Testing For Exogeneity In Time-Varying Instrumental Variable Regressions: A Bootstrap Approach

Charisios Grivas[†]

February 14, 2022

Abstract

This paper considers bootstrap versions of a Hausman-type test statistic for econometric models with endogenous regressors whose coefficients are allowed to vary over time both deterministically or stochastically. I compare the finite sample performance of the asymptotic and the bootstrap version of the test by means of Monte Carlo simulations. The bootstrap test statistic appears to have proper size and higher power. More importantly, it is shown that the size and the power of the bootstrap test, are invariant to the choice of the bandwidth parameters and the number of instruments.

Keywords: Instrumental variables, Time-varying parameters, Endogeneity, Hausman test, Non-parametric methods, Bootstrap

1. Introduction

Instrumental variable (IV) regressions are fundamental in applied economic research. By construction, IV methods exploit the variation in the endogenous variables caused by shifts in the instrumental variable. Should one use these estimators to alleviate bias in case of endogeneity or use standard estimators that, although biased under endogeneity, they are more efficient if there is no endogeneity? Clearly, a Hausman (1978) exogeneity test which compares a standard estimator, like the LS, with those that account for endogeneity can address the issue.

A usual assumption often made in the IV literature is that both the entertained model and the parameters remain constant over time. In turn, this implies that the relationship between the endogenous variables and the instruments remains unchanged as well. This assumption,

[†]PhD candidate at Birkbeck College.All correspondence should be addressed to:Charisios.grivas@gmail.com.

although crucial, still remains highly susceptible to criticism. Zhang et al., 2008 argue that the conflicting conclusions about the importance of key variables in the determination of inflation based on a New Keynesian Phillips Curve(NKPC) may be due to neglected parameter variation.

A recent strand of literature also considers the possibility of a time-varying endogeneity status of a variable. Giraitis et al. (2020)(GKM), building upon the seminal works by Chen (2015) and Giraitis et al. (2014)(GKY), proposed a non-parametric, kernel-based, estimation and inferential theory for time-varying IV regression allowing for both deterministic and random coefficients. In their article, GKM introduced time-varying types of Hausman exogeneity tests allowing for possible changes in the endogeneity status of the variables. Although the finite sample bias of the estimators is found to be small, the time-varying Hausman tests appear to experience size distortions and have low power. More importantly, these size distortions are quite sensitive to the choice of the bandwidth parameter and the number of instruments.

In this paper, I consider a bootstrap version of the time-varying Hausman exogeneity test, which compares time-varying OLS and IV estimators. The ability of bootstrapping procedures to approximate the small sample distribution is well known in the literature(see for instance Beran and Ducharme (1991)) and Hall (1992)). In a related study, fu Wong (1996) finds that the bootstrap-based Hausman test can provide substantial refinements over the asymptotic test. As shown in the Monte Carlo simulations section, bootstrap appears to deal with the size distortions in an effective manner and most importantly, render test statistic insensitive to the choices of the bandwidth parameter and the number of instruments. To the best of my knowledge, this is the first paper to use bootstrap in time-varying instrumental variable estimation to address the issue of bandwidth selection for hypothesis testing.

Finally, I employ a time-varying IV estimation of a Phillips curve for the USA to illustrate both the qualitative and quantitative differences of the two versions of the test in an empirical application. Considerable attention has been focused on studying the NKPC as it is used to identify the forward-looking components of inflation and the trade-off between inflation and unemployment over the cycle. Following GKM, unemployment is used as a forcing variable for inflation. The time-varying bootstrap version of the Hausman test seems to be less sensitive to the choice of the bandwidth parameter than the asymptotic test, while also suggesting time endogeneity of unemployment for a longer period of time around 2000 when compared to the asymptotic test.

The paper is organized as follows: In Section 2, I briefly present an overview of the problem addressed by GKM and review the time-varying Hausman test proposed. In Section 3, I discuss the bootstrap approach in this context. Section 4 examines the behaviour of the test statistic by means of Monte Carlo simulations. Section 5 presents the empirical application. Finally, some concluding remarks and some directions on further research are given in Section 6.

2. Theory

Giraitis et al. (2014) introduced a non-parametric time-varying OLS estimation method that is based on a kernel generalisation of a rolling window. (GKM) expanded the results in the IV context with either deterministic or random coefficients, and derived a time-varying version of the Hausman exogeneity test comparing the time-varying OLS and IV estimators, allowing for a shift in endogeneity status over time.

To fix ideas, I consider the following regression model for a univariate series, y_t:

$$y_t = x_t^{'}\beta_t + u_t \tag{1}$$

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

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

Assume that endogenous variables x_t are correlated with u_t but there exist some exogenous instruments z_t such that:

$$E[z_t u_t] = 0, \quad E[z_t v_t^{'}] = 0, \quad t \geq 1 \tag{3}$$

GKM introduced a kernel type estimator for β_t

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

where $b_{H,|j-t|} = K(\frac{|j-t|}{H})$ are the kernel weights with bandwidth parameter H and $\hat{\Psi}_j$ is the kernel OLS estimator

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

which is a consistent estimate of Ψ_j.

Note the different bandwidth parameters L and H used in (5) and (4) respectively. The kernel weights are of the form:

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

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

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

I also consider the OLS estimator

$$\hat{\beta}_t = (\sum_{j=1}^{T} b_{H,|j-t|} x_j x_j^{'})^{-1} (\sum_{j=1}^{T} b_{H,|j-t|} x_j y_j) \tag{8}$$

under the assumption of exogenous regressors.

Under assumptions (1)-(5) in Giraitis et al. (2020), GKM proposed a time-varying version of the Hausman exogeneity test, which compares the time-varying IV and OLS estimators defined above, allowing for changes in the endogeneity status over time. The test statistic takes the form of

$$\frac{K_t^2}{K_{2,t}} V_{T,t}^{'} \hat{\Sigma}_{\hat{v}\hat{v},t}^{-1} V_{T,t} \hat{\sigma}_{\hat{u},t}^{-2} \tag{9}$$

where $\hat{\Sigma}_{\hat{v}\hat{v},t} := K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \hat{v}_j \hat{v}_j^{'}$, $\hat{\sigma}_{\hat{u},t}^2 := K_t^{-1} \sum_{j=1}^{T} b_{H,|j-t|} \hat{u}_j^2$ based on residuals $\hat{u}_j = y_j - x_j^{'} \tilde{\beta}_{1,j}$ and $\hat{v}_j = x_j - \hat{\Psi}_j^{'} z_j$ which can be used to test the null hypothesis $H_0 : E[v_t u_t] = 0$ that x_j is exogenous at time t. Under the null, the statistic is asymptotically distributed as χ_p^2.

As shown in GKM, size distortions of the test statistic in (9) can be substantial even for large sample sizes. These size distortions seem to be quite sensitive to the choice of the bandwidth parameters h_1 and h_2 and also on the number of instruments. In the next section, I present an overview of the bootstrap approaches used and discuss a number of issues regarding their implementation.

3. The Bootstrap

Originally proposed by Efron (1979), bootstrap constitutes a major tool in the hands of statisticians for approximating the sampling distribution and variance of complicated statistics. Also, as pointed out by Politis (2003), Beran and Ducharme (1991) among many, bootstrap tests can often provide significant refinements to asymptotic tests. However, bootstrap involves a number of choices to be made ex ante by the researcher.

The first issue that needs to be addressed is the resampling of x and z. Here, I do not make any assumption regarding the parametric model for x and z and hence I use nonparametric bootstrap. Depending on the presence of dependence between these variables one might want to resample blocks of rows of data from (x_t, z_t) using block bootstrap[1] as proposed by Kunsch (1989). Resampling blocks of rows instead of resampling each variable independently is performed so as to retain the dependency between these variables. A well known difficulty of the block bootstrap is that the block size b needs to be determined in advance by the practitioner.

