Title

Heteroscedasticity and Autocorrelation Robust F and t Tests in Stata

Authors

Ye, Xiaoqing
Sun, Yixiao

Publication Date

2018-07-09

Heteroscedasticity and Autocorrelation Robust F and t Tests in Stata

Xiaoqing Ye
School of Mathematics and Statistics
South-Central University for Nationalities
Wuhan, Hubei 430074, China
yshtim@126.com

Yixiao Sun
Department of Economics
UC San Diego
La Jolla, CA 92093-0508, USA
yisun@ucsd.edu

Abstract. In this article, we consider time series OLS and IV regressions and introduce a new pair of commands, `har` and `hart`, which implement a more accurate class of heteroscedasticity and autocorrelation robust (HAR) F and t tests. These tests represent part of the recent progress on HAR inference. The F and t tests are based on the convenient F and t approximations and are more accurate than the conventional chi-squared and normal approximations. The underlying smoothing parameters are selected to target the type I and type II errors, the two fundamental objects in every hypothesis testing problem. The estimation command `har` and the post-estimation test command `hart` allow for both kernel HAR variance estimators and orthonormal series HAR variance estimators. In addition, we introduce another pair of new commands, `gmmhar` and `gmmhart` which implement the recently developed F and t tests in a two-step GMM framework. For this command we opt for the orthonormal series HAR variance estimator based on the Fourier bases, as it allows us to develop convenient F and t approximations as in the first-step GMM framework. Finally, we present several examples to demonstrate the use of these commands.

Keywords: st0001, har, gmmhar, fixed-smoothing, kernel function, orthonormal series, testing-optimal, AMSE, OLS/IV, Two-step GMM, J statistic

1 Introduction

The last two decades have witnessed substantial progress in heteroskedasticity and autocorrelation robust (HAR) inference.

First, the fixed-smoothing asymptotic theory, a new class of asymptotic theory, has been developed. See, for example, Kiefer and Vogelsang (2005) and Sun (2014a) and the references therein. It is now well known that fixed-smoothing asymptotic approximations are more accurate than conventional increasing-smoothing asymptotic approximations, i.e., the chi-squared and normal approximations. The higher accuracy, which is supported by ample numerical evidence, has been established rigorously via high order Edgeworth expansions in Jansson (2004) and Sun et al. (2008). The source of the accuracy is that the new asymptotic approximations capture the estimation uncertainty in the nonparametric HAR variance estimator. Both the effect of the smoothing parameter and the form of the variance estimator are retained in the fixed-smoothing asymptotic approximations. In addition, the estimation error in the model parameter estimator is also partially reflected in the new asymptotic approximations.

Second, a new rule for selecting the smoothing parameter that is optimal for the HAR testing has been developed. It has been pointed out that the mean squared error of the variance estimator is not the most suitable criterion to use in the testing context. For hypothesis testing, the ultimate goals are the type I error and the type II error. One should choose the smoothing parameter to minimize a loss function that is a weighted sum of the type I and type II errors with the weights reflecting the relative consequences of committing these two types of errors. Alternatively and equivalently, one should minimize one type of error subject to the control of the other type of error. See Sun et al. (2008) and Sun (2014a) for the choices of the smoothing parameter that are oriented toward the testing problem at hand.

Finally, while kernel methods are widely used in practice, there is a renewed interest in using a different nonparametric variance estimator that involves a sequence of orthonormal basis functions. In a special case, this gives rise to the simple average of periodograms as an estimator of the spectral density at zero. Such an estimator is a familiar choice in the literature on spectral density estimation. The advantage of using the orthonormal series (OS) HAR variance estimator is that the fixed-smoothing asymptotic distribution is the standard F or t distribution. There is no need to simulate any critical value. This is in contrast to the case with the usual kernel HAR variance estimator where nonstandard critical values have to be simulated.

The fixed-smoothing asymptotic approximations have been established in various settings. For the kernel HAR variance estimators, the smoothing parameter can be parametrized as the ratio of the truncated lag (for truncated kernels) to the sample size. This ratio is often denoted by b, and the fixed-smoothing asymptotics are referred to as the fixed-b asymptotics in the literature. The fixed-b asymptotics have been developed by Kiefer and Vogelsang (2002a), Kiefer and Vogelsang (2002b), Kiefer and Vogelsang (2005), Jansson (2004), Sun et al. (2008), and Gonçlaves and Vogelsang (2011) in the time series setting, Bester et al. (2016) and Sun and Kim (2015) in the spatial setting, and Gonçlaves (2011), Kim and Sun (2013), and Vogelsang (2012) in the panel data setting. For the OS HAR variance estimators, the smoothing parameter is the number of basis functions used. This smoothing parameter is often denoted by K, and the fixed-smoothing asymptotics are often called the fixed-K asymptotics. For its theoretical development and related simulation evidence, see, for example, Phillips (2005), Müller (2007), Sun (2011), and Sun (2013). A recent paper by Lazarus et al. (2016) shows that tests based on the OS HAR variance estimator have competitive power compared to tests based on the kernel HAR variance estimator with the optimal kernel.

Most research on fixed-smoothing asymptotics has been devoted to first-step GMM estimation and inference. More recently, fixed-smoothing asymptotics have been established in a general two-step GMM framework. See Sun and Kim (2012), Sun (2013), Sun (2014b), and Hwang and Sun (2017). The key difference between first-step GMM and two-step GMM is that in the latter case the HAR variance estimator not only appears in the covariance estimator but also plays the role of the optimal weighting matrix in the second-step GMM criterion function.

While the fixed-smoothing approximations are more accurate than the conventional increasing-smoothing approximations, they have not been widely adopted in empirical applications. There are two possible reasons. First, the fixed-smoothing asymptotic distributions based on popular kernel variance estimators are nonstandard, and therefore critical values have to be simulated. Second, there is no Stata command that implements the new and more accurate approximations.

In this article, we describe the new estimation command `har` and the post-estimation test command `hart`, which implement the fixed-smoothing Wald and t tests of Sun (2013) and Sun (2014a) for linear regression models with possibly endogenous covariates. These two commands automatically select the testing-optimal smoothing parameter. In addition, we provide another pair of commands `gmmhar` and `gmmhart` that implement the fixed-smoothing Wald and t tests in a two-step efficient GMM setting, introduced in Hwang and Sun (2017). Under the fixed-smoothing asymptotics, Hwang and Sun (2017) show that the modified Wald statistic is asymptotically F distributed, and the modified t statistic is asymptotically t distributed. So the new tests are very convenient to use. In addition, Sun and Kim (2012) show that under the fixed-smoothing asymptotics, the J statistic for testing over-identification is also asymptotically F distributed.

The remainder of the article is organized as follows: In Sections 2 and 3, we present the fixed-smoothing inference based on the first-step estimator and the two-step estimator, respectively. In Sections 4 and 5, we describe the syntaxes of `har` and `gmmhar` and illustrate their usage. Section 6 presents some simulation evidence and Section 7 describes the two post-estimation test commands: `hart` and `gmmhart`. The last section concludes and discusses future work.

2 Fixed-smoothing Asymptotics: First-step GMM

2.1 OLS and IV Regressions

Consider the regression model

$$Y_t = X_t\theta_0 + e_t,\ t = 1, ..., T,$$

where $\{e_t\}$ is a zero mean process that may be correlated with the covariate process $\{X_t \in \mathbb{R}^{1\times d}\}$. There are instruments $\left\{Z_t \in \mathbb{R}^{1\times m}\right\}$ such that the moment conditions

$$EZ_t'\left(Y_t - X_t\theta\right) = 0$$

hold if and only if $\theta = \theta_0$. When X_t is exogenous, we take $Z_t = X_t$, leading to the moment conditions behind the OLS estimator. Note that the first elements of X_t and Z_t are typically 1. We allow the process $\{Z_t'e_t\}$ to have autocorrelation of unknown forms. The model may be over-identified with the degree of over-identification $q = m - d \geq 0$.

Define

$$S_{ZX} = \frac{1}{T}\sum_{t=1}^{T} Z_t'X_t,\ S_{ZZ} = \frac{1}{T}\sum_{t=1}^{T} Z_t'Z_t, S_{ZY} = \frac{1}{T}\sum_{t=1}^{T} Z_t'Y_t.$$

Then the IV estimator of θ_0 is given by

$$\hat{\theta}_{IV} = \left[S'_{ZX}W_{0T}^{-1}S_{ZX}\right]^{-1}\left[S'_{ZX}W_{0T}^{-1}S_{ZY}\right], \quad (1)$$

where $W_{0T} = S_{ZZ} \in \mathbb{R}^{m\times m}$. For the asymptotic results that follow, we can allow W_{0T} to be a general weighting matrix. It suffices to assume that $\text{Plim}_{T\to\infty}W_{0T} = W_0$ for a positive definite nonrandom matrix W_0. When $Z_t = X_t$, the IV estimator reduces to the OLS estimator.

Suppose we are interested in testing the null $H_0 : R\theta_0 = r$ against the alternative $H_1 : R\theta_0 \neq r$, where $r \in \mathbb{R}^{p\times 1}$ and $R \in \mathbb{R}^{p\times d}$ is a matrix of full row rank. Nonlinear restrictions can be converted into linear ones via the delta method. Under some standard high-level conditions, we have

$$\sqrt{T}R(\hat{\theta}_{IV} - \theta_0) = \sqrt{T}(R\hat{\theta}_{IV} - r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t + o_p(1),$$

where, for $G_0 = ES_{ZX} \in \mathbb{R}^{m\times d}$, $u_t = R\left(G_0'W_0^{-1}G_0\right)^{-1}G_0'W_0^{-1}Z_t'e_t$ is the transformed moment process. It then follows that

$$\sqrt{T}R(\hat{\theta}_{IV} - \theta_0) \xrightarrow{d} N(0, \Omega),$$

where $\Omega = \sum_{j=-\infty}^{j=+\infty} Eu_tu'_{t-j}$ is the long run variance (LRV) of $\{u_t\}$.

