The Stata Journal (2018)
18, Number 4, pp. 951–980

Heteroskedasticity- and autocorrelation-robust F and t tests in Stata

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

Yixiao Sun
Department of Economics
University of California, San Diego
La Jolla, CA
yisun@ucsd.edu

Abstract. In this article, we consider time-series, ordinary least-squares, and instrumental-variable regressions and introduce a new pair of commands, **har** and **hart**, that implement more accurate heteroskedasticity- 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, which are the two fundamental objects in every hypothesis testing problem. The estimation command **har** and the postestimation 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**, that implement the recently developed F and t tests in a two-step generalized method of moments framework. For these commands, we opt for the orthonormal-series HAR variance estimator based on the Fourier bases because it allows us to develop convenient F and t approximations as in the first-step generalized method of moments framework. Finally, we present several examples to demonstrate these commands.

Keywords: st0548, har, hart, gmmhar, gmmhart, heteroskedasticity- and autocorrelation-robust inference, fixed-smoothing, kernel function, orthonormal series, testing-optimal, AMSE, OLS/IV, two-step GMM, J statistic

1 Introduction

During the last two decades, there has been substantial progress in heteroskedasticity- and autocorrelation-robust (HAR) inference.

First, researchers developed fixed-smoothing asymptotic theory, which is a new class of asymptotic theory. 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, that is, 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, Philips, and Jin (2008). The accuracy is because the new asymptotic approximations capture the estimation uncertainty in the nonparametric HAR variance estimator. Both the effect of the smoothing

 st0548

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, researchers developed a new rule for selecting the smoothing parameter that is optimal for the HAR testing. Researchers have pointed out that the mean squared error (MSE) 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, Philips, and Jin (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, unlike the usual kernel HAR variance estimator, in which nonstandard critical values must be simulated.

The fixed-smoothing asymptotic approximations have been established in various settings. For the kernel HAR variance estimators, the smoothing parameter can be parameterized 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,b, 2005), Jansson (2004), Sun, Philips, and Jin (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), and Sun (2011, 2013). A recent article by Lazarus et al. (2016) shows that tests based on the OS HAR variance estimator have competitive power compared with tests based on the kernel HAR variance estimator with the optimal kernel.

Most research on fixed-smoothing asymptotics has been devoted to first-step generalized method of moments (GMM) estimation and inference. More recently, researchers established fixed-smoothing asymptotics in a general two-step GMM framework. See Sun and Kim (2012), Sun (2013, 2014b), and Hwang and Sun (2017). The key difference between first- and two-step GMM is that in two-step GMM, 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 for two possible reasons. First, the fixed-smoothing asymptotic distributions based on popular kernel variance estimators are nonstandard, and therefore critical values must be simulated. Second, no Stata command implements the new and more accurate approximations.

In this article, we describe the new estimation command `har` and the new postestimation test command `hart`, which implement the fixed-smoothing Wald and t tests of Sun (2013, 2014a) for linear regression models with possibly endogenous covariates. These two commands automatically select the testing-optimal smoothing parameter. We also 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. Thus, the new tests are convenient to use. In addition, Sun and Kim (2012) show that under the fixed-smoothing asymptotics, the J statistic for testing overidentification is also asymptotically F distributed.

The remainder of this 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. In section 6, we present some simulation evidence. In section 7, we describe the two postestimation test commands `hart` and `gmmhart`. In the last section, we conclude and discuss future work.

2 Fixed-smoothing asymptotics: First-step GMM

2.1 Ordinary least-squares and instrumental-variable regressions

Consider the regression model

$$\mathbf{Y}_t = \mathbf{X}_t\boldsymbol{\theta}_0 + \mathbf{e}_t \qquad t = 1, \ldots, T$$

where $\{\mathbf{e}_t\}$ is a zero-mean process that may be correlated with the covariate process $\{\mathbf{X}_t \in \mathbb{R}^{1\times d}\}$. There are instruments $\left\{\mathbf{Z}_t \in \mathbb{R}^{1\times m}\right\}$ such that the moment conditions

$$E\mathbf{Z}_t'\left(\mathbf{Y}_t - \mathbf{X}_t\boldsymbol{\theta}\right) = 0$$

hold if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. When $\mathbf{X}_t$ is exogenous, we take $\mathbf{Z}_t = \mathbf{X}_t$, leading to the moment conditions behind the ordinary least-squares (OLS) estimator. Note that the first elements of $\mathbf{X}_t$ and $\mathbf{Z}_t$ are typically 1. We allow the process $\{\mathbf{Z}_t'\mathbf{e}_t\}$ to have autocorrelation of unknown forms. The model may be overidentified with the degree of overidentification $q = m - d \geq 0$.

Define

$$\mathbf{S}_{ZX} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t'\mathbf{X}_t, \quad \mathbf{S}_{ZZ} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t'\mathbf{Z}_t, \quad \mathbf{S}_{ZY} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t'\mathbf{Y}_t$$

Then the instrumental-variable (IV) estimator of $\boldsymbol{\theta}_0$ is given by

$$\widehat{\boldsymbol{\theta}}_{\mathrm{IV}} = \left(\mathbf{S}_{ZX}'\mathbf{W}_{0T}^{-1}\mathbf{S}_{ZX}\right)^{-1}\left(\mathbf{S}_{ZX}'\mathbf{W}_{0T}^{-1}\mathbf{S}_{ZY}\right) \tag{1}$$

where $\mathbf{W}_{0T} = \mathbf{S}_{ZZ} \in \mathbb{R}^{m\times m}$. For the asymptotic results that follow, we can allow $\mathbf{W}_{0T}$ to be a general weighting matrix. One can assume that $\mathrm{Plim}_{T\to\infty}\mathbf{W}_{0T} = \mathbf{W}_0$ for a positive definite nonrandom matrix $\mathbf{W}_0$. When $\mathbf{Z}_t = \mathbf{X}_t$, the IV estimator reduces to the OLS estimator.

