	0.4	100	0.033	0.013	0.148	0.187	0.835	0.921	0.723	0.833
	0.4	200	0.029	0.001	0.120	0.154	0.914	0.985	0.739	0.846
	0.4	400	0.032	0.008	0.102	0.128	0.949	0.990	0.729	0.836
	0.4	1000	0.032	0.004	0.081	0.104	0.970	1.013	0.728	0.839
0.5	0.5	100	0.030	0.001	0.147	0.194	0.866	0.982	0.724	0.843
	0.5	200	0.035	0.012	0.123	0.162	0.915	1.018	0.730	0.847
	0.5	400	0.032	0.006	0.102	0.136	0.958	1.035	0.731	0.844
	0.5	1000	0.032	0.005	0.080	0.105	0.972	1.029	0.727	0.841

Table 8 Over-Identified Case, $s = 0.5$

h_1	h_2	T	Median Deviation		Abs. Med. Dev.		Decile Range		95% Coverage	
			$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$	$\tilde{\beta}_t$	$\tilde{\beta}_{1,t}$
0.4	0.4	100	0.231	0.061	0.244	0.188	0.996	1.101	0.513	0.834
	0.4	200	0.240	0.049	0.247	0.160	1.043	1.139	0.440	0.852
	0.4	400	0.237	0.039	0.239	0.136	1.045	1.105	0.381	0.868
	0.4	1000	0.237	0.029	0.238	0.109	1.052	1.107	0.282	0.883
0.4	0.5	100	0.237	0.045	0.250	0.198	0.984	1.156	0.506	0.845
	0.5	200	0.236	0.030	0.242	0.166	1.026	1.184	0.445	0.856
	0.5	400	0.241	0.020	0.243	0.138	1.030	1.153	0.373	0.877
	0.5	1000	0.240	0.014	0.241	0.112	1.035	1.125	0.278	0.888
0.5	0.4	100	0.232	0.063	0.249	0.184	0.904	0.940	0.443	0.793
	0.4	200	0.233	0.048	0.241	0.150	0.919	0.950	0.361	0.790
	0.4	400	0.241	0.044	0.244	0.126	0.979	0.998	0.275	0.794
	0.4	1000	0.239	0.031	0.241	0.100	0.986	0.995	0.186	0.792
0.5	0.5	100	0.230	0.047	0.246	0.185	0.863	0.917	0.446	0.803
	0.5	200	0.231	0.030	0.238	0.152	0.916	0.959	0.368	0.801
	0.5	400	0.237	0.022	0.241	0.128	0.969	0.997	0.282	0.805
	0.5	1000	0.240	0.017	0.241	0.099	0.998	0.999	0.183	0.807

Table 9 Table 7, Hausman test rejection rate at $t = T/2$

s	h_1	h_2	$T = 100$	$T = 200$	$T = 400$	$T = 1000$
0	0.4	0.4	0.013	0.024	0.015	0.019
	0.4	0.5	0.017	0.011	0.027	0.028
	0.5	0.4	0.020	0.023	0.033	0.027
	0.5	0.5	0.015	0.017	0.028	0.022
0.2	0.4	0.4	0.022	0.020	0.022	0.024
	0.4	0.5	0.021	0.023	0.031	0.029
	0.5	0.4	0.033	0.034	0.042	0.048
	0.5	0.5	0.026	0.027	0.041	0.056
0.5	0.4	0.4	0.174	0.253	0.377	0.468
	0.4	0.5	0.190	0.250	0.359	0.479
	0.5	0.4	0.282	0.422	0.529	0.652
	0.5	0.5	0.303	0.406	0.512	0.665

Table 10 Global Hausman test rejection rate, $T_0 = 5$ and $T_1 = T - 5$

s	h_1	h_2	$T = 100$	$T = 200$	$T = 400$	$T = 1000$
0	0.4	0.4	0.070	0.070	0.093	0.085
	0.4	0.5	0.062	0.071	0.074	0.100
	0.5	0.4	0.075	0.087	0.107	0.121
	0.5	0.5	0.053	0.081	0.110	0.102
0.2	0.4	0.4	0.091	0.105	0.121	0.245
	0.4	0.5	0.073	0.093	0.129	0.212
	0.5	0.4	0.090	0.121	0.159	0.258
	0.5	0.5	0.078	0.120	0.164	0.230
0.5	0.4	0.4	0.687	0.850	0.949	0.997
	0.4	0.5	0.626	0.792	0.936	0.996
	0.5	0.4	0.677	0.873	0.963	0.998
	0.5	0.5	0.668	0.835	0.952	0.997