The Wald statistic for testing H_0 against H_1 is:

$$F_{IV} = \left[\sqrt{T}(R\hat{\theta}_{IV} - r)\right]'\hat{\Omega}^{-1}\left[\sqrt{T}(R\hat{\theta}_{IV} - r)\right]/p, \quad (2)$$

where $\hat{\Omega}$ is an estimator of Ω. When $p = 1$, we can construct the t statistic

$$t_{IV} = \frac{\sqrt{T}(R\hat{\theta}_{IV} - r)}{\sqrt{\hat{\Omega}}}.$$

Let $G_T = S_{ZX}$,

$$\hat{u}_t = R(G_T'W_{0T}^{-1}G_T)^{-1}G_T'W_{0T}^{-1}Z_t'(Y_t - X_t\hat{\theta}_{IV}), \text{ and } \hat{u}^{ave} = T^{-1}\sum_{s=1}^{T}\hat{u}_s. \quad (3)$$

We consider the estimator $\hat{\Omega}$ of the form

$$\hat{\Omega} = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} Q_h(\frac{s}{T}, \frac{t}{T})\left(\hat{u}_t - \hat{u}^{ave}\right)\left(\hat{u}_s - \hat{u}^{ave}\right)', \quad (4)$$

where $Q_h(r, s)$ is a weighting function, and h is the smoothing parameter.

The above estimator includes the kernel HAR variance estimators and the OS HAR variance estimators as special cases. For the kernel LRV estimator, we let $Q_h(r, s) =$

$k\left((r-s)/b\right)$ and $h=1/b$ for a kernel function $k\left(\cdot\right)$. In this case, the estimator $\hat{\Omega}$ can be written in a more familiar form that involves a weighted sum of autocovariances:

$$\hat{\Omega}=\sum_{j=-(T-1)}^{T-1}k\left(\frac{j}{M_T}\right)\hat{\Gamma}_j, \tag{5}$$

where

$$\hat{\Gamma}_j=\begin{cases} T^{-1}\sum_{t=j+1}^{T}[\hat{u}_t-\hat{u}^{ave}][\hat{u}_{t-j}-\hat{u}^{ave}]' \text{ for } j\geq 0, \\ T^{-1}\sum_{t=j+1}^{T}[\hat{u}_{t+j}-\hat{u}^{ave}][\hat{u}_t-\hat{u}^{ave}]' \text{ for } j<0, \end{cases}$$

and $M_T=bT$ is the so-called truncation lag. This is a misnomer, as the kernel function may not have bounded support. Nevertheless, we follow the literature and refer to M_T as the truncation lag.

For the OS HAR variance estimator, we let

$$Q_h\left(r,s\right)=K^{-1}\sum_{j=1}^{K}\phi_j\left(r\right)\phi_j\left(s\right)$$

and $h=K$, where $\{\phi_j\left(\cdot\right)\}_{j=1}^{K}$ are orthonormal basis functions on $L^2[0,1]$ satisfying $\int_0^1\phi_j\left(r\right)dr=0$ for $j=1,...,K$. Here we assume that K is even and focus only on the Fourier basis functions:

$$\phi_{2j-1}(x)=\sqrt{2}\cos(2j\pi x) \text{ and } \phi_{2j}(x)=\sqrt{2}\sin(2j\pi x) \text{ for } j=1,...,K/2.$$

In this case, $\hat{\Omega}$ is equal to the average of the first $K/2$ periodograms multiplied by 2π. Other basis functions can be used, but the form of the basis functions does not seem to make a difference.

For both the kernel and OS HAR variance estimators, we parametrize h in such a way so that h indicates the amount of smoothing. We consider the fixed-smoothing asymptotics under which $T\to\infty$ for a fixed h. Let

$$Q_h^*(r,s)=Q_h(r,s)-\int_0^1Q_h(\tau,s)d\tau-\int_0^1Q_h(r,\tau)d\tau+\int_0^1\int_0^1Q_h(\tau_1,\tau_2)d\tau_1d\tau_2.$$

It follows from Kiefer and Vogelsang (2005) and Sun (2014a,b) that when h is fixed,

$$F_{IV}\to^d F_\infty(p,h) \text{ and } t_{IV}\to^d t_\infty(p,h),$$

where

$$F_\infty(p,h)=W_p'\left(1\right)C_{pp}^{-1}W_p\left(1\right)/p, \tag{6}$$

$$t_\infty(p,h)=W_p\left(1\right)/\sqrt{C_{pp}}, \tag{7}$$

$$C_{pp}=\int_0^1\int_0^1Q_h^*(r,s)dW_p(r)dW_p'(s),$$

and $W_p(r)$ is the standard p-dimensional Brownian motion process.

2.2 The kernel case

For the kernel case, the limiting distributions $F_\infty(p,h)$ and $t_\infty(p,h)$ are nonstandard. The critical values, i.e., the quantiles of $F_\infty(p,h)$ and $t_\infty(p,h)$, have to be simulated. This hinders the use of the new approximation in practice. Sun (2014a) establishes a standard F approximation to the nonstandard distribution $F_\infty(p,h)$. In particular, Sun (2014a) shows that the $100(1-\alpha)\%$ quantile of the distribution $F_\infty(p,h)$ can be approximated well by

$$\mathcal{F}_{IV}^{1-\alpha} := \kappa \mathcal{F}_{p,K}^{1-\alpha}$$

where $\mathcal{F}_{p,K}^{1-\alpha}$ is the $100\,(1-\alpha)\,\%$ quantile of the standard $F_{p,K}$ distribution,

$$K = \max\left(\left\lceil \frac{1}{bc_2} \right\rceil, p\right) - p + 1 \tag{8}$$

is the equivalent degrees of freedom ($\lceil . \rceil$ is the ceiling function), and

$$\kappa = \frac{\exp\left(b\left[c_1 + (p-1)\,c_2\right]\right) + \left(1 + b\left[c_1 + (p-1)\,c_2\right]\right)}{2} \tag{9}$$

is a correction factor. In the above, $c_1 = \int_{-\infty}^{\infty} k(x)dx, c_2 = \int_{-\infty}^{\infty} k^2(x)dx$. For the Bartlett kernel, $c_1 = 1$, $c_2 = 2/3$; For the Parzen kernel, $c_1 = 3/4$, $c_2 = 0.539285$; For the quadratic spectral (QS) kernel, $c_1 = 1.25$, $c_2 = 1$.

For the fixed-smoothing test based on the t statistic, we can use the approximate critical value

$$\mathfrak{t}_{IV}^{1-\alpha} = \begin{cases} \sqrt{\kappa \mathcal{F}_{1,K}^{1-2\alpha}}, & \alpha < 0.5, \\ -\sqrt{\kappa \mathcal{F}_{1,K}^{2\alpha-1}}, & \alpha \geq 0.5. \end{cases} \tag{10}$$

To see this, consider the case $\alpha < 0.5$. Since $t_\infty(p,h)$ is symmetric, its $1-\alpha$ quantile $t_\infty^{1-\alpha}$ is positive. By definition,

$$\begin{aligned} 1-\alpha &= P\left(t_\infty(p,h) < t_\infty^{1-\alpha}\right) \\ &= 1 - P\left(t_\infty(p,h) \geq t_\infty^{1-\alpha}\right) = 1 - \frac{1}{2} P\left(\left|t_\infty(p,h)\right|^2 \geq \left|t_\infty^{1-\alpha}\right|^2\right) \\ &= 1 - \frac{1}{2} P\left(F_\infty(1,h) \geq \left|t_\infty^{1-\alpha}\right|^2\right) = \frac{1}{2} + \frac{1}{2} P\left(F_\infty(1,h) < \left|t_\infty^{1-\alpha}\right|^2\right). \end{aligned}$$

So $P(F_\infty(1,h) < \left|t_\infty^{1-\alpha}\right|^2) = 1-2\alpha$, which implies that $\left|t_\infty^{1-\alpha}\right|^2$ is the $(1-2\alpha)$ quantile of the distribution $F_\infty(1,h)$. Therefore, we can take $\mathfrak{t}_{IV}^{1-\alpha} = \sqrt{\kappa \mathcal{F}_{1,K}^{1-2\alpha}}$ as the approximate critical value. The result for $\alpha \geq 0.5$ can be similarly proved. For a two-sided t test, we use the $(1-\alpha/2)$ quantile of $t_\infty(p,h)$ as the critical value for a test with nominal size α. This quantile can be approximated by $\sqrt{\kappa \mathcal{F}_{1,K}^{1-\alpha}}$.

The test based on the scaled F critical value $\kappa \mathcal{F}_{p,K}^{\alpha}$ is an approximate fixed-smoothing test. Sun (2014a) establishes asymptotic approximations to the type I and type II errors of this test. Given the approximate type I and type II errors $e_I(b)$ and $e_{II}(b)$, we can select the bandwidth parameter b to solve the constrained minimization problem:

$$b_{\text{opt}} = \arg\min e_{II}(b), \text{ s.t. } e_I(b) \leq \tau\alpha,$$

for some tolerance parameter $\tau > 1$. For our new commands, we take $\tau = 1.15$.