The choice of the block size is critical for the performance of the bootstrap. If the block size chosen is too small then the dependency among the blocks is broken and hence it cannot be expected that the bootstrap samples will mimic closely the original data. If, on the other hand, the block size is chosen to be too large, the bootstrap samples are no longer random enough. However, only in few cases literature provides guidance regarding the selection of the block size (c.f Bühlmann and Künsch (1999), Lahiri (1999) and Hall et al. (1995) just to name a few). In most of the cases, simple rules specifying the rate at which b should increase with sample size are provided for specific applications. Usually those include $b = O(T^{1/3})$ or $b = O(T^{1/2})$. In my case, there is no simple rule guiding my choice of the block size and so different values for b are examined.

Another important issue, in testing for exogeneity, is the mode of resampling so that the null distribution is imposed, as discussed in Kapetanios (2010). This is especially important if, for example, the whole distribution is bootstrapped and used instead of the asymptotic χ^2 approximation. In particular, not imposing the null can lead to an inconsistent test since, in the case of the Hausman test, endogeneity prevails in the bootstrapped samples and hence the bootstrap test statistics will be unbounded asymptotically.

Then, the practitioner needs to resample y_t using parametric bootstrap. This, of course requires to choose whether to use OLS estimates or 2SLS estimates. In my paper, I use[2] OLS estimates as suggested by Kapetanios (2010) and resample the OLS residuals $\hat{u}_t$. For each bootstrap sample I must impose the null $E^\star(x_t^\star u_t^\star) = 0$ where $E^\star$ is conditional on

[1] Alternatively, one could use variants of the MBB as in Politis and Romano, 1991 who propose the Circular Bootstrap to deal with the end effects. This path was also explored. See the robustness analysis in Appendix A.

[2] A preliminary analysis suggested that similar results are obtained using either OLS or IV estimator. Results are available upon request from the author.

the realized sample. I suggest two alternatives for that. The first alternative is the wild bootstrap as proposed by Wu (1986)(see also Goncalves and Kilian (2004) for an application in time series). The wild bootstrap uses a transformation of the residuals to construct the bootstrap error term $u_t^\star = \eta_t \hat{u}_t$ where η_t is a random variable with mean 0 and variance 1. Different distributions have been proposed for η_t including those of Rademacher as suggested by Davidson and Flachaire (2008) and Mammen as proposed by Mammen (1993). Using wild bootstrap guarantees that the null is imposed as I now have $E^\star(x_t^\star u_t^\star) = E^\star(x_t^\star \hat{u}_t \eta_t) = 0$. The Wild bootstrap algorithm used, stems from Kapetanios (2010) and the algorithm is the following:

Step 1. Estimate (1) by OLS and obtain the residuals and the estimates of $\hat{u}$ and $\hat{\beta}_t$ respectively.

Step 2. Resample blocks of rows of (z_t, x_t) using a block bootstrap approach to produce bootstrap samples of size T, for the regressors, as suggested by Kunsch (1989).

Step 3. The wild bootstrap is applied to OLS residuals $\hat{u}_t$ by premultiplying them with η_t so that I have $u^\star = \eta_t \hat{u}_t$,where $\{\eta\}_{t=1}^T$ is an i.i.d standard normal distribution, independent of all other random processes. A bootstrap sample of $\{u^\star\}_{t=1}^T$ is obtained.

Step 4. Use $\hat{\beta}$,$(z_t^\star, x_t^\star)$ and $\{u^\star\}_{t=1}^T$ in (1) to obtain bootstrap samples of size T for y_t and denote them $y_t^\star$.

Step 5. Use $(z_t^\star, x_t^\star, y_t^\star)$ to obtain OLS and IV estimates given by (8) and (4) respectively, and calculate the Hausman test for the bootstrap sample.

Step 6. Repeat 2-5 B times to obtain B bootstrap tests.

The second alternative is to resample OLS residual itself. By virtue of OLS, the residuals are guaranteed to be orthogonal to the covariates even in case of endogeneity. Once the bootstrap error term has been generated, (1) can be used to construct $y_t^\star$. The Residual Bootstrap[3] is similar to that of fu Wong (1996) and the algorithm employed is the following:

Step 1. Estimate (1) using OLS. Coefficients $\hat{\beta}_t$ and residuals $\hat{u}_t = y_t - \hat{\beta}_t x_t$ are obtained.

Step 2. Recenter the residuals by $\hat{u}_{new} = \hat{u}_t - \bar{\hat{u}}$ where $\bar{\hat{u}} = \frac{1}{T}\sum_{t=1}^T \hat{u}_t$ and obtain

$$y_t^\star = \hat{\beta}_t x_t + \hat{u}_{new}^\star$$

[3]Henceforth, RB.

by resampling with replacement from $\hat{u}_{new}$ to produce a bootstrap sample of size T, for the dependent variable $y_t^\star$. By resampling in this way, any links between u_t and (z_t, x_t) are broken while the relationship between z_t and x_t is maintained. Note that the mean adjustment is necessary, since there is no guarantee that the mean of the bootstrap samples is zero.

Step 3. The bootstrap sample is then used to estimate the model by OLS and 2SLS and produce a bootstrap test statistic for the exogeneity test.

Step 4. Repeat steps 2-3 B times to produce B bootstrap test statistics.

Finally, an issue that needs to be addressed, is the number of bootstrap replications. Hall (1992) proposed to choose B, the number of bootstrap repetitions such that $\nu/(B+1) = 1-\alpha$ for a positive integer ν. This implies in turn that α is a rational such that $\alpha = \alpha_1/\alpha_2$ for positive integers α_1 and α_2 with no common integer divisors. Then $B = \alpha_2 h - 1$ and $\nu = (\alpha_2 - \alpha_1)/h$ for positive integer h[4]. As Davidson and MacKinnon (2000) showed, using $B = 399$ is about the minimum for a test that guarantees a loss of power less than 1% at 0.05 level.

4. Simulation Results

In this Section, Monte Carlo experiments are carried out to investigate the finite-sample performance of the time-varying local Hausman test proposed by GKM and the bootstrap version of it. In all simulations[5], in this section, I use 5000 Monte Carlo replications and 999 bootstrap replications. The experiments found in this section are the same as those employed in GKM.

As data generating process (DGP) under the exactly identified case, I consider the following model:

$$y_t = \beta_t x_t + u_t \tag{10}$$

$$x_t = \psi_t z_t + v_t \tag{11}$$

for $t = 1, ..., T$. Following GKM, correlation between u_t and v_t is introduced by specifying

[4]For instance, for $\alpha = 0.05$ I have $\alpha_1 = 1$, $\alpha_2 = 20$,$B = 20h - 1$ and $\nu = 19h$ for a positive integer h. So $B = 19, 39, 59, ...$ etc.

[5]All simulations were performed using an Apple M1 with a 8 GB unified memory. The code was written and executed in JuliaPro-1.5.3

them as

$$u_t = se_{1,t} + (1-s)e_{2,t} \quad v_t = se_{1,t} + (1-s)e_{3,t} \tag{12}$$

where $s = 0, 0.2, 0.5, 0.8, 0.9$ and $\{e_{1,t}\}$,$\{e_{2,t}\}$ and $\{e_{3,t}\}$ are mutually independent $NIID(0,1)$ sequences.