Appendix D: Phillips curves: exact replication

Figure 8 reports the parameters estimates of the "traditional" Phillips curve obtained setting $h_1 = h_2 = 0.5$.

Appendix E: Phillips curve: extended replication

Figure 9 shows estimated coefficients and confidence intervals for both OLS and IV estimators, setting $h_1 = h_2 = 0.5$. See Sect. 4.1.

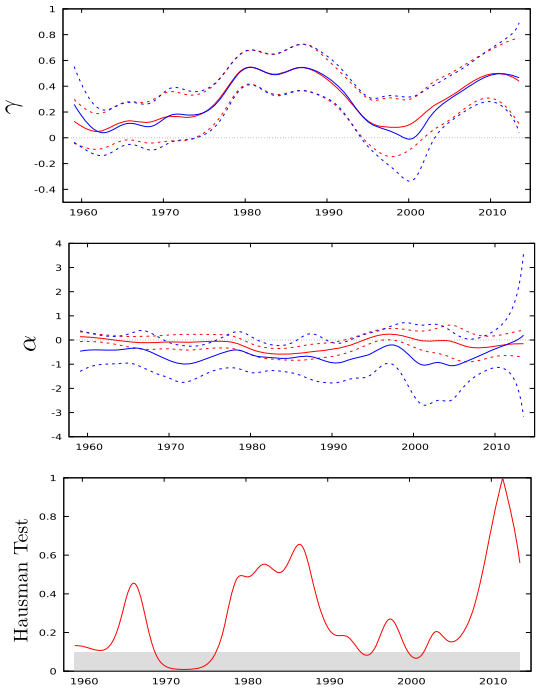

Fig. 8 PC Replication, $h_1 = h_2 = 0.5$. Note: see Fig. 3

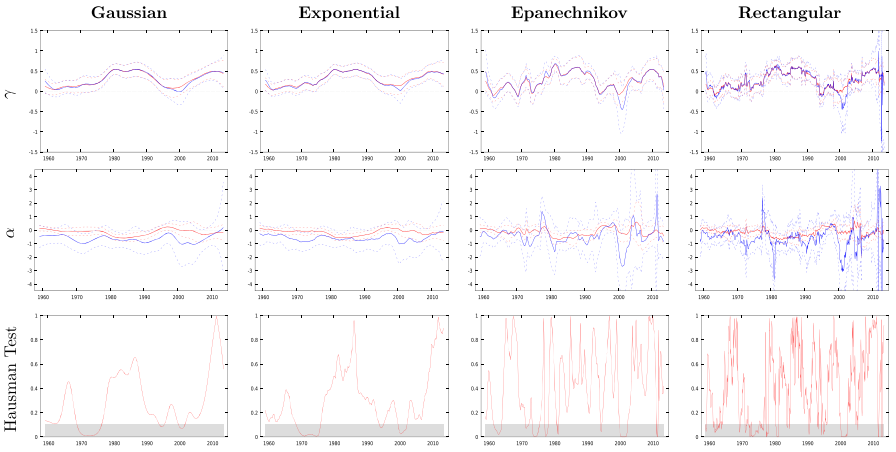

Fig. 9 Traditional PC: 1959:02–2013:07, $h_1 = h_2 = 0.5$. Note: see Fig. 3

Appendix F: The Kalman-filter apparatus: a comparison

In this appendix, we provide a comparison between the time-varying OLS (TV-OLS) estimator discussed in GKM and an estimator obtained by the Kalman-filter apparatus (KF hereafter). To this aim, we compare TV-OLS and KF estimates on simulated data and on the "traditional" Phillips curve proposed in Sect. 3.2.