Consider the local alternative $H_1(\delta) : R\theta_0 = r + \Omega^{1/2}\tilde{c}/\sqrt{T}$ for $\tilde{c}$ uniformly distributed on $\mathcal{S}_p(\delta^2) = \{\tilde{c} \in \mathbb{R}^p : \|\tilde{c}\|^2 = \delta^2\}$. Let $G'_{p,\delta^2}(\cdot)$ be the pdf of the noncentral $\chi^2_p(\delta^2)$ distribution with degrees of freedom p and noncentrality parameter δ^2. It is shown that the test-optimal smoothing parameter b for testing H_0 against the alternative $H_1(\delta)$ at the significance level α is given by

$$b_{\text{opt}} = \begin{cases} \left[\frac{2qG'_{p,\delta^2}\left(\mathcal{X}_p^{1-\alpha}\right)|\bar{\mathcal{B}}|}{\delta^2 G'_{(p+2),\delta^2}\left(\mathcal{X}_p^{1-\alpha}\right)c_2}\right]^{\frac{1}{q+1}} T^{-\frac{q}{q+1}}, & \bar{\mathcal{B}} > 0, \\ \left[\frac{G'_p\left(\mathcal{X}_p^{1-\alpha}\right)\mathcal{X}_p^{1-\alpha}|\bar{\mathcal{B}}|}{(\tau-1)\alpha}\right]^{1/q} \frac{1}{T}, & \bar{\mathcal{B}} \leq 0, \end{cases} \tag{11}$$

where $\mathcal{X}_p^{1-\alpha}$ is the $(1-\alpha)$ quantile of the chi-squared distribution χ^2_p with p degrees of freedom, δ^2 is chosen according to $\Pr\left(\chi_p\left(\delta^2\right) > \mathcal{X}_p^{1-\alpha}\right) = 75\%$,

$$\bar{\mathcal{B}} = tr\left\{\mathcal{B}\Omega^{-1}\right\}/p, \ \mathcal{B} = -\rho_q \sum_{h=-\infty}^{\infty} |h|^q \, E u_t u'_{t-h},$$

q is the order of the kernel used, and ρ_q is the Parzen characteristic exponent of the kernel. For the Bartlett kernel $q = 1$, $\rho_q = 1$; For the Parzen kernel $q = 2$, $\rho_q = 6$; For the QS kernel $q = 2$, $\rho_q = 1.421223$.

For a one-sided fixed-smoothing t test, the testing-optimal b is not available from the literature. We suggest using the rule in (11).

The parameter $\bar{\mathcal{B}}$ can be estimated by a standard VAR(1) plug-in procedure. This is what we opt for in the new commands. Plugging the estimate of $\bar{\mathcal{B}}$ into (11) yields $\hat{b}_{\text{temp}}$. The data-driven choice of b_{opt} is then given by $\hat{b}_{\text{opt}} = \min(\hat{b}_{\text{temp}}, 0.5)$. We do not use a b larger than 0.5 in order to avoid large power loss.

2.3 The orthonormal series case

For the orthonormal series case, Sun (2013) shows that under the fixed-smoothing asymptotics,

$$F_{IV} \xrightarrow{d} \frac{K}{K-p+1} \cdot \mathfrak{F}_{p,K-p+1}, \tag{12}$$

where $\mathfrak{F}_{p,K-p+1} \sim F_{p,K-p+1}$ and $F_{p,K-p+1}$ is the F distribution with degrees of freedom $(p, K-p+1)$. This is a very convenient result, as the fixed-smoothing asymptotic approximation is a standard distribution. There is no need to simulate critical values. Let $\mathcal{F}_{p,K-p+1}^{1-\alpha}$ be the $1-\alpha$ quantile of the F distribution $F_{p,K-p+1}$, then we can use

$$\mathcal{F}_{IV}^{1-\alpha} = \frac{K}{K-p+1}\mathcal{F}_{p,K-p+1}^{1-\alpha}$$

as the critical value to perform the fixed-smoothing Wald test when an OS HAR variance estimator is used.

Similarly,

$$t_{IV} \xrightarrow{d} t_K,$$

where t_K is the t distribution with degrees of freedom K. We can therefore use the quantile from the t_K distribution to carry out the fixed-smoothing t test.

The testing-optimal choice of K in the OS case is similar to the testing-optimal choice of b is the kernel case. We can first compute the optimal b^* for the following configuration: $q=2$, $c_2=1$, $\rho_d=\pi^2/6$. These are characteristic values associated with the Daniell kernel, the equivalent kernel behind the OS HAR variance estimator using Fourier bases. We then take $K=\left\lceil \frac{1}{bc_2} \right\rceil$. More specifically, we employ the following K value:

$$K_{\text{opt}} = \begin{cases} \left[\frac{\delta^2 G'_{(p+2),\delta^2}\left(\mathcal{X}_p^{1-\alpha}\right)}{4G'_{p,\delta^2}\left(\mathcal{X}_p^{1-\alpha}\right)|\bar{\mathcal{B}}|}\right]^{\frac{1}{3}} T^{\frac{2}{3}}, & \text{if } \bar{\mathcal{B}} > 0 \\ \left[\frac{(\tau-1)\alpha}{G'_p\left(\mathcal{X}_p^{1-\alpha}\right)\mathcal{X}_p^{1-\alpha}|\bar{\mathcal{B}}|}\right]^{1/2} T, & \text{if } \bar{\mathcal{B}} \leq 0. \end{cases}$$

As before, the parameter $\bar{\mathcal{B}}$ is estimated by a standard VAR(1) plug-in procedure. Plugging the estimate of $\bar{\mathcal{B}}$ into K_{opt} yields $\hat{K}_{\text{temp}}$. We truncate $\hat{K}_{\text{temp}}$ to be between $p+4$ and T. That is, we take

$$\tilde{K}_{\text{temp}} = \begin{cases} p+4, & \text{if } \hat{K}_{\text{temp}} \leq p+4 \\ \hat{K}_{\text{temp}}, & \text{if } \hat{K}_{\text{temp}} \in (p+4, T] \\ T, & \text{if } \hat{K}_{\text{temp}} > T \end{cases}.$$

Imposing the lower bound $p+4$ ensures that the variance of the approximating distribution $F_{p,K-p+1}$ is finite and that power loss is not very large. Finally, we round $\tilde{K}_{\text{temp}}$ to the greatest even number less than $\tilde{K}_{\text{temp}}$. We take this greatest even number, denoted by $\hat{K}_{\text{opt}}$, to be our data-driven and testing-optimal choice for K.

2.4 The test procedure

The fixed-smoothing Wald test involves the following steps:

1. Specify the null hypothesis of interest $H_0: R\theta_0 = r$ and the significance level α.

2. Estimate the model using the estimator in (1). Construct

$$\hat{u}_t = R(G_T' W_{0T}^{-1} G_T)^{-1} G_T' W_{0T}^{-1} Z_t'(Y_t - X_t \hat{\theta}_{IV}).$$

3. Fit a VAR(1) model into $\{\hat{u}_t\}$ and obtain a plug-in estimator $\bar{\mathcal{B}}^{est}$. Compute $\hat{b}_{\text{opt}}$ or $\hat{K}_{\text{opt}}$ as described in the previous two subsections.

4. For the kernel case, plug $\hat{b}_{\text{opt}}$ into (8) and (9) to obtain $\hat{K}$ and $\hat{\kappa}$ and compute $\hat{\mathcal{F}}_{IV}^{1-\alpha} = \hat{\kappa}\mathcal{F}_{p,\hat{K}}^{1-\alpha}$. For the OS case, compute

$$\hat{\mathcal{F}}_{IV}^{1-\alpha} = \frac{\hat{K}_{\text{opt}}}{\hat{K}_{\text{opt}} - p + 1} \mathcal{F}_{p,\hat{K}_{\text{opt}}-p+1}^{1-\alpha}.$$

5. Calculate

$$\hat{\Omega} = \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} k\left(\frac{t-s}{\hat{b}T}\right) (\hat{u}_t - \hat{u}^{ave}) (\hat{u}_s - \hat{u}^{ave})', \tag{13}$$

$$\hat{\Omega} = \frac{1}{\hat{K}_{\text{opt}}} \sum_{j=1}^{\hat{K}_{\text{opt}}} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_j\left(\frac{t}{T}\right) \hat{u}_t\right] \left[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \phi_j\left(\frac{s}{T}\right) \hat{u}_s\right]', \tag{14}$$

respectively for the kernel case and the OS case.

6. Construct the test statistic:

$$F_{IV} = \left[\sqrt{T}(R\hat{\theta}_{IV} - r)\right]' \hat{\Omega}^{-1} \left[\sqrt{T}(R\hat{\theta}_{IV} - r)\right] / p. \tag{15}$$

Reject the null if $F_{IV} > \hat{\mathcal{F}}_{IV}^{1-\alpha}$.

We can follow similar steps to perform the fixed-smoothing t test.

To construct two-sided confidence intervals for any individual slope coefficient, we can choose the restriction matrix R to be the selection vector. For example, to select the second element of θ, we can let $R = (0, 1, 0, ..., 0)$. The $100\,(1-\alpha)\,\%$ confidence interval for $R\theta_0$ is

$$\left[R\hat{\theta}_{IV} - \mathsf{t}_{IV}^{1-\alpha/2} \times \sqrt{\hat{\Omega}_R},\ R\hat{\theta}_{IV} + \mathsf{t}_{IV}^{1-\alpha/2} \times \sqrt{\hat{\Omega}_R}\right]$$

where $\mathsf{t}_{IV}^{1-\alpha/2}$ is defined in (10). Here we have added a subscript 'R' to $\hat{\Omega}$ to indicate its dependence on the restriction vector R.