The parameters $\beta_t = T^{-1/2}\xi_{1,t}$, $\psi_t = T^{-1/2}\xi_{2,t}$,$t = 1, ..., T$ are generated as two independent rescaled random walks, such that $\xi_{l,t} - \xi_{l,t-1} \sim N(0,1)$ for $l = 1,2$ that are also independent of $\{\psi_t\}$,$\{u_t\}$ and $\{v_t\}$. This implies that both the structural and the reduced form regressions have time-varying coefficients. Exogeneity of x_t is implied by $s = 0$, while for $s = 0.2, 0.5, 0.8, 0.9$, x_t is endogenous. The magnitude of s hence, provides a means for controlling the extent of endogeneity.

I cxamine two estimators of β_t using the notation of GKM: the time-varying $\hat{\beta}_t$(OLS) and the time-varying $\hat{\beta}_{1,t}$(IV). These are computed using the Gaussian kernel[6] $K(x) = exp(-x^2/2)$ with a variety of bandwidth values H for estimation of β_t and L for ψ_t. Specifically, I set $H = T^{h_1}$ and $L = T^{h_2}$ with $h_1, h_2 = 0.4$ and 0.5 as in GKM[7]. Results for values of 0.7 are also reported in the Appendix A. Lower values for the bandwidth increase robustness of estimates to parameter changes but decrease efficiency. Further, I consider three sample sizes of length $100 + T$ with $T \in \{100, 200, 400\}$. The first 100 observations are then discarded in order to eliminate initial value effects and only the remaining T observations are used. I now discuss the bootstrap implementation that is employed in the rest of the section.

I consider two different bootstrap procedures. In the first one, which I call RB as an abbreviation for bootstrap based on residuals, I use a standard model bootstrap together with the OLS estimates where residuals are resampled with replacement in a similar manner as in fu Wong (1996) to obtain $y_t^\star$. No resampling of (x_t, z_t) is taking place in this procedure. The second bootstrap method uses block resampling of rows of (x_t, z_t) where I examine multiple values for the block size b, that are proportional or multiples to the sample size T. This method is combined with wild bootstrap to obtain bootstrap replications of $y_t^\star$ under the null. In this case, I construct bootstrap error term $u_t^\star = \eta_t \hat{u}_t$ where η_t is an i.i.d standard Normal[8] distributed sequence as suggested in Kapetanios (2010).

[6] Similar results were obtained using both the Epanechnikov kernel $K(x) = 0.75(1 - x^2)\ for\ |x| < 1$ and the exponential kernel $K(x) = exp(-cx^\alpha)$ where $c > 0$ and $\alpha > 0$.

[7] Other combinations of values for the bandwidth parameters produced similar evidence and are available upon request from the author.

[8] The Rademacher latice distribution was also used as an alternative to the standard Normal but provided similar results.

To evaluate the performance of the test statistic, I examine the rejection frequencies of the local time varying Hausman test at 5% significance level and for $t = T/2$. These frequencies are reported in Table 1-Table 3.

Table 1 shows the rejection frequencies for the local Hausman test for $T = 100$. The asymptotic test appears to exhibit considerable size distortions for all bandwidth parameters and also low power. On the other hand, bootstrap procedures seem to outperform the asymptotic test both in terms of size and power irrespectively of the bandwidth parameter chosen. The size is close to the nominal 5% while the power is at least 38.5% larger than the power of the asymptotic test statistic.

Table 1: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$. Model (10)-(12).

			T=100					
s	h_1	h_2	***Asymptotic***	***RB***	$b = 2$	$b = 4$	$b = 6$	$b = 8$
0	0.4	0.4	0.024	0.047	0.060	0.057	0.054	0.054
	0.4	0.5	0.029	0.047	0.052	0.052	0.050	0.049
	0.5	0.4	0.034	0.048	0.056	0.057	0.055	0.053
	0.5	0.5	0.033	0.047	0.056	0.056	0.053	0.055
0.2	0.4	0.4	0.032	0.049	0.060	0.058	0.058	0.059
	0.4	0.5	0.039	0.048	0.056	0.054	0.053	0.053
	0.5	0.4	0.040	0.054	0.064	0.064	0.064	0.063
	0.5	0.5	0.042	0.055	0.060	0.059	0.058	0.059
0.5	0.4	0.4	0.305	0.362	0.376	0.373	0.368	0.369
	0.4	0.5	0.306	0.358	0.353	0.347	0.344	0.338
	0.5	0.4	0.452	0.506	0.528	0.527	0.526	0.524
	0.5	0.5	0.452	0.509	0.524	0.518	0.513	0.516
0.8	0.4	0.4	0.705	0.846	0.805	0.801	0.798	0.797
	0.4	0.5	0.687	0.814	0.771	0.764	0.763	0.761
	0.5	0.4	0.804	0.909	0.881	0.877	0.876	0.875
	0.5	0.5	0.785	0.899	0.875	0.873	0.870	0.870
0.9	0.4	0.4	0.705	0.867	0.810	0.805	0.803	0.803
	0.4	0.5	0.684	0.821	0.772	0.766	0.764	0.763
	0.5	0.4	0.807	0.919	0.883	0.880	0.879	0.877
	0.5	0.5	0.786	0.907	0.877	0.875	0.870	0.869

Table 2 presents the rejection frequencies for the local Hausman test for $T = 200$. The size of the bootstrap test is close to the nominal level irrespective of the bandwidth parameters. In terms of power, there is at least a 10.5% increase through the use of bootstrap. This is especially the case for small values of s where the power refinements through the use of bootstrap account for at least a 41% gain.

Table 2: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$. Model (10)-(12).

			T=200					
s	h_1	h_2	***Asymptotic***	***RB***	$b = 4$	$b = 6$	$b = 8$	$b = 16$
0	0.4	0.4	0.025	0.054	0.062	0.061	0.061	0.057
	0.4	0.5	0.029	0.050	0.053	0.050	0.051	0.048
	0.5	0.4	0.033	0.054	0.066	0.064	0.065	0.061
	0.5	0.5	0.031	0.053	0.062	0.063	0.063	0.060
0.2	0.4	0.4	0.032	0.053	0.069	0.066	0.066	0.063
	0.4	0.5	0.037	0.056	0.059	0.058	0.055	0.052
	0.5	0.4	0.043	0.062	0.072	0.070	0.068	0.064
	0.5	0.5	0.042	0.059	0.070	0.066	0.068	0.063
0.5	0.4	0.4	0.369	0.443	0.460	0.457	0.455	0.449
	0.4	0.5	0.368	0.435	0.424	0.419	0.415	0.410
	0.5	0.4	0.551	0.609	0.626	0.625	0.626	0.617
	0.5	0.5	0.542	0.607	0.621	0.619	0.619	0.613
0.8	0.4	0.4	0.721	0.850	0.805	0.804	0.803	0.801
	0.4	0.5	0.701	0.811	0.770	0.765	0.764	0.758
	0.5	0.4	0.823	0.921	0.887	0.883	0.883	0.880
	0.5	0.5	0.804	0.911	0.876	0.873	0.873	0.868
0.9	0.4	0.4	0.713	0.869	0.810	0.808	0.805	0.801
	0.4	0.5	0.683	0.810	0.769	0.763	0.760	0.758
	0.5	0.4	0.822	0.930	0.890	0.887	0.885	0.882
	0.5	0.5	0.792	0.912	0.875	0.872	0.869	0.867

Similar results are obtained from Table 3 for $T = 400$. The power is at least 9.8% larger than the power of the asymptotic test. Overall, residual bootstrap seems to outperform the asymptotic test statistic in terms of power while retaining size close to nominal.

The Block bootstrap underperforms both the Asymptotic and the RB test for all sample sizes and all bandwidth parameters independently of the block size chosen. Its size varies substantially from 0.048 to 0.066 along different values of block size. Likewise, its power ranges from 0.051 to 0.078 and 0.338 to 0.693 for $s = 0.2$ and $s = 0.5$ respectively.