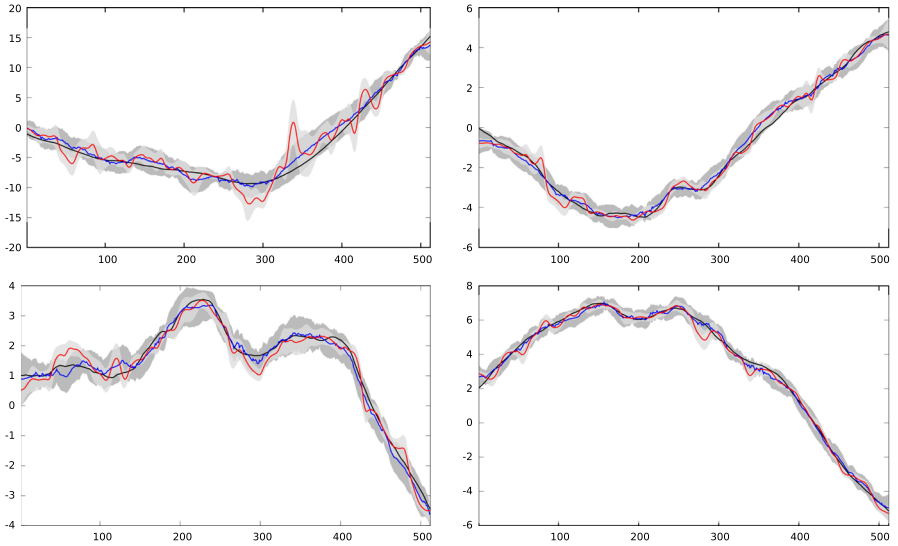

Fig. 10 Comparing TV-OLS and KF: simulated data. Note: results for $\beta_{1,t}$ and $\beta_{2,t}$ (above); $\beta_{3,t}$ and $\beta_{4,t}$ (below). Each panel reports the true value of the parameter (black line), the estimates from TV-OLS (red line) and KF (blue line) and their 95% confidence intervals (light gray and dark gray shaded areas, respectively). TV-OLS bandwidth parameter $h \approx 0.30$

In order to obtain the KF estimator, we consider, for $t = 1, \ldots, T$, the following state-space representation of model

$$\begin{aligned} y_t &= \boldsymbol{x}_t' \boldsymbol{\beta}_t + u_t, \\ \boldsymbol{\beta}_{t+1} &= \boldsymbol{\beta}_t + \eta_t, \end{aligned} \tag{F26}$$

where the states $\boldsymbol{\beta}_t$ are the time-varying parameters, which are assumed to be independent random walk processes[11] and the rest of notation coincides with the one in Sect. 2.

In the first exercise, we generated data according to the model in Eq. (F26), where the u_t terms are random draws from a Gaussian distribution and $x_{k,t}$, with $k = 2, \ldots, 4$,[12] are generated as independent AR(1) processes with first-order autocorrelation equal to 0.95. In order to simulate time-varying coefficients with a high degree of smoothness, the $\boldsymbol{\beta}_t$ coefficients were generated as detrended cumulated random walk (that is, $I(2)$) processes. Therefore, we are using a state-space representation that is technically misspecified, given the way we generated the $\boldsymbol{\beta}_t$ processes. Estimation is carried out by maximum likelihood, with a diffuse prior for the state as in Durbin and Koopman (2012); the estimate of the $\boldsymbol{\beta}_t$ parameters can then be recovered by smoothing the state vector.

[11] This choice is not the only one present in the literature, but it is very common among practitioners. For a fuller description, see Harvey (1990), Section 7.7.1.

[12] We denote the constant term as $x_{1,t}$.