3 Fixed-smoothing Asymptotics: the Two-step GMM

When any element of X_t is endogenous and there are more instruments than the number of regressors, we have an overidentified model. In this case, for efficiency considerations, we may employ a two-step GMM estimator and conduct inferences based on this estimator.

The two-step GMM estimator is given by

$$\begin{aligned}\hat{\theta}_{GMM} &= \arg\min_{\theta\in\Theta} g_T(\theta)' \left[W_T(\hat{\theta}_{IV})\right]^{-1} g_T(\theta) \\ &= \{S'_{ZX}[W_T(\hat{\theta}_{IV})]^{-1}S_{ZX}\}^{-1}\left[S'_{ZX}[W_T(\hat{\theta}_{IV})]^{-1}S_{ZY}\right] \end{aligned} \tag{16}$$

where

$$W_T(\theta) = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} Q_h(\frac{s}{T},\frac{t}{T})\left(v_t(\theta)-\bar{v}(\theta)\right)\left(v_s(\theta)-\bar{v}(\theta)\right)', \tag{17}$$

$v_t(\theta) = Z_t'(Y_t - X_t\theta)$, and $\bar{v}(\theta) = \sum_{t=1}^{T} v_t(\theta)/T.$

Note that $W_T(\hat{\theta}_{IV})$ is an estimator of the long run variance of moment process $\{v_t(\theta_0)\}$. It takes the same form as $\hat{\Omega}$ given in (4) but is based on the (estimated) moment process $\{v_t(\hat{\theta}_{IV})\}$ instead of the (estimated) transformed moment process $\{\hat{u}_t\}$.

The Wald statistic is given by

$$F_{GMM} = \sqrt{T}(R\hat{\theta}_{GMM}-r)'\left\{R[G_T'W_T^{-1}(\hat{\theta}_{GMM})G_T]^{-1}R'\right\}^{-1}\sqrt{T}(R\hat{\theta}_{GMM}-r)/p, \tag{18}$$

and the t statistic is given by

$$t_{GMM} = \frac{\sqrt{T}(R\hat{\theta}_{GMM}-r)}{\sqrt{R[G_T'W_T^{-1}(\hat{\theta}_{GMM})G_T]^{-1}R'}}.$$

Let $B_p(r)$, $B_{d-p}(r)$ and $B_q(r)$ be independent standard Brownian motion processes of dimensions p, $d-p$, and q, respectively. Denote

$$\begin{aligned} C_{pp} &= \int_0^1\int_0^1 Q_h^*(r,s)dB_p(r)dB_p(s)',\ C_{pq} = \int_0^1\int_0^1 Q_h^*(r,s)dB_p(r)dB_q(s)' \\ C_{qq} &= \int_0^1\int_0^1 Q_h^*(r,s)dB_q(r)dB_q(s)',\ D_{pp} = C_{pp} - C_{pq}C_{qq}^{-1}C_{pq}'. \end{aligned} \tag{19}$$

Under some conditions, Sun (2014b) shows that under the fixed-smoothing asymptotics

$$\begin{aligned} F_{GMM} &\to^d \left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]' D_{pp}^{-1}\left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]/p, \\ t_{GMM} &\to^d \frac{\left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]}{\sqrt{D_{pp}}}. \end{aligned}$$

The fixed-smoothing asymptotic distributions are nonstandard in both kernel and OS cases. For the OS case, Hwang and Sun (2017) shows that a modified Wald statistic is asymptotically F distributed and that a modified t statistic is asymptotically t distributed. More specifically, the modified Wald and t statistics are given by

$$F_{GMM}^{c} = \frac{K-p-q+1}{K} \frac{F_{GMM}}{1+\frac{1}{K}J_T}$$
$$t_{GMM}^{c} = \sqrt{\frac{K-q}{K}} \frac{t_{GMM}}{\sqrt{1+\frac{1}{K}J_T}}$$

where

$$J_T = T g_T(\hat{\theta}_{GMM})' \left[W_T(\hat{\theta}_{GMM}) \right]^{-1} g_T(\hat{\theta}_{GMM})$$

is the usual J statistic for testing overidentification restrictions. It is shown that

$$F_{GMM}^{c} \to^{d} F_{p,K-p-q+1} \text{ and } t_{GMM}^{c} \to^{d} t_{K-q}.$$

So, we can employ

$$\left[1+\frac{1}{K}J_T\right] \left[\frac{K}{K-p-q+1}\right] \mathcal{F}_{p,K-p-q+1}^{1-\alpha}$$

as the critical value for the original Wald statistic F_{GMM} and

$$\sqrt{1+\frac{1}{K}J_T}\sqrt{\frac{K}{K-q}}\mathrm{t}_{K-q}^{1-\frac{\alpha}{2}}$$

as the critical value for the t statistic $|t_{GMM}|$. As in the case with the first-step GMM, as long as the OS HAR variance estimator is used, there is no need to simulate any critical value.

We note that Sun and Kim (2012) establish that the modified J statistic is asymptotically F distribution:

$$J_T^c := \frac{K-q+1}{qK} J_T \to^{d} F\left(q, K-q+1\right).$$

For the two-step GMM with an estimated weighting matrix, a testing-optimal choice of K has not been established in the literature, but see Sun and Phillips (2009) for a suggestion for the smoothing parameter choice that is oriented towards interval estimation. For practical implementations, Hwang and Sun (2017) suggest selecting K based on the conventional AMSE criterion implemented by using the VAR(1) plug-in procedure. More specifically,

$$\hat{K}_{\text{tmp}} = \left\lceil \left(\frac{tr(I_{m^2}+\mathbb{K}_{mm})(\hat{\Omega}_v \otimes \hat{\Omega}_v)}{4vec\left(\hat{\mathcal{B}}_v\right)' vec\left(\hat{\mathcal{B}}_v\right)} \right)^{1/5} T^{4/5} \right\rceil,$$

where $\mathbb{K}_{mm}$ is the $m^2 \times m^2$ commutation matrix and I_{m^2} is the $m^2 \times m^2$ identity matrix. In the above, $\hat{\mathcal{B}}_v$ is the plug-in estimator of

$$\mathcal{B}_v = -\frac{\pi^2}{6} \sum_{j=-\infty}^{\infty} j^2 E v_t v'_{t-j}$$

and $\hat{\Omega}_v$ is the plug-in estimator of the LRV of $\{v_t\}$. The formulae for $\hat{\Omega}_v$ and $\hat{\mathcal{B}}_v$ in terms of the estimated VAR(1) matrix and the error variance are available from Andrews (1991). We then obtain $\tilde{K}_{\text{tmp}}$ by truncating $\hat{K}_{\text{tmp}}$ to be between $p+q+4$ and T. Finally, we round $\tilde{K}_{\text{tmp}}$ to $\hat{K}_{\text{mse}}$, the greatest even number less than $\tilde{K}_{\text{tmp}}$ and use $\hat{K}_{\text{mse}}$ throughout the two-step procedure.

To conduct the two-step fixed-smoothing Wald test, we follow the steps below:

1. Specify the null hypothesis of interest $H_0 : R\theta_0 = r$ and the significance level α.
2. Estimate θ_0 by the IV estimator and construct $\hat{v}_t = Z'_t(Y_t - X_t\hat{\theta}_{IV})$.
3. Fit a VAR (1) model into $\{\hat{v}_t\}$ and compute the data-driven choice $\hat{K}_{\text{mse}}$.
4. On the basis of $\hat{K}_{\text{mse}}$, construct the weighting matrix $\hat{W}_T = W_T(\hat{\theta}_{IV})$ in (17).
5. Estimate θ_0 by
$$\hat{\theta}_{GMM} = \left[S'_{ZX}\hat{W}_T^{-1}S_{ZX}\right]^{-1}\left[S'_{ZX}\hat{W}_T^{-1}S_{ZY}\right]. \tag{20}$$
6. Calculate the test statistic F_{GMM} defined in (18) and the critical value
$$\hat{\mathcal{F}}^{1-\alpha}_{GMM} = \left[1 + \frac{1}{\hat{K}_{\text{mse}}} J_T(\hat{\theta}_{GMM})\right]\left[\frac{\hat{K}_{\text{mse}}}{\hat{K}_{\text{mse}} - p - q + 1}\right]\mathcal{F}^{1-\alpha}_{p,\hat{K}_{\text{mse}}-p-q+1}$$
7. If $F_{GMM} > \hat{\mathcal{F}}^{1-\alpha}_{GMM}$, then we reject the null. Otherwise, we fail to reject the null.

With some simple modifications, the above steps can be followed to perform the fixed-smoothing t test.

Following the same procedure as in the IV case, we can employ the two-step GMM estimator and construct the associated $100\,(1-\alpha)\,\%$ confidence interval for $R\theta_0$ as

$$R\hat{\theta}_{GMM} \pm \mathfrak{t}^{1-\alpha/2}_{GMM} \times \sqrt{R[G'_T W_T^{-1}(\hat{\theta}_{GMM})G_T]R'}.$$

where

$$\mathfrak{t}^{1-\alpha/2}_{GMM} = \sqrt{1 + \frac{1}{K}J_T}\sqrt{\frac{K}{K-q}}\mathfrak{t}^{1-\alpha/2}_{K-q}$$

4 The har command

4.1 Stata syntax

har *depvar* [*varlist*1] (*varlist*2 = *instlist*) [*if*] [*in*] *kernel(string)* [, noconstant level(#)]

4.2 Options

level(#) specifies the confidence level, as a percentage, for confidence intervals. The default is level(95).

kernel(*string*) set the type of kernel. For the Bartlett kernel, any of the four usages — kernel(bartlett), kernel(BARTLETT), kernel(B), or kernel (b) — produces the same results. Similarly, for the Parzen, QS, and orthonormal series LRV estimators, we can use any of the respective choices: (PARZEN, parzen, P, p), (QUADRATIC, quadratic, Q, q), and (ORTHOSERIES, orthoseries, O, o). kernel() is required.

noconstant suppresses the constant term.