Table 3: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$. Model (10)-(12).

			T=400					
s	h_1	h_2	***Asymptotic***	***RB***	$b = 6$	$b = 8$	$b = 16$	$b = 32$
0	0.4	0.4	0.023	0.046	0.058	0.059	0.054	0.052
	0.4	0.5	0.028	0.047	0.053	0.052	0.048	0.050
	0.5	0.4	0.028	0.050	0.062	0.064	0.060	0.055
	0.5	0.5	0.028	0.051	0.063	0.063	0.059	0.056
0.2	0.4	0.4	0.029	0.055	0.071	0.071	0.063	0.061
	0.4	0.5	0.035	0.056	0.062	0.060	0.051	0.053
	0.5	0.4	0.041	0.060	0.078	0.076	0.075	0.067
	0.5	0.5	0.041	0.062	0.078	0.078	0.073	0.069
0.5	0.4	0.4	0.433	0.500	0.522	0.521	0.515	0.509
	0.4	0.5	0.425	0.494	0.478	0.476	0.460	0.465
	0.5	0.4	0.622	0.678	0.693	0.689	0.687	0.676
	0.5	0.5	0.615	0.672	0.688	0.684	0.681	0.679
0.8	0.4	0.4	0.729	0.857	0.812	0.808	0.804	0.796
	0.4	0.5	0.701	0.815	0.771	0.766	0.762	0.760
	0.5	0.4	0.832	0.929	0.893	0.891	0.885	0.876
	0.5	0.5	0.810	0.918	0.876	0.877	0.870	0.867
0.9	0.4	0.4	0.725	0.879	0.818	0.814	0.807	0.800
	0.4	0.5	0.694	0.814	0.769	0.766	0.757	0.758
	0.5	0.4	0.835	0.941	0.896	0.896	0.885	0.878
	0.5	0.5	0.809	0.929	0.878	0.877	0.871	0.864

I also examine the overidentified case as in GKM, where I have

$$y_t = \beta_t x_t + u_t \tag{13}$$

$$x_t = \psi_{1,t} z_{1,t} + \psi_{2,t} z_{2,t} + v_t \tag{14}$$

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

Using (13)-(14) I now report the rejection frequencies in Table 4-Table 6. Table 4 reports the rejection frequencies for the local Hausman test for $T = 100$. The size of the asymptotic test is close to the nominal value alghough it varies more than the RB. Note for $h_1 = 0.5$ and $h_2 = 0.4$ the size of the asymptotic test is 0.069 while for the RB it is close to nominal. In terms of power, the asymptotic test and RB are quite close in the case of weak endogeneity while for strong endogeneity, RB has higher power.

Table 4: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$. Model (13)-(14).

			T=100					
s	h_1	h_2	***Asymptotic***	***RB***	$b=2$	$b=4$	$b=6$	$b=8$
0	0.4	0.4	0.055	0.052	0.073	0.074	0.076	0.073
	0.4	0.5	0.051	0.052	0.071	0.070	0.067	0.067
	0.5	0.4	0.069	0.056	0.085	0.085	0.084	0.084
	0.5	0.5	0.056	0.047	0.069	0.069	0.071	0.069
0.2	0.4	0.4	0.065	0.058	0.079	0.081	0.080	0.079
	0.4	0.5	0.061	0.058	0.074	0.072	0.072	0.070
	0.5	0.4	0.084	0.062	0.095	0.094	0.094	0.094
	0.5	0.5	0.069	0.059	0.075	0.077	0.077	0.076
0.5	0.4	0.4	0.379	0.391	0.410	0.410	0.408	0.410
	0.4	0.5	0.386	0.414	0.399	0.398	0.395	0.395
	0.5	0.4	0.535	0.525	0.555	0.555	0.557	0.555
	0.5	0.5	0.542	0.556	0.562	0.559	0.560	0.563
0.8	0.4	0.4	0.889	0.929	0.914	0.914	0.912	0.912
	0.4	0.5	0.896	0.944	0.923	0.920	0.919	0.918
	0.5	0.4	0.944	0.953	0.957	0.957	0.956	0.957
	0.5	0.5	0.950	0.968	0.965	0.964	0.964	0.963
0.9	0.4	0.4	0.907	0.951	0.935	0.934	0.932	0.934
	0.4	0.5	0.906	0.958	0.936	0.935	0.933	0.933
	0.5	0.4	0.956	0.968	0.968	0.968	0.967	0.968
	0.5	0.5	0.958	0.976	0.972	0.972	0.971	0.972

Table 5 shows the rejection frequencies for $T = 200$. Again similar results to $T = 100$ are obtained. The asymptotic test varies substantially in terms of size ranging from 0.047 to 0.067 for different values of the bandwidth parameter while the size of the RB has smaller variation ranging from 0.053 to 0.060. The power of the asymptotic test is close to the power of the RB with RB having larger power than asymptotic as s increases.

Table 5: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$. Model (13)-(14).

			T=200					
s	h_1	h_2	***Asymptotic***	***RB***	$b=2$	$b=4$	$b=6$	$b=8$
0	0.4	0.4	0.049	0.055	0.075	0.075	0.074	0.075
	0.4	0.5	0.047	0.053	0.064	0.064	0.064	0.063
	0.5	0.4	0.067	0.060	0.087	0.088	0.089	0.088
	0.5	0.5	0.054	0.057	0.074	0.074	0.074	0.071
0.2	0.4	0.4	0.053	0.055	0.078	0.077	0.075	0.074
	0.4	0.5	0.052	0.056	0.064	0.061	0.062	0.061
	0.5	0.4	0.078	0.065	0.095	0.095	0.096	0.094
	0.5	0.5	0.064	0.063	0.079	0.079	0.079	0.078
0.5	0.4	0.4	0.474	0.500	0.522	0.521	0.518	0.517
	0.4	0.5	0.476	0.517	0.503	0.499	0.498	0.493
	0.5	0.4	0.666	0.672	0.696	0.694	0.698	0.693
	0.5	0.5	0.675	0.700	0.704	0.700	0.700	0.699
0.8	0.4	0.4	0.927	0.963	0.952	0.951	0.950	0.949
	0.4	0.5	0.920	0.960	0.943	0.942	0.940	0.938
	0.5	0.4	0.970	0.986	0.982	0.982	0.981	0.980
	0.5	0.5	0.967	0.988	0.982	0.981	0.981	0.981
0.9	0.4	0.4	0.930	0.975	0.958	0.959	0.956	0.955
	0.4	0.5	0.920	0.966	0.947	0.946	0.944	0.944
	0.5	0.4	0.975	0.992	0.985	0.984	0.984	0.982
	0.5	0.5	0.970	0.990	0.984	0.983	0.982	0.981

Finally, in Table 6 I present the results for $T = 400$. The asymptotic test has small size distortions while the size of the RB is close to the nominal value. On the other hand, the block bootstrap's size depends somewhat on what block size is. In terms of power, asymptotic bootstrap and residual bootstrap's performance is quite close for low values of s while the difference increases for larger values of s in favour of the RB. As expected, the size distortions of the asymptotic test decrease with sample size. Overall, RB retains the proper size across all sample sizes, bandwidth parameter values and number of instruments whilst also outperforms the asymptotic test in terms of power.

Table 6: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$. Model (13)-(14).