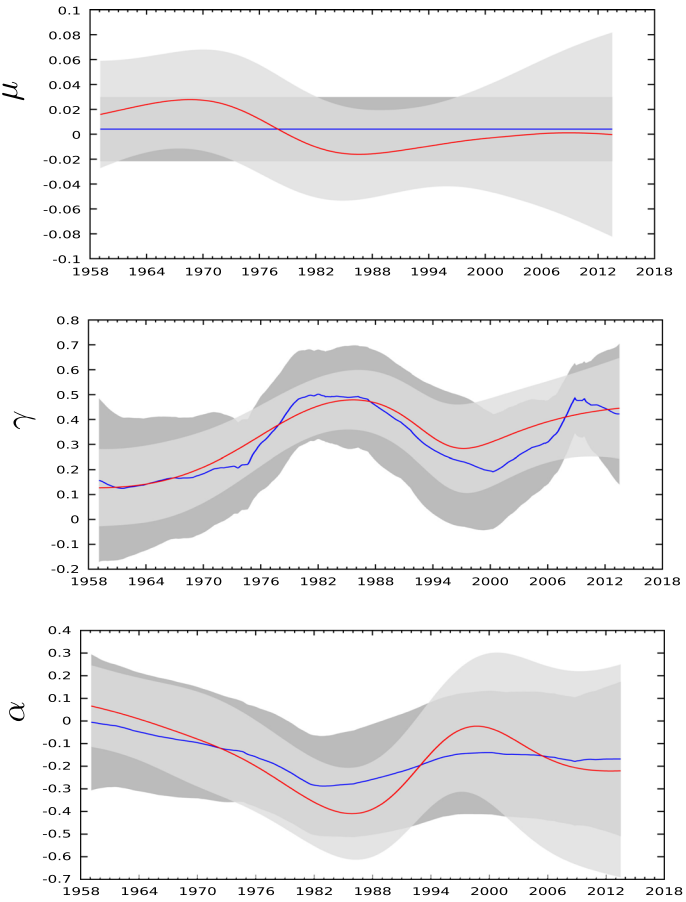

Fig. 11 Comparing TV-OLS and KF: "traditional" Phillips curve. Note: each panel reports the true value of the parameter (black line), the estimates from TV-OLS (red line) and KF (blue line) and their 95% confidence intervals (light gray and dark gray shaded areas, respectively). TV-OLS bandwidth parameter $h \approx 0.65$

Figure 10 reports the simulated $\boldsymbol{\beta}_t$ parameters and their estimates via TV-OLS and KF. Here, for the TV-OLS we choose the automatic bandwidth selection outlined in Sect. 4.2. KF and TV-OLS estimates appear to be generally similar with the latter exhibiting a path slightly more volatile then the KF one. Moreover, it must be remarked that TV-OLS is several order of magnitudes less demanding than KF in terms of computational complexity: KF estimates entail a full Maximum-Likelihood estimation routine, which could take a long CPU time to converge, if at all. TV-OLS, on the other hand, requires no numerical optimization.

Finally, Fig. 11 shows the estimated coefficients of the "traditional" Phillips curve in Eq. (2). The results of the KF and TV-OLS estimators are very much alike, as they both provide evidence of very little time-variation of the parameter μ_t, while the time paths of γ_t and α_t are very similar to each other.