You must `tsset` your data before using `har`, see [TS] `tsset`.

Time series operators are allowed.

4.3 Saved results

The `har` uses `ivregress` to get the estimates of the model parameters. In addition to the standard stored results from `ivregress`, `har` also stores the following results in `e()`:

Scalars

e(N)	number of observations	∗e(sF)	adjusted F statistic
∗e(ssdf)	second degrees of freedom	∗e(kopt)	data-driven optimal K of orthonormal bases
⋆e(kF)	adjusted F statistic	⋆e(ksdf)	second degrees of freedom
⋆e(lag)	data-driven truncation lag	∗⋆ e(fdf)	first degrees of freedom

Macros

e(cmd)	har	e(cmdline)	command as typed
e(carg)	nocons or " " if specified	e(varline)	variable line as typed
e(depvar)	name of dependent variable	e(title)	title in the estimation output
e(kerneltype)	kernel in the estimation	e(vcetype)	title used to label Std. Err.

Matrices

e(b)	coefficient vector	∗e(sstderr)	adjusted std error for each individual coefficient
∗e(sdf)	degrees of freedom of t statistic	∗e(st)	the t statistic
∗e(sbetahat)	the IV coefficient vector	⋆e(kbetahat)	the IV coefficient vector
⋆e(kstderr)	adjusted std error for each individual coefficient	⋆e(kdf)	degrees of freedom of the t statistic
⋆e(kt)	t statistic		

Functions

e(sample)	marks the estimation sample

note: ∗ for orthonormal series; ⋆ for Bartlett, Parzen, QS kernels

We use the time series data downloaded from Stata's official website http://www.stata-press.com/data/r14/idle2.dta to illustrate the use of `har`. We illustrate `har` by analyzing the influence of *idle* and *wio* on *usr*. The data consists of time series of 30 observations covering the periods from 08:20 to 18:00. We have to 'tsset' the dataset before using the command `har`.

Case 1: nonparametric Bartlett kernel approach, default confidence level 95%, testing-optimal automatic bandwidth selection.

```
. webuse idle2

. tsset time
        time variable:  time, 1 to 30
                delta:  1 unit

. har usr idle wio, kernel(bartlett)

Regression with HAR standard errors                 Number of obs  =        30
Kernel: Bartlett                                    F(  2,    17)  =     47.66
Data-driven optimal lag: 2                          Prob > F       =    0.0000

------------------------------------------------------------------------------------
                                 HAR
          usr       Coef.     Std.Err.       t       df   P>|t|     [95% Conf. Interval]
------------------------------------------------------------------------------------
         idle   -.6670978    .0715786    -9.32       22   0.000    -.8155428   -.5186529
          wio   -.7792461      .11897    -6.55       13   0.000    -1.036265    -.522227
        _cons    66.21805    6.984346     9.48       19   0.000     51.59965    80.83646
------------------------------------------------------------------------------------
```

The header consists of the kernel type, the data-driven testing-optimal truncation lag, and the F statistic for the Wald test. The column title of the matrix reports coefficients, HAR standard errors, t statistics, the equivalent degrees of freedom, p-values, and confidence intervals. Each covariate is associated with its own asymptotic t distribution. This is different from the regular Stata commands `regress` and `newey` where a single standard normal distribution is used. The reason is that the testing-optimal smoothing parameter b depends on the null restriction vector R. Each model parameter corresponds to a different vector R and hence a different data-driven b and a different t approximation.

Case 2: nonparametric Bartlett kernel approach, confidence level 99%, testing-optimal automatic bandwidth selection, noconstant.

```
. webuse idle2

. tsset time
        time variable:  time, 1 to 30
                delta:  1 unit

. har usr idle wio, kernel(bartlett) l(99) nocons

Regression with HAR standard errors                 Number of obs  =        30
Kernel: Bartlett                                    F(  2,     3)  =      8.88
Data-driven optimal lag: 13                         Prob > F       =    0.0549
```

usr	Coef.	HAR Std.Err.	t	df	P>|t|	[99% Conf.	Interval]
idle	.0186886	.0101968	1.83	5	0.126	-.0224265	.0598037
wio	.2759991	.0954198	2.89	5	0.034	-.1087473	.6607454

Case 3: nonparametric Parzen kernel approach, confidence level 95%, testing-optimal automatic bandwidth selection.

```
. webuse idle2

. tsset time
        time variable:  time, 1 to 30
                delta:  1 unit

. har usr idle wio, kernel(parzen)

Regression with HAR standard errors                 Number of obs  =        30
Kernel: Parzen                                      F(  2,     4)  =     50.87
Data-driven optimal lag: 10                         Prob > F       =    0.0014
```

usr	Coef.	HAR Std.Err.	t	df	P>|t|	[95% Conf.	Interval]
idle	-.6670978	.071317	-9.35	15	0.000	-.8191065	-.5150892
wio	-.7792461	.1143269	-6.82	12	0.000	-1.028343	-.5301492
_cons	66.21805	6.922399	9.57	14	0.000	51.37099	81.06512

Case 4: nonparametric Quadratic Spectral kernel approach, confidence level 95%, testing-optimal automatic bandwidth selection.

```
. webuse idle2

. tsset time
        time variable:  time, 1 to 30
                delta:  1 unit

. har usr idle wio, kernel(quadratic)

Regression with HAR standard errors                 Number of obs  =        30
Kernel: Quadratic Spectral                          F(  2,     4)  =     46.84
Data-driven optimal lag: 5                          Prob > F       =    0.0017
```

	HAR

usr	Coef.	Std.Err.	t	df	P>|t|	[95% Conf.	Interval]
idle	-.6670978	.0697384	-9.57	16	0.000	-.8149366	-.5192591
wio	-.7792461	.1131035	-6.89	13	0.000	-1.023591	-.5349009
_cons	66.21805	6.834698	9.69	15	0.000	51.65024	80.78587

Case 5: nonparametric orthonormal series approach, confidence level 95%, testing-optimal automatic bandwidth selection.

```
. webuse idle2

. tsset time
        time variable:  time, 1 to 30
                delta:  1 unit

. har usr idle wio, kernel(orthoseries)

Regression with HAR standard errors                  Number of obs  =        30
Kernel: Orthonormal Series                           F(  2,     5)  =     43.17
Data-driven optimal K: 6                             Prob > F       =    0.0007
```

usr	Coef.	HAR Std.Err.	t	df	P>|t|	[95% Conf.	Interval]
idle	-.6670978	.0706388	-9.44	14	0.000	-.8186029	-.5155927
wio	-.7792461	.1122118	-6.94	12	0.000	-1.023735	-.5347576
_cons	66.21805	6.838414	9.68	14	0.000	51.55111	80.88499

In this case, the header reports data-driven testing-optimal K. This is different from the nonparametric kernel approach.

5 The gmmhar command

5.1 Stata syntax

gmmhar *depvar* [*varlist1*] (*varlist2* = *instlist*) [*if*] [*in*] [, noconstant level(#)]

5.2 Options

level(#) specifies the confidence level, as a percentage, for confidence intervals. The default is level(95).

noconstant suppress constant term.

You must `tsset` your data before using `gmmhar`, see [TS] `tsset`.

Time-series operators are allowed.

5.3 Saved results

The `gmmhar` uses `ivregress` to get the colname in `e(b)` for the output table in gmmhar_tab.ado. In addition to the standard stored results from `ivregress`, `gmmhar` also stores the following results in `e()`:

Scalars

`e(N)`	number of observations	`e(sF)`	adjusted F statistic
`e(sfdf)`	first degrees of freedom	`e(ssdf)`	second degrees of freedom
`e(kopt)`	data-driven optimal K for the OS variance estimator	`e(J)`	J statistic for testing the over identification

Macros

`e(cmd)`	`gmmhar`	`e(cmdline)`	command as typed
`e(carg)`	nocons or " " if specified	`e(varline)`	variable line as typed
`e(vcetype)`	orthonormal series	`e(title)`	title in estimation output
`e(depvar)`	name of the dependent variable	`e(exog)`	exogenous variables
`e(endog)`	endogenous variables	`e(inst)`	instrument variables

Matrices

`e(betahat)`	two-step gmm coefficient vector	`e(sstderr)`	adjusted std error for each individual coefficient
`e(sdf)`	degrees of freedom of the t statistic	`e(st)`	t statistic

Functions

`e(sample)`	marks the estimation sample

5.4 Examples

To illustrate the use of `gmmhar` in the two-step GMM framework, we estimate a quarterly time-series model relating the change in the U.S. inflation rate (D.inf) to the unemployment rate (UR) for 1959q1-2000q4. As instruments, we use the second lag of quarterly GDP growth, the lagged values of the Treasury bill rate, the trade-weighted exchange rate, and the Treasury medium-term bond rate. We fit our model using the two-step efficient GMM method.