			T=400					
s	h_1	h_2	***Asymptotic***	***RB***	$b = 2$	$b = 4$	$b = 6$	$b = 8$
0	0.4	0.4	0.042	0.048	0.066	0.066	0.065	0.060
	0.4	0.5	0.045	0.051	0.061	0.061	0.059	0.057
	0.5	0.4	0.049	0.052	0.075	0.072	0.073	0.069
	0.5	0.5	0.046	0.052	0.061	0.062	0.062	0.064
0.2	0.4	0.4	0.049	0.056	0.077	0.075	0.074	0.071
	0.4	0.5	0.053	0.059	0.064	0.064	0.061	0.062
	0.5	0.4	0.065	0.059	0.085	0.084	0.084	0.080
	0.5	0.5	0.060	0.062	0.073	0.075	0.075	0.075
0.5	0.4	0.4	0.601	0.629	0.661	0.661	0.659	0.649
	0.4	0.5	0.607	0.644	0.636	0.633	0.626	0.631
	0.5	0.4	0.815	0.825	0.847	0.847	0.843	0.837
	0.5	0.5	0.819	0.839	0.848	0.848	0.845	0.843
0.8	0.4	0.4	0.935	0.973	0.957	0.957	0.956	0.954
	0.4	0.5	0.925	0.965	0.945	0.944	0.942	0.944
	0.5	0.4	0.981	0.993	0.990	0.989	0.988	0.987
	0.5	0.5	0.972	0.990	0.982	0.982	0.981	0.980
0.9	0.4	0.4	0.936	0.979	0.960	0.960	0.960	0.957
	0.4	0.5	0.920	0.969	0.945	0.944	0.942	0.942
	0.5	0.4	0.978	0.992	0.985	0.986	0.984	0.982
	0.5	0.5	0.972	0.991	0.982	0.981	0.981	0.979

5. Empirical Application

In this section, I follow GKM and employ the local Hausman exogeneity test in a time varying version of the traditional Phillips curve. The goal here is to compare the asymptotic test with its bootstrap counterpart and see whether different results are obtained.

The original article[9] by GKM does not mention the dataset used and hence I consider data obtained from y. Louis FRED. Inflation π_t is computed as 100 times the seasonal log difference of the *CPIAUCSL* variable and the variable *UNRATE* is used for unemployment u_t. The sample period ranges from 1959:1 to 2021:12 to include COVID-19 pandemic. The model used is

$$\Delta\pi_t = c_t + \gamma_t \Delta\pi_{t-1} + \alpha_t \Delta u_t + e_t \tag{15}$$

where change in inflation is the dependent variable and change in unemployment together with one lag of the change in inflation are the independent variables. A Gaussian kernel is employed with bandwidth parameters $H = L = T^{0.7}$ since for these values the asymptotic Hausman test is shown to suffer the most in terms of size and power compared to its bootstrap version. It is worth noting that for my sample, there appears to be significant serial correlation and hence results should be viewed cautious because neither asymptotic test nor the bootstrap test allow for serial correlation.

Figure 1 shows the time varying OLS and IV coefficient estimates of α_t and γ_t with their associated 90% confidence intervals, respectively. The time-varying IV estimator is quite different from the time-varying OLS for the parameter α untill 2000 while for the remaining, the two overlap. The average values over time of $\hat{\alpha}$ and $\tilde{\alpha}$ are about -0.157 and -0.649 which are comparable to the full sample constant parameters OLS and 2SLS values, -0.109 and 0.127 respectively.

The lower panel of Figure 1 graphs $\hat{\gamma}_t$ and $\tilde{\gamma}_t$ with the associated 90% confidence intervals. The two estimators provide similar results and the two lines seem almost indistinguishable. The average value over time for $\hat{\gamma}_t$ is 0.356 while the full sample constant parameter OLS value amounts to 0.376.

[9]Following the suggestion by Lucchetti and Valentini (2021).

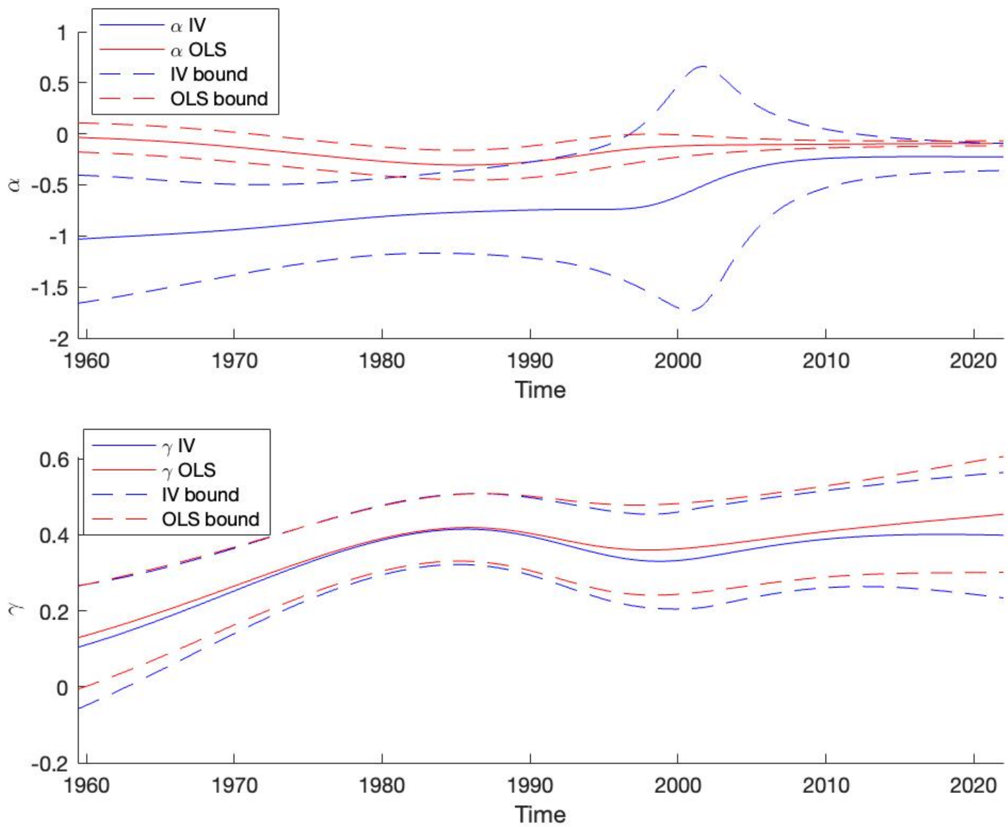


Figure 1: Empirical results for model (15).The two panels graph the OLS and IV coefficient estimates for α_t and γ_t respectively using $H = L = T^{0.7}$.

Figure 2 presents the p-values of the asymptotic time-varying Hausman test and its bootstrap version. The upper panel shows the empirical p-values of the two tests for $H = L = T^{0.7}$. For most of the sample, the two tests seem to provide the same results while for the period around 2000 the two tests show conflicting results. Specifically, the bootstrap version of the Hausman test rejects the null of exogeneity at 10% significance level while the asymptotic does not reject.

Turning now to the bandwidth parameter values of $H = L = T^{0.5}$ the lower panel of Figure 2 shows the p-values of the asymptotic test and its bootstrap counterpart. Given the Monte Carlo simulations in Section 4, these values were the values for which the asymptotic Hausman test experienced the smallest distortions in terms of size and power and hence can serve as a benchmark to compare the differences found in the two tests.

The two versions of the test in Figure 2 seem to be quite similar for most parts, except for the period around 2002 to 2006 where the bootstrap version of the test points out to rejecting the null while the asymptotic test shows exogeneity of the regressor. This finding is consistent with the Monte Carlo simulations considered in Section 4, since bootstrap test rejects the null more often that the asymptotic test.