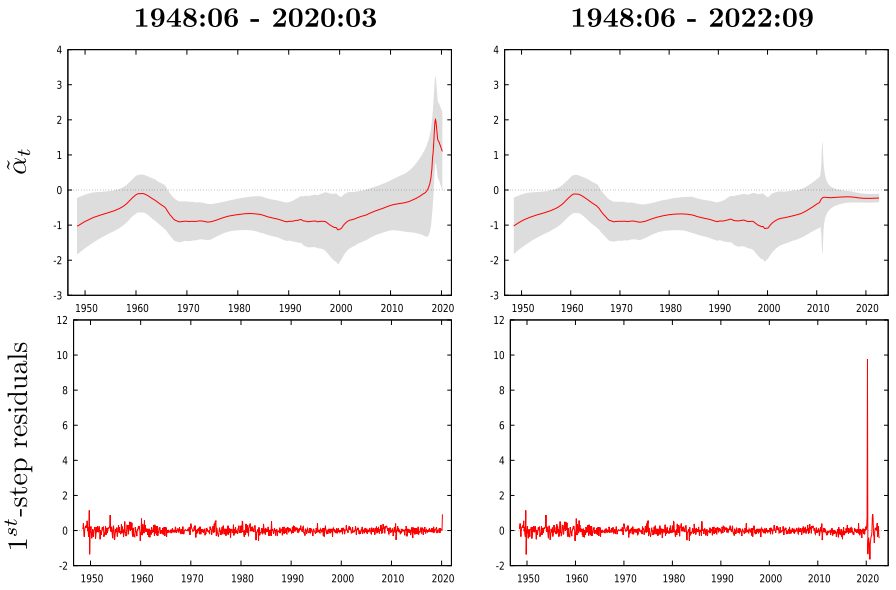

Fig. 12 Excluding Covid-19: Epanechnikov kernel. Note: the two columns report IV estimates and first-stage residuals of the endogenous variable Δu_t, obtained by excluding (left) the Covid-19 period and with the full sample (right). Epanechnikov kernel function, bandwidth parameters $h_1 = h_2 = 0.7$. Shaded area denotes 90% confidence interval

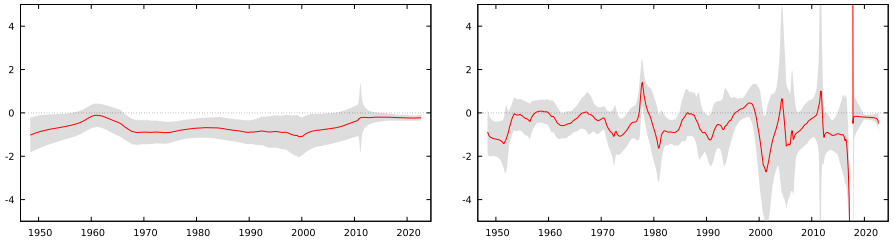

Fig. 13 Bandwidth: local effects on standard errors. Note: the two panels report estimated $\tilde{\alpha}_t$ and 90% confidence intervals for the model in Eq. (2), Epanechnikov kernel. The left panel is obtained with $h_1 = h_2 = 0.7$ and right one with $h_1 = h_2 = 0.5$. Shaded area denotes 90% confidence interval

Appendix G: Additional figures

We report an alternative version of Figs. 6 and 7 where estimation is performed by relying on the Epanechnikov kernel function rather than the Gaussian kernel (see Figs. 12, 13).

References

Bai J, Perron P (2003) Critical values for multiple structural change tests. Econom J 6(1):72–78
Cai Z (2007) Trending time-varying coefficient time series models with serially correlated errors. J Econom 136(1):163–188
Cai Z, Tiwari RC (2000) Application of a local linear autoregressive model to BOD time series. Environmetrics 11(3):341–350
Chow GC (1960) Tests of equality between sets of coefficients in two linear regressions. Econometrica 28:591–605
Durbin J, Koopman SJ (2012) Time series analysis by state space methods, 2nd edn. Oxford University Press, Oxford
Giraitis L, Kapetanios G, Yates T (2014) Inference on stochastic timevarying coefficient models. J Econom 179(1):46–65. https://doi.org/10.1016/j.jeconom.2013.10.009
Giraitis L, Kapetanios G, Yates T (2018) Inference on multivariate heteroscedastic time varying random coefficient models. J Time Ser Anal 39(2):129–149. https://doi.org/10.1111/jtsa.12271
Giraitis L, Kapetanios G, Marcellino M (2021) Time-varying instrumental variable estimation. J Econom 224(2):394–415
Hamilton J (1994) Time series econometrics. Princeton University Press, Princeton
Harvey AC (1990) Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge
Hastie T, Tibshirani R (1993) Varying-coefficient models. J R Stat Soc: Ser B (Methodol) 55(4):757–779
Kalaba R, Tesfatsion L (1989) Time-varying linear regression via flexible least squares. Comput Math Appl 17(8–9):1215–1245
Kapetanios G, Marcellino M, Venditti F (2019) Large time-varying parameter VARs: a nonparametric approach. J Appl Econom 34(7):1027–1049
Ruisi G (2019) Time-varying local projections (Working Papers No. 891). Queen Mary University of London, School of Economics and Finance
Schlicht E (2021) Vc: a method for estimating time-varying coefficients in linear models. J Korean Stat Soc 50:1–33
Vogt M (2012) Nonparametric regression for locally stationary time series. Ann Stat 40(5):2601–2633