Case 6: nonparametric orthonormal series approach, confidence level 95%, AMSE automatic bandwidth selection.

```
. use http://fmwww.bc.edu/ec-p/data/stockwatson/macrodat
. generate inf =100 * log( CPI / L4.CPI )
(4 missing values generated)
. generate ggdp=100 * log( GDP / L4.GDP )
(10 missing values generated)
. gmmhar D.inf (UR=L2.ggdp L.TBILL L.ER L.TBON)
Two-step Efficient GMM Estimation                    Number of obs   =       158
Data-driven optimal K: 46                            F(  1,    43)   =      2.05
                                                     Prob > F        =    0.1597
```

D.inf	Coef.	HAR std.Err.	t	df	P>|t|	[95% Conf.	Interval]
UR	-.0971458	.067901	-1.43	43	0.160	-.2340812	.0397895
_cons	.5631061	.3936908	1.43	43	0.160	-.2308471	1.357059

```
HAR J statistic  =  .92614349
Reference Dist for the J test: F(  3,    44)
P-value of the J test =    0.4361
Instrumented: UR
Instruments: L2.ggdp L.TBILL L.ER L.TBON
```

In this case, the header reports the data-driven K value by the AMSE method. In the above table, the negative coefficient on the unemployment rate is consistent with the basic macroeconomic theory: lowering unemployment below the natural rate will cause an acceleration of price inflation. The fixed-smoothing J test is now far from rejecting the null, giving us greater confidence that our instrument set is appropriate.

Case 7: nonparametric orthonormal series approach, noconstant, confidence level 99%, AMSE automatic bandwidth selection.

```
. use http://fmwww.bc.edu/ec-p/data/stockwatson/macrodat
. generate inf =100 * log( CPI / L4.CPI )
(4 missing values generated)
. generate ggdp=100 * log( GDP / L4.GDP )
(10 missing values generated)
. gmmhar D.inf (UR=L2.ggdp L.TBILL L.ER L.TBON),nocons l(99)
Two-step Efficient GMM Estimation                  Number of obs  =       158
Data-driven optimal K: 40                          F(  1,    37)  =      0.01
                                                   Prob > F       =    0.9119
```

D.inf	Coef.	HAR std.Err.	t	df	P>|t|	[99% Conf.	Interval]
UR	.0014583	.0130865	0.11	37	0.912	-.0340768	.0369934

```
HAR J statistic  =  .95768181
Reference Dist for the J test: F(  3,    38)
P-value of the J test =    0.4226
Instrumented: UR
Instruments: L2.ggdp L.TBILL L.ER L.TBON
```

6 Monte Carlo Evidence

In this section, we use the commands `har` and `gmmhar` to evaluate the coverage accuracy of the 95% confidence intervals based on the fixed-smoothing asymptotic approximations. If the coverage rate, i.e., the percentage of confidence intervals in repeated experiments that contain the true value, is close to 95%, the nominal coverage probability, then the confidence intervals so constructed have accurate coverage, and the asymptotic approximations are reliable in finite samples. For comparison, we include the results from the commands `newey` and `ivregress` in our report.

6.1 Specifications

DGP for `har`

We consider the data generating process

$$y_t = x_{0,t}\gamma + x_{1,t}\beta_1 + x_{2,t}\beta_2 + \varepsilon_t, \tag{21}$$

where $x_{0,t} \equiv 1$ and $x_{1,t}$, $x_{2,t}$ and ε_t follow independent AR(1) processes:

$$x_{j,t} = \rho x_{j,t-1} + \sqrt{1-\rho^2} e_{j,t}, j = 1, 2;\ \varepsilon_t = \rho \varepsilon_{t-1} + \sqrt{1-\rho^2} e_{0,t},$$

or MA(1) processes:

$$x_{j,t} = \rho e_{j,t-1} + \sqrt{1-\rho^2} e_{j,t}, j = 1, 2;\ \varepsilon_t = \rho e_{t-1,0} + \sqrt{1-\rho^2} e_{t,0}.$$

The error term $e_{j,t} \sim iidN(0,1)$ across j and t. In the AR case, the processes are initialized at zero. We consider $\rho = 0.25, 0.5, 0.75$.

DGP for `gmmhar`

We follow Hwang and Sun (2017) and consider a linear model of the form:

$$y_t = x_{0,t}\gamma + x_{1,t}\beta_1 + x_{2,t}\beta_2 + \varepsilon_{y,t}, \tag{22}$$

where $x_{0,t} \equiv 1$ and $x_{1,t}, x_{2,t}$ are scalar endogenous regressors. The unknown parameter vector is $\theta = (\gamma, \beta_1, \beta_2)' \in \mathbb{R}^3$. We have m instruments $z_{0,t}, z_{1,t}, z_{2,t}, \cdots, z_{m-1,t}$ with $z_{0,t} \equiv 1$. The reduced form equations for $x_{1,t}$ and $x_{2,t}$ are given by

$$x_{j,t} = z_{j,t} + \sum_{i=d}^{m-1} z_{i,t} + \varepsilon_{x_j,t}, j = 1, 2.$$

We consider two different experiment designs: the autoregressive (AR) design and the centered moving average (CMA) design. In the AR design, each $z_{i,t}$ follows an AR(1) process of the form

$$z_{i,t} = \rho z_{i,t-1} + \sqrt{1-\rho^2} e_{z_i,t} \text{ for } i = 1, 2, ..., m$$

where $e_{z_i,t} = \left(e^i_{zt} + e^0_{zt}\right)/\sqrt{2}$ and $e_t = [e^0_{zt}, e^1_{zt}, \ldots, e^{m-1}_{zt}]' \sim iidN(0, I_m)$. By construction, each non-constant z_{it} has unit variance and the correlation coefficient between the non-constant $z_{i,t}$ and $z_{j,t}$ for $i \neq j$ is 0.5. The GDP for $\varepsilon_t = (\varepsilon_{y,t}, \varepsilon_{x_1,t}, \varepsilon_{x_2,t})$ is the same as that for $(z_{1,t}, \cdots, z_{m-1,t})$ except for the dimensional difference. We take $\rho = -0.5, 0.5, 0.8$.

In the CMA design, $\varepsilon_{y,t}$ is a scaled and centered moving average of an iid sequence $\varepsilon_{y,t} = \sum_{j=-L}^{j=L} e_{t+j}/\sqrt{2L+1}$, where $e_t \sim iidN(0,1)$ and L is the number of leads and lags in the average. The instruments are generated according to

$$z_{i,t} = \left[e_{t-L+i-1} - (2L+1)^{-1} \sum_{j=-L}^{L} e_{t+j} \right] \sqrt{(2L+1)/2L}, i = 1, 2, \cdots, m-1.$$

The error term in the reduced form equation is given by $\varepsilon_{x_j,t} = \left(\varepsilon_{y,t} + e_{x_j,t}\right)/\sqrt{2}$, where $e_{x_j,t} \sim iidN(0,1)$ and is independent of the sequence $\{e_t\}$. We take $L = 3$. The number of moment conditions is set to be $m = 3, 4, 5$ with the corresponding degrees of overidentification being $q = 0, 1, 2$. We consider the sample size $T = 100$ and the significance level 5%. Throughout we are concerned with testing the slope coefficients β_1 and β_2. We employ the HAR variance estimators based on the Bartlett, Parzen, QS kernels and orthonormal Fourier series. The number of simulation replications is 1000.

In both sets of Monte Carlo experiments, we set $\gamma = 1$, $\beta_1 = 3$, $\beta_2 = 2$ without loss of generality.

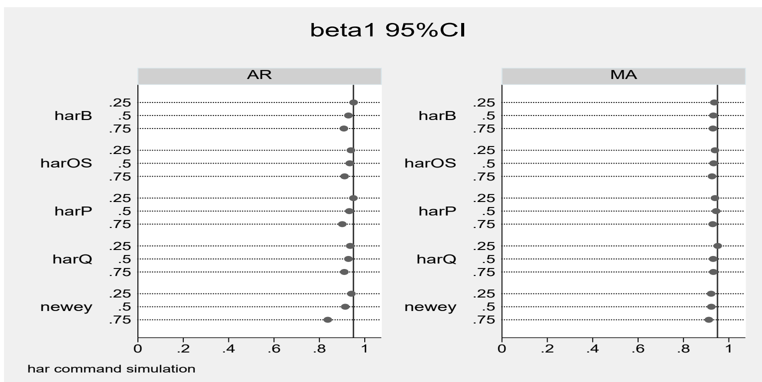

Figure 1: Empirical coverage rates of 95% confidence intervals of β_1 in model (21): the y-labels 0.25, 0.5, and 0.75 indicate the AR or MA parameter

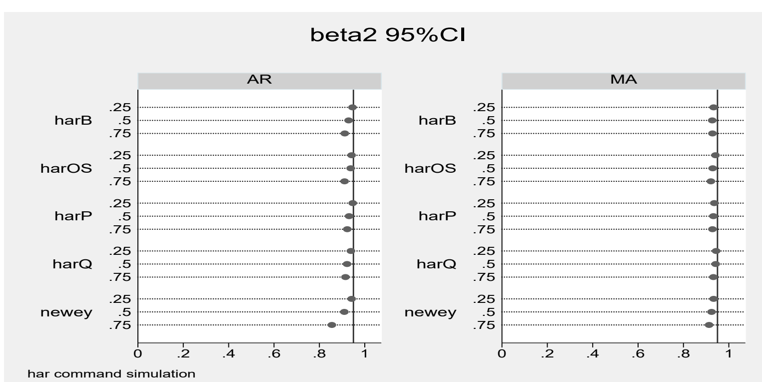

Figure 2: Empirical coverage rates of 95% confidence intervals of β_2 in model (21): the y-labels 0.25, 0.5 and 0.75 denote the AR or MA parameter

6.2 Results

Figures 1 and 2 report the empirical coverage rates of the 95% confidence intervals for β_1 and β_2, respectively. The results are based on the command `har` applied to the data generated by the model in (21). It is clear from these two figures that confidence intervals based on the fixed-smoothing approximations have more accurate coverage than those based on the normal approximation, which is adopted in the command `newey`. As ρ increases, coverage accuracy deteriorates in each case. When ρ is equal to 0.75, the confidence intervals based on the fixed-smoothing asymptotic approximations are still reasonably accurate. In contrast, confidence intervals produced by `newey` undercover the true value substantially.