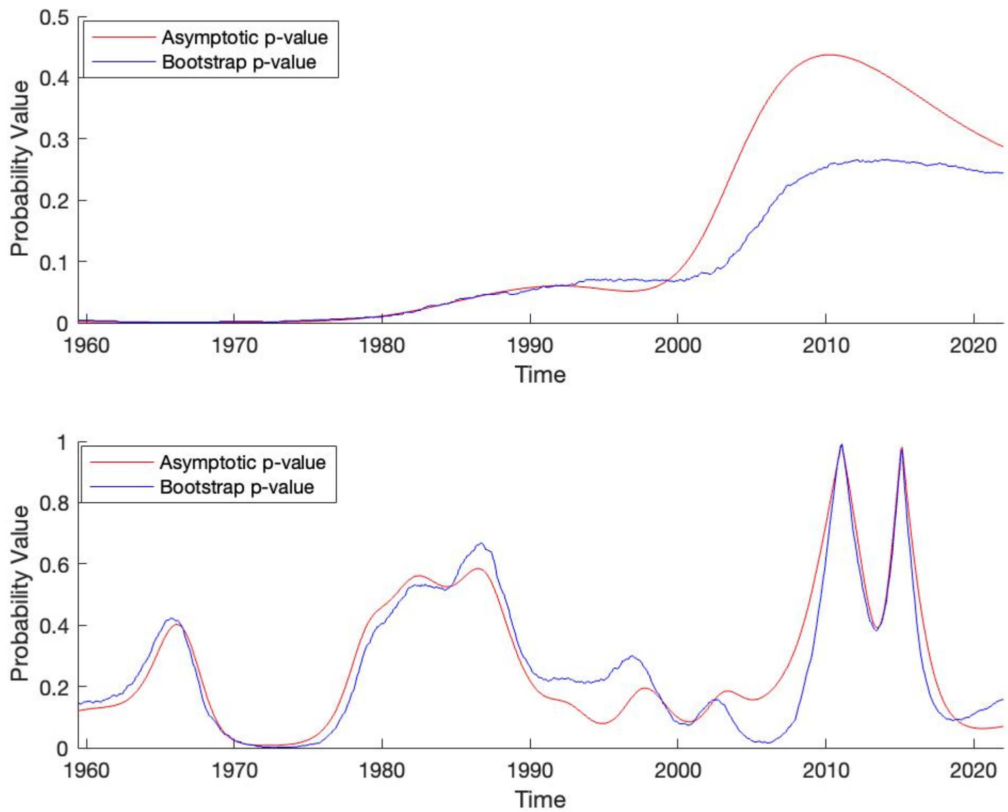


Figure 2: Empirical results for model (15).The two panels graph the empirical p-values of the aymptotic and bootstrap Hausman tests using $H = L = T^{0.7}$ and $H = L = T^{0.5}$ respectively.

Next, a forward looking (New-Keynesian) Phillips curve is also considered, as found in GKM and along the lines of Galí and Gertler, 1999. The New-Keynesian Phillips curve is given by:

$$\Delta\pi_t = c_t + \rho_t \Delta\pi^e_{t+1} + \gamma_t \Delta\pi_{t-1} + \alpha_t \Delta u_t + v_t \tag{16}$$

or written differently

$$\Delta\pi_t = c_t + \rho_t \Delta\pi_{t+1} + \gamma_t \Delta\pi_{t-1} + \alpha_t \Delta u_t + \epsilon_t \tag{17}$$

where $\epsilon_t = \rho_t(\Delta\pi^e_{t+1} - \Delta\pi_{t+1}) + v_t$, $\Delta\pi^e_{t+1}$ is the optimal one-step ahead forecast of $\Delta\pi_{t+1}$ made in period t, and v_t is an i.i.d error which is uncorrelated with all leads and lags with the forecast error $(\Delta\pi^e_{t+1} - \Delta\pi_{t+1})$. Obviously $\Delta\pi_{t+1}$ is correlated with the error term ϵ_t and hence a time-varying IV estimator is employed. I repeat the same experiment as above, using four lags of the change in unemployment and inflation as instruments, a Gaussian kernel and $H = L = 0.7$.

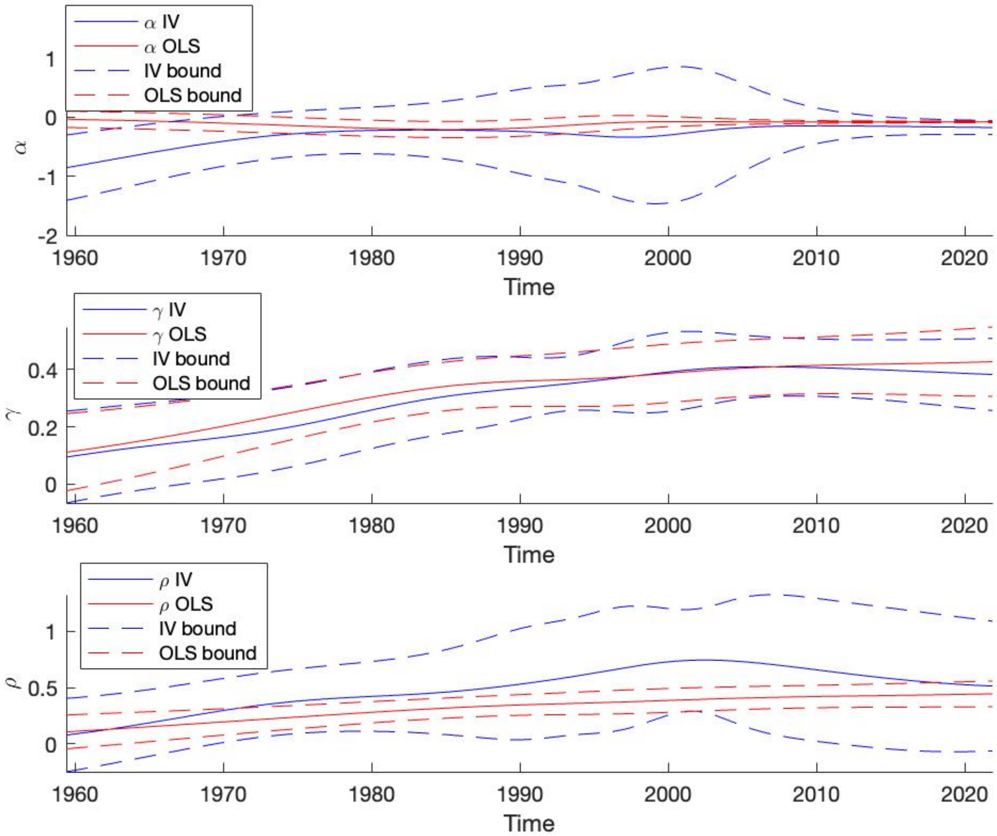


Figure 3: Empirical results for model (17).The three panels graph the OLS and IV coefficient estimates for α_t, γ_t and ρ_t respectively using $H = L = T^{0.7}$.

Figure 3 reports the coefficient estimates for this model. The upper panel of Figure 3 shows the coefficient estimates of $\hat{\alpha}_t$ and $\tilde{\alpha}_t$. The coefficient is close to 0 when estimated by time-varying OLS and never significant. Similarly, the time-varying IV estimator provides estimates close to 0 except for a small period around '60s.

The middle panel of Figure 3 depicts the estimates of $\hat{\gamma}_t$ and $\tilde{\gamma}_t$. They both seem to perform quite similarly as the two lines seem almost identical. The lower panel graphs the results for $\hat{\rho}_t$ and $\tilde{\rho}_t$. The two lines appear to be deviating only for a short period around the '2000s.

Finally, the upper panel of Figure 4 graphs the p-values of the time-varying Hausman test and its bootstrap version for $H = L = T^{0.7}$. The asymptotic test provides lower p-values for the first half of the sample than the bootstrap test. Interestingly, the bootstrap test rejects the null of exogeneity around the '2000s while the asymptotic does not.