Figures 3-6 report the simulation results based on the commands `gmmhar` and `ivregress gmm` for the IV regression. For the `ivregress` command, the weighting matrix is based on the option "`wmatrix (hac kernel opt (#))`", that is, the weighting matrix is based on a kernel HAC estimator using the data-driven truncation lag proposed by Newey and West (1994).

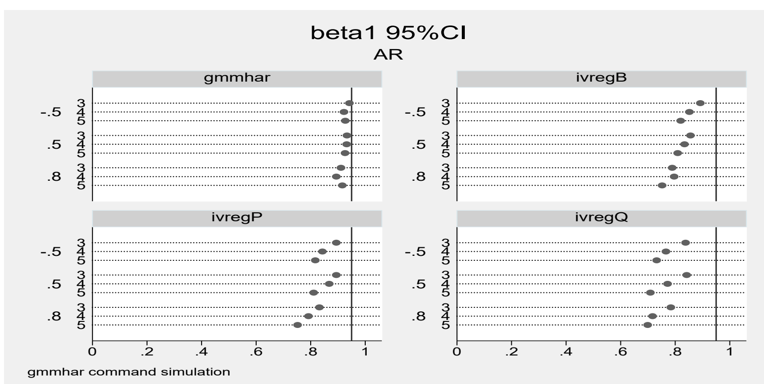

Figure 3: Empirical coverage rates of 95% confidence intervals of β_1 in model (22): the y-labels -0.5, 0.5 and 0.8 indicate the values of the AR parameter, and the y-sublables 3, 4, and 5 indicate the number of instruments used

Figures 3 and 4 contain the results for the AR design, and Figures 5 and 6 contain the results for the CMA design. Under both designs, the confidence intervals based the command `gmmhar`, which uses the fixed-smoothing t approximations, are much more accurate than those based on the command `ivregress gmm`, which uses the normal approximation. Under both designs, the coverage accuracy of the confidence intervals produced by `ivregress gmm` deteriorates quickly as the number of instruments increases. This is especially true when the Parzen and QS kernels are used. In contrast, the coverage accuracy of the confidence intervals produced by `gmmhar` does not appear to be affected by the number of instruments.

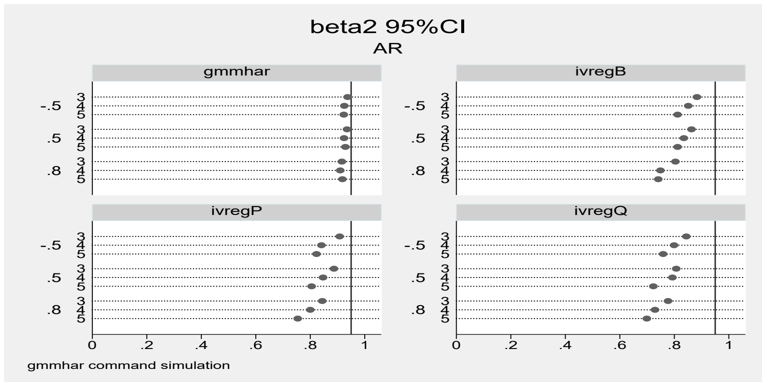

Figure 4: Empirical coverage rates of 95% confidence intervals of β_2 in model (22): the y-labels -0.5, 0.5 and 0.8 indicate the values of the AR parameter, and the y-sublables 3, 4, and 5 indicate the number of instruments used

7 The `hart` and `gmmhart` commands

`hart` and `gmmhart` are the post-estimation commands that should be used immediately after the respective estimation commands `har` and `gmmhar`. These two commands perform the Wald type of tests but employ more accurate fixed-smoothing critical values. The test statistics are given in (2) and (18), respectively.

7.1 Stata syntax

The syntaxes of `hart` and `gmmhart` are as follows:

`hart coeflist, kernel(`*string*`)` [`, accumulate level(#)`] `syntax 1`

`hart exp=`*exp*[`=...`]`, kernel(`*string*`)` [`, accumulate level(#)`] `syntax 2`

`gmmhart coeflist,` [`, accumulate`] `syntax 1`

`gmmhart exp=`*exp*[`=...`]`,` [`, accumulate`] `syntax 2`

Syntax 1 tests that the listed coefficient are jointly 0;

Syntax 2 tests a single or multiple linear restrictions.

`hart` implements the test described in Section 2 for testing the null $H_0 : R\theta_0 = r$ against the alternative $H_1 : R\theta_0 \neq r$.

`gmmhart` implements the test described in Section 3 for the same null and alternative

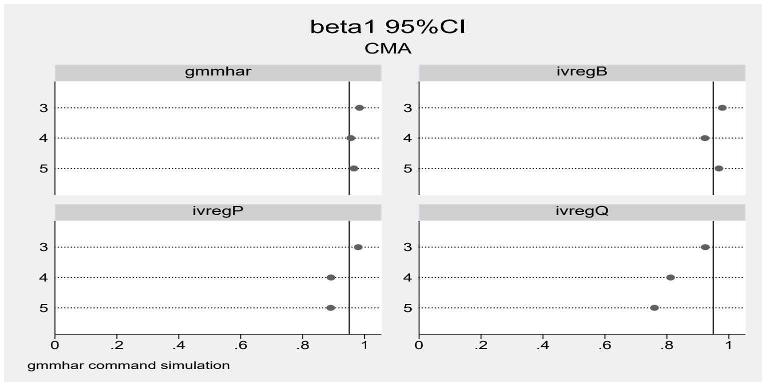

Figure 5: Empirical coverage rates of 95% confidence intervals of β_1 in model (22): the y-labels $3, 4$, and 5 indicate the number of instruments used.

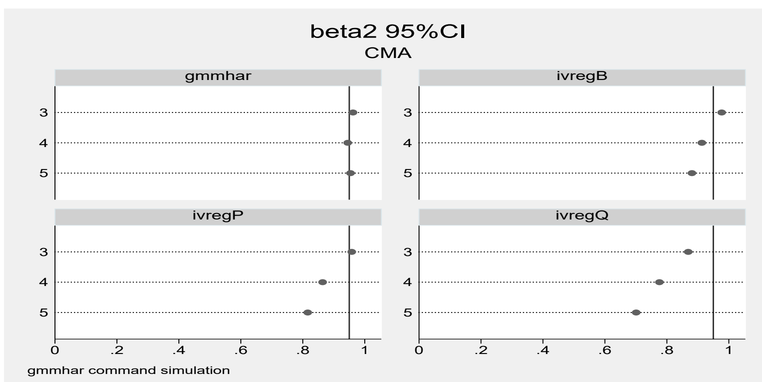

Figure 6: Empirical coverage rates of 95% confidence intervals of β_2 in model (22): the y-labels $3, 4$, and 5 indicate the number of instruments used.

hypotheses.

The options kernel(*string*) and level(#) in `hart` must be consistent with those in `har`.

7.2 Options

Three options are available to the `hart` command.

level(#) sets the confidence level $1 - \alpha$ (or the significance level α). The default is level(95), which corresponds to confidence level 95% and significance level 5%.

accumulate tests the hypothesis jointly with previously tested hypotheses.

kernel(*string*) sets the type of kernel.

7.3 Saved results

The commands `hart` and `gmmhart` store the following in `r()`:

Scalars

⋆∗r(firdf)	the first degrees of freedom	⋆∗r(secdf)	the second degrees of freedom
⋆∗r(kopt)	the datadriven optimal K	⋆r(lag)	the datadriven optimal truncation lag
⋆∗r(F)	the adjusted F statistic		

Matrices

⋆r(thetaiv)	the iv coefficient vector	∗r(thetagmm)	the two-step gmm coefficient vector

⋆ for hart; ∗ for gmmhart.

7.4 Examples

We provide some examples to illustrate the use of `hart` and `gmmhart`. We will use the data in Section 4 for `hart` and data in Section 5 for `gmmhart`.