The lower panel of Figure 4 now depicts the p-values for the two tests for $H = L = T^{0.5}$. The two lines seem almost exact for most of the period except for '2000s where again the bootstrap rejects the exogeneity for a small period while the asymptotic does not.

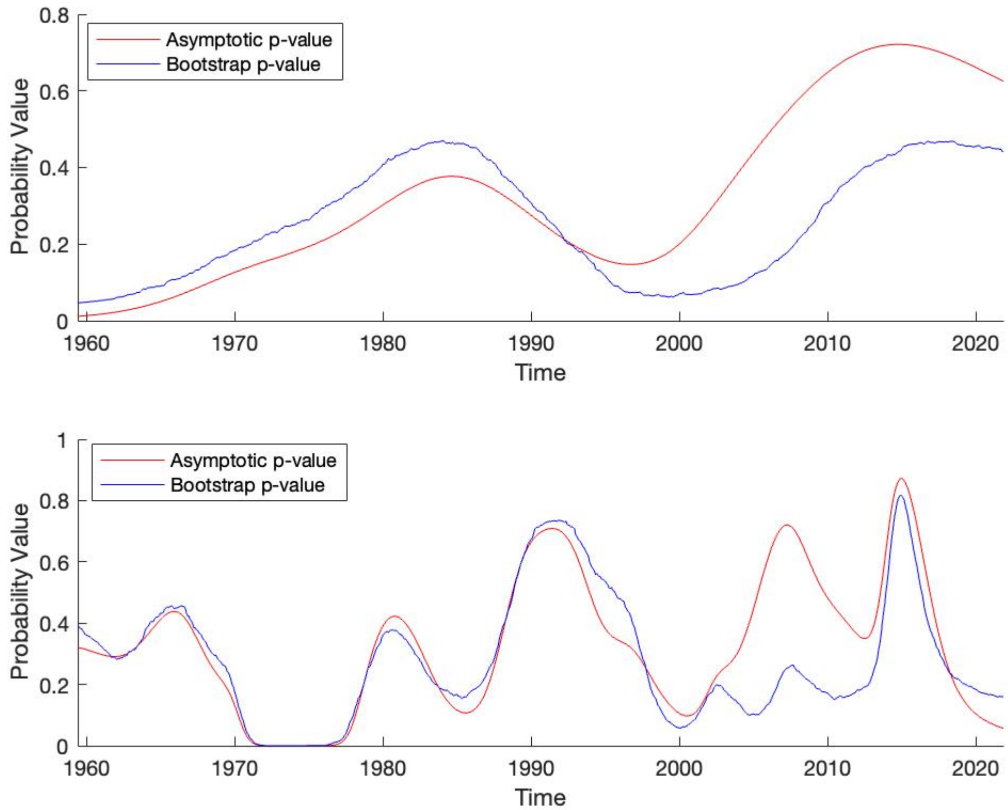


Figure 4: Empirical p-values of the time-varying Hausman test for model (17) and its bootstrap version for $H = L = T^{0.7}$ and $H = L = T^{0.5}$ respectively.

In summary, both the asymptotic and the time-varying boostrap version of the Hausman exogeneity test appear to perform quite similarly for most parts of the sample. However, contrary to the asymptotic test, its bootstrap version points out to endogeneity of unemployment for a short period around '2000s. In light of the results obtained in Section 4, this is consistent with my findings that the bootstrap test rejects the null more often than the asymptotic.

6. Conclusion

A usual assumption made when carrying out IV estimation is that the model, and hence the parameter vector, does not change through time. This in turn, also implies that the endogeneity status of the variable remains constant through time. This assumption, although crucial is often highly susceptible. A new strand of literature lead primarily by Giraitis et al. (2020) has proposed a non-parametric IV estimation based on kernels, and allowing for both deterministic or random coefficients. Consequently, a time-varying Hausman exogeneity test has been developed to test for a possible switching endogeneity status at a specific point. However, this test appears to experience size distortions and have low power.

In this paper, I propose to bootstrap this test. The performance of the bootstrapped test is evaluated through Monte Carlo simulations. As it is shown, bootstrap provides considerable refinements over the asymptotic test statistic with higher power, and size close to nominal. More importantly, the size and power of the bootstraped test are invariant to the choice of the bandwidth parameters and the number of instruments.

Revisiting the empirical application by GKM, I estimate a Phillips curve for the USA, using unemployment as the forcing variable for inflation, and examine whether similar results are obtained using the asymptotic test and its bootstrap counterpart. The two tests seem to perform quite similarly for most parts except for a period around 2000 where the bootstrap test points out to endogeneity of unemployment. These results seem consistent with the Monte Carlo simulations, since the bootstrap test rejects the null more often that the asymptotic.

I finish with some open questions for further research. First, this article has focused only on the case of the local Hausman exogeneity test, but the seminal paper by Giraitis et al. (2020) also proposes a Global Hausman test for testing for possible endogeneity in a specified interval. Preliminary analysis suggests that bootstrap refinements could be obtained also in the case of the Global Hausman test. Second, in this article I have addressed the issue of bootstrapping the local Hausman test so that its size and power does not depend on the bandwidth parameter and the number of instruments used. However, one could also select the bandwidth parameter using some form of calibration as in Shao and Politis (2012) and examine whether similar results are obtained.

Appendices

A. Appendix

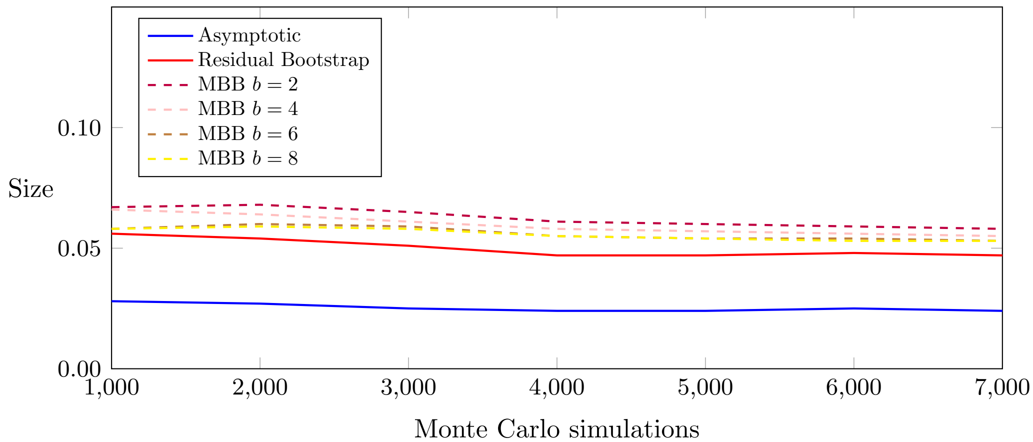
In Appendix A, I assess the robustness of the results with respect to a number of different attributes of the experiments employed in the paper.

A.1 Robustness Analysis

i) **Number of Monte Carlo Simulations**

Figure 5 shows the results for the size of the asymptotic and bootstrap versions of the test for (10)-(12), $T = 100$ and $h_1 = h_2 = 0.4$ across different number of Monte Carlo simulations. Convergence is attained at 5000 simulations.

Figure 5: Coverage-size convergence for bootstrap and asymptotic versions of the Hausman test for model (10)-(12), $T = 100$ and $h_1 = h_2 = 0.4$.



ii) **Bootstrap Alternatives**

Although in this article I employ the MBB and the RB, I also consider aanother alternative. To begin, by employing the MBB I implicitly assume that the size of each block is equal and fixed to b and that the sample size of each replication is equal to T which requires some sort of trimming in case T/b is not an integer and hence I have ignored any end effects. For example, because there is no data after $\{y_T, x_T\}$ the moving blocks method does not define a block of length b beginning at the end points. Politis and Romano (1991) proposed the Circular Bootstrap to "wrap" the data around in a circle so that $\{y_1, x_1\}$ follows directly after $\{y_T, x_T\}$. This path has also been explored in a preliminary analysis but provided similar results to MBB.

iii) Further Monte Carlo results for higher values of the bandwidth parameters H and L are provided here.

Table 7: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$. Model (10)-(12).

				T=100				
s	**h1**	***h2***	***Asymptotic***	***Residual***	***b=2***	***b=4***	***b=6***	***b=8***
0	0.4	0.7	0.043	0.045	0.050	0.048	0.048	0.046
	0.5	0.7	0.041	0.048	0.051	0.048	0.048	0.046
	0.7	0.4	0.074	0.067	0.101	0.098	0.099	0.097
	0.7	0.5	0.070	0.065	0.094	0.093	0.092	0.094
	0.7	0.7	0.052	0.057	0.061	0.063	0.062	0.061
0.2	0.4	0.7	0.048	0.051	0.051	0.049	0.047	0.047
	0.5	0.7	0.053	0.056	0.055	0.054	0.055	0.052
	0.7	0.4	0.094	0.080	0.118	0.118	0.117	0.115
	0.7	0.5	0.087	0.076	0.108	0.109	0.105	0.106
	0.7	0.7	0.068	0.069	0.078	0.076	0.077	0.076
0.5	0.4	0.7	0.281	0.348	0.277	0.273	0.268	0.265
	0.5	0.7	0.408	0.481	0.415	0.406	0.401	0.400
	0.7	0.4	0.704	0.739	0.758	0.754	0.755	0.755
	0.7	0.5	0.692	0.734	0.749	0.747	0.749	0.747
	0.7	0.7	0.639	0.708	0.691	0.688	0.687	0.683
0.8	0.4	0.7	0.642	0.768	0.666	0.659	0.653	0.647
	0.5	0.7	0.720	0.832	0.761	0.754	0.748	0.745
	0.7	0.4	0.942	0.969	0.969	0.969	0.968	0.967
	0.7	0.5	0.928	0.967	0.965	0.964	0.963	0.962
	0.7	0.7	0.868	0.951	0.933	0.932	0.932	0.929
0.9	0.4	0.7	0.642	0.769	0.675	0.666	0.661	0.656
	0.5	0.7	0.720	0.834	0.767	0.762	0.755	0.754
	0.7	0.4	0.946	0.975	0.972	0.972	0.971	0.971
	0.7	0.5	0.931	0.972	0.966	0.966	0.965	0.965
	0.7	0.7	0.869	0.958	0.934	0.933	0.933	0.930

Table 8: Rejection frequencies for the local Hausman test at $t = T/2$ and for $\alpha = 0.05$. Model (10)-(12).

			T=200					
s	***h1***	***h2***	***Asymptotic***	***Residual***	***b=4***	***b=6***	***b=8***	***b=16***
0	0.4	0.7	0.042	0.048	0.050	0.046	0.044	0.043
	0.5	0.7	0.043	0.050	0.051	0.048	0.049	0.044
	0.7	0.4	0.100	0.080	0.124	0.123	0.122	0.114
	0.7	0.5	0.098	0.081	0.120	0.118	0.118	0.114
	0.7	0.7	0.072	0.074	0.088	0.085	0.084	0.082
0.2	0.4	0.7	0.050	0.054	0.051	0.050	0.049	0.045
	0.5	0.7	0.054	0.057	0.058	0.054	0.053	0.050
	0.7	0.4	0.135	0.105	0.154	0.152	0.151	0.144
	0.7	0.5	0.131	0.105	0.150	0.147	0.147	0.139
	0.7	0.7	0.100	0.093	0.110	0.108	0.105	0.102
0.5	0.4	0.7	0.325	0.427	0.326	0.315	0.313	0.299
	0.5	0.7	0.493	0.581	0.503	0.497	0.489	0.469
	0.7	0.4	0.815	0.835	0.853	0.852	0.851	0.846
	0.7	0.5	0.806	0.836	0.851	0.849	0.848	0.843
	0.7	0.7	0.748	0.806	0.800	0.799	0.794	0.789
0.8	0.4	0.7	0.651	0.776	0.692	0.684	0.679	0.669
	0.5	0.7	0.734	0.844	0.787	0.783	0.778	0.767
	0.7	0.4	0.975	0.985	0.988	0.988	0.987	0.987
	0.7	0.5	0.967	0.984	0.985	0.985	0.985	0.983
	0.7	0.7	0.913	0.965	0.957	0.955	0.953	0.949
0.9	0.4	0.7	0.639	0.768	0.685	0.678	0.673	0.659
	0.5	0.7	0.725	0.843	0.785	0.778	0.778	0.767
	0.7	0.4	0.979	0.987	0.990	0.989	0.988	0.987
	0.7	0.5	0.968	0.985	0.984	0.984	0.983	0.982
	0.7	0.7	0.911	0.967	0.955	0.953	0.953	0.949

References

Beran, R. and Ducharme, G. (1991). Asymptotic theory for bootstrap methods in statistics.

Bickel, P. J. and Freedman, D. A. (1981). Some Asymptotic Theory for the Bootstrap. *The Annals of Statistics*, 9(6):1196 – 1217.

Bühlmann, P. and Künsch, H. R. (1999). Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis*, 31(3):295–310.

Chen, B. (2015). Modeling and testing smooth structural changes with endogenous regressors. *Journal of Econometrics*, 185(1):196–215.

Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, 146(1):162–169.

Davidson, R. and MacKinnon, J. G. (2000). Bootstrap tests: how many bootstraps? *Econometric Reviews*, 19(1):55–68.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

fu Wong, K. (1996). Bootstrapping hausman's exogeneity test. *Economics Letters*, 53(2):139–143.

Galí, J. and Gertler, M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, 44(2):195–222.

Giraitis, L., Kapetanios, G., and Marcellino, M. (2020). Time-varying instrumental variable estimation. *Journal of Econometrics*.

Giraitis, L., Kapetanios, G., and Yates, T. (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics*, 179(1):46–65.

Goncalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, 123(1):89–120.

Hall, P. (1992). The bootstrap and edgeworth expansion.

Hall, P., HOROWITZ, J. L., and JING, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

Kapetanios, G. (2010). Testing for exogeneity in threshold models. *Econometric Theory*, 26(1):231–259.

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1):386–404.

Lucchetti, R. and Valentini, F. (2021). Kernel-based time-varying iv estimation: handle with care. Technical report, University Library of Munich, Germany.

Mammen, E. (1993). Bootstrap and Wild Bootstrap for High Dimensional Linear Models. *The Annals of Statistics*, 21(1):255 – 285.

Politis, D. N. (2003). The Impact of Bootstrap Methods on Time Series Analysis. *Statistical Science*, 18(2):219 – 230.

Politis, D. N. and Romano, J. P. (1991). *A circular block-resampling procedure for stationary data*. Purdue University. Department of Statistics.

Shao, X. and Politis, D. N. (2012). Fixed-b subsampling and block bootstrap: Improved confidence sets based on p-value calibration.

Wu, C. F. J. (1986). Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics*, 14(4):1261 – 1295.

Zhang, C., Osborn, D. R., and Kim, D. H. (2008). The new keynesian phillips curve: From sticky inflation to sticky prices. *Journal of Money, Credit and Banking*, 40(4):667–699.