Case 8: We use `hart` to test different null hypotheses based on the Bartlett kernel. The first two commands test that the coefficients on *idle* and *wio* are jointly zero. These two commands produce numerically identical results. The last command tests the null that the coefficient for *wio* is equal to 1.168 times the coefficient for *idle.*

```
. webuse idle2

. tsset time
        time variable:  time, 1 to 30
                delta:  1 unit

. har usr idle wio, kernel(bartlett)

Regression with HAR standard errors                 Number of obs  =        30
Kernel: Bartlett                                    F(  2,    17)  =     47.66
Data-driven optimal lag: 2                          Prob > F       =    0.0000

------------------------------------------------------------------------------
             |                HAR
         usr |      Coef.   Std.Err.      t        df   P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        idle |  -.6670978   .0715786    -9.32      22   0.000    -.8155428   -.5186529
         wio |  -.7792461     .11897    -6.55      13   0.000    -1.036265    -.522227
       _cons |   66.21805   6.984346     9.48      19   0.000     51.59965    80.83646
------------------------------------------------------------------------------

. hart idle=wio=0, kernel(bartlett)
       F(  2,    17) =   47.66
            Prob > F =    0.0000

. qui hart idle=0, kernel(bartlett)

.     hart idle=wio, kernel(bartlett) acc
       F(  2,    17) =   47.66
            Prob > F =    0.0000

.     hart 1.168*idle=wio, kernel(bartlett)
       F(  1,    14) =    0.00
```

```
             Prob > F =    0.9989
```

Case 9: We use `hart` to test that the coefficients on *idle* and *wio* are jointly zero again, but now we employ the orthonormal series LRV estimator.

```
. webuse idle2

. tsset time
        time variable:  time, 1 to 30
                delta:  1 unit

.     har usr idle wio, kernel(O)

Regression with HAR standard errors                  Number of obs  =        30
Kernel: Orthonormal Series                           F(  2,     5)  =     43.17
Data-driven optimal K: 6                             Prob > F       =    0.0007

------------------------------------------------------------------------------
             |                HAR
         usr |      Coef.   Std.Err.       t        df   P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        idle |  -.6670978   .0706388    -9.44       14   0.000    -.8186029   -.5155927
         wio |  -.7792461   .1122118    -6.94       12   0.000    -1.023735   -.5347576
       _cons |   66.21805   6.838414     9.68       14   0.000     51.55111    80.88499
------------------------------------------------------------------------------

.     hart (idle=0) (wio=0), kernel(O)
       F(  2,     5) =   43.17
             Prob > F =    0.0007

.
```

Case 10: The case is the same as Case 9, but no constant is included in the `har` regression.

```
. webuse idle2

. tsset time
        time variable:  time, 1 to 30
                delta:  1 unit

.     har usr idle wio, kernel(o) nocons

Regression with HAR standard errors                  Number of obs  =        30
Kernel: Orthonormal Series                           F(  2,     5)  =     12.00
Data-driven optimal K: 6                             Prob > F       =    0.0123

------------------------------------------------------------------------------
             |                HAR
         usr |      Coef.   Std.Err.       t        df   P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        idle |   .0186886   .0084701     2.21        8   0.058    -.0008434    .0382206
         wio |   .2759991   .1206479     2.29        8   0.051    -.0022156    .5542137
------------------------------------------------------------------------------

.     hart idle wio, kernel(o)
       F(  2,     5) =   12.00
             Prob > F =    0.0123

.
```

Case 11: We use `gmmhart` to test three different hypotheses based on the two-step GMM estimator with the (inverse) weighting matrix estimated by the orthonormal series approach.

1. Test that the coefficient on UR is 0;
2. Test that the coefficient on UR is 0 again but with a shorter command;
3. Test that the coefficient on UR is -0.09715.

```
. use http://fmwww.bc.edu/ec-p/data/stockwatson/macrodat
. generate inf =100 * log( CPI / L4.CPI )
(4 missing values generated)
. generate ggdp=100 * log( GDP / L4.GDP )
(10 missing values generated)
. gmmhar D.inf (UR=L2.ggdp L.TBILL L.ER L.TBON)
Two-step Efficient GMM Estimation                    Number of obs  =       158
Data-driven optimal K: 46                            F(  1,    43)  =      2.05
                                                     Prob > F       =    0.1597
```

D.inf	Coef.	HAR std.Err.	t	df	P>|t|	[95% Conf.	Interval]
UR	-.0971458	.067901	-1.43	43	0.160	-.2340812	.0397895
_cons	.5631061	.3936908	1.43	43	0.160	-.2308471	1.357059

```
HAR J statistic  =  .92614349
Reference Dist for the J test: F(  3,    44)
P-value of the J test =    0.4361
Instrumented: UR
Instruments: L2.ggdp L.TBILL L.ER L.TBON
. gmmhart UR=0
       F(  1,    43) =    2.05
            Prob > F =    0.1597
. gmmhart UR
       F(  1,    43) =    2.05
            Prob > F =    0.1597
. gmmhart UR=-0.09715
       F(  1,    43) =    0.00
            Prob > F =    1.0000
```

8 Conclusion

In this article, we present the new estimation command `har` and the post-estimation test command `hart` in Stata. These commands extend the existing commands for linear regression models with time series data. We use the more accurate fixed-smoothing asymptotic approximations to construct the confidence intervals and conduct various tests. For the OLS and IV regressions, there are two main differences between the tests based on the new commands `har`/`hart` and the tests based on the Stata commands `newey`/`test`. First, the bandwidth parameter is selected in different ways. While `newey`/`test` use a single data-driven smoothing parameter for all tests, `har`/`hart` use different smoothing parameters for different tests. The smoothing parameter behind `har`/`hart` is tailored towards each test or parameter under consideration. Second, for the case with a single restriction, `newey` uses the standard normal approximation while `har` uses a t approximation. For joint tests with more than one restrictions, `newey/test` use a chi-squared approximation while `har/hart` use an F approximation.

We also introduce another new pair of Stata commands `gmmhar/gmmhart` to be used in an over-identified linear IV regression. In this case, the efficient estimator minimizes a GMM criterion function that uses a long run variance estimator as the weighting matrix. So the underlying nonparametric LRV estimator plays two different roles: it is a part of the GMM criterion function and a part of the asymptotic variance estimator. Recent research has established more accurate distributional approximations that account for the estimation uncertainty in the LRV estimator in both occurrences. Given that the new approximations are less convenient when a kernel LRV estimator is used, we recommend using an OS LRV estimator, in which case the modified F and t statistics converge to standard F and t distributions, respectively.

The Monte Carlo evidence shows that the fixed-smoothing confidence intervals are more accurate than the conventional confidence intervals. The simulation results produced by the commands `har` and `gmmhar` are consistent with those produced by the authors using Matlab.

Both estimation commands `har` and `gmmhar` and the corresponding post-estimation test commands are designed for linear OLS or IV regressions. While these two regressions are most popular in practice, it is worthwhile to update the commands in the future so that they can accommodate general GMM models with nonlinear moment conditions.

9 Acknowledgements

We thank David M. Drukker, the Executive Director of Econometrics at StataCorp, for feedback and encouragement. We also thank Tim Vogelsang whose Stata codes at https://msu.edu/˜tjv/working.html provide partial motivation for this article.

10 References

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858.

Bester, A., Conley, T., Hansen, C., and Vogelsang, T. (2016). Fixed-b asymptotics for spatially dependent robust nonparametric covariance matrix estimators. *Econometric Theory*, 32:154–186.

Gonçlaves, S. (2011). The moving blocks bootstrap for panel linear regression models with individual fixed effects. *Econometric Theory*, 27:1048–1082.

Gonçlaves, S. and Vogelsang, T. (2011). Block bootstrap HAC robust tests: The sophistication of the naive bootstrap. *Econometric Theory*, 27:745–791.

Hwang, J. and Sun, Y. (2017). Asymptotic F and t tests in an efficient GMM setting. *Journal of Econometrics*, 198:277–295.

Jansson, M. (2004). On the error of rejection probability in simple autocorrelation robust tests. *Econometrica*, 72:937–946.

Kiefer, N. and Vogelsang, T. (2002a). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory*, 18:1350–1366.

Kiefer, N. and Vogelsang, T. (2002b). heteroskedasticity-autocorrelation robust standard errors using the bartlett kernel without truncation. *Econometrica*, 70:2093–2095.

Kiefer, N. and Vogelsang, T. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21:1130–1164.

Kim, M. and Sun, Y. (2013). Heteroskedasticity and spatiotemporal dependence robust inference for linear panel models with fixed effects. *Journal of Econometrics*, 177:85–108.

Lazarus, E., Lewis, D., Stock, J., and Watson, M. (2016). HAR inference: Kernel choice, size distortions, and power losses. Working Paper, Harvard University.

Müller, U. (2007). A theory of robust long-run variance estimation. *Journal of Econometrics*, 141:1331–1352.

Phillips, P. (2005). HAC estimation by automated regression. *Econometric Theory*, 21:116–142.

Sun, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *Journal of Econometrics*, 164:345–366.

Sun, Y. (2013). A heteroskedasticity and autocorrelation robust f test using orthonormal series variance estimator. *Econometrics Journal*, 16:1–26.

Sun, Y. (2014a). Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. *Journal of Econometrics*, 178:659–677.

Sun, Y. (2014b). Fixed-smoothing asymptotics in a two-step GMM framework. *Econometrica*, 82:2327–2370.

Sun, Y. and Kim, M. (2012). Simple and powerful GMM over-identification tests with accurate size. *Journal of Econometrics*, 166:267–281.

Sun, Y. and Kim, M. (2015). Asymptotic F test in a GMM framework with cross sectional dependence. *Review of Economics and Statistics*, 97:210–223.

Sun, Y., Philips, P., and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica*, 76:175–194.

Sun, Y. and Phillips, P. (2009). Bandwidth choice for interval estimation in GMM regression. Working Paper, UC San Diego.

Vogelsang, T. (2012). Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. *Journal of Econometrics*, 166:303–319.

About the authors

Xiaoqing Ye is an associate professor at School of Mathematics and Statistics at South-Central University for Nationalities of China. Her research interests focus on topics in panel data and time series econometrics.

Yixiao Sun is a professor at Department of Economics at the University of California, San Diego. His research interests focus on topics in theoretical econometrics including heteroscedasticity and autocorrelation robust inference in the presence of temporal, spatial, and temporal-spatial dependence, and inference in nonparametric and semiparametric models.