Suppose we are interested in testing the null $H_0 : \mathbf{R}\boldsymbol{\theta}_0 = \mathbf{r}$ against the alternative $H_1 : \mathbf{R}\boldsymbol{\theta}_0 \neq \mathbf{r}$, where $\mathbf{r} \in \mathbb{R}^{p\times 1}$ and $\mathbf{R} \in \mathbb{R}^{p\times d}$ is a matrix of full row rank. Nonlinear restrictions can be converted into linear restrictions via the delta method. Let $G_0 = E\mathbf{S}_{ZX} \in \mathbb{R}^{m\times d}$, and let $\mathbf{u}_t = \mathbf{R}\left(\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0\right)^{-1}\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{Z}_t'\mathbf{e}_t$ be the transformed moment process. Under some standard high-level conditions, we have

$$\sqrt{T}\mathbf{R}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{IV}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Omega}\right)$$

where $\boldsymbol{\Omega} = \sum_{j=-\infty}^{j=+\infty} E\mathbf{u}_t\mathbf{u}_{t-j}'$ is the long-run variance (LRV) of $\{\mathbf{u}_t\}$.

The Wald statistic for testing H_0 against H_1 is

$$F_{\mathrm{IV}} = \left\{\sqrt{T}\left(\mathbf{R}\widehat{\boldsymbol{\theta}}_{\mathrm{IV}} - \mathbf{r}\right)\right\}'\widehat{\boldsymbol{\Omega}}^{-1}\left\{\sqrt{T}\left(\mathbf{R}\widehat{\boldsymbol{\theta}}_{\mathrm{IV}} - \mathbf{r}\right)\right\}/p \tag{2}$$

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

$$t_{\mathrm{IV}} = \sqrt{T}\left(\mathbf{R}\widehat{\theta}_{\mathrm{IV}} - r\right)/\sqrt{\widehat{\Omega}}$$

Let $\mathbf{G}_T = \mathbf{S}_{ZX}$, $\widehat{\mathbf{u}}_t = \mathbf{R}\left(\mathbf{G}_T'\mathbf{W}_{0T}^{-1}\mathbf{G}_T\right)^{-1}\mathbf{G}_T'\mathbf{W}_{0T}^{-1}\mathbf{Z}_t'\left(\mathbf{Y}_t - \mathbf{X}_t\widehat{\boldsymbol{\theta}}_{\mathrm{IV}}\right)$, and $\widehat{\mathbf{u}}^{ave} = T^{-1}\sum_{s=1}^{T}\widehat{\mathbf{u}}_s$. We consider the estimator $\widehat{\boldsymbol{\Omega}}$ of the form

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} Q_h\left(\frac{s}{T}, \frac{t}{T}\right)\left(\widehat{\mathbf{u}}_t - \widehat{\mathbf{u}}^{ave}\right)\left(\widehat{\mathbf{u}}_s - \widehat{\mathbf{u}}^{ave}\right)' \tag{3}$$

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

$$\widehat{\boldsymbol{\Omega}} = \sum_{j=-(T-1)}^{T-1} k\left(\frac{j}{M_T}\right)\widehat{\boldsymbol{\Gamma}}_j$$

where

$$\widehat{\mathbf{\Gamma}}_j = \begin{cases} T^{-1} \sum_{t=j+1}^{T} \left(\widehat{\mathbf{u}}_t - \widehat{\mathbf{u}}^{ave}\right)\left(\widehat{\mathbf{u}}_{t-j} - \widehat{\mathbf{u}}^{ave}\right)' & \text{for} \quad j \geq 0 \\ T^{-1} \sum_{t=j+1}^{T} \left(\widehat{\mathbf{u}}_{t+j} - \widehat{\mathbf{u}}^{ave}\right)\left(\widehat{\mathbf{u}}_t - \widehat{\mathbf{u}}^{ave}\right)' & \text{for} \quad j < 0 \end{cases}$$

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

The OS HAR variance estimator has a long history. There is a renewed interest in this type of estimator in econometrics, starting from Phillips (2005), Sun (2006), and Müller (2007). 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 $j = 1,\ldots,K$. Here we assume that K is even and focus only on the Fourier basis functions:

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

In this case, $\widehat{\mathbf{\Omega}}$ 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 parameterize h so that h indicates the amount of smoothing. We consider the fixed-smoothing asymptotics under which $T \to \infty$ for a fixed h. Let

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

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

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

where

$$\begin{aligned} F_\infty(p,h) &= \mathbf{W}_p'\left(1\right)C_{pp}^{-1}\mathbf{W}_p\left(1\right)/p \\ t_\infty(p,h) &= \mathbf{W}_p\left(1\right)/\sqrt{C_{pp}} \\ C_{pp} &= \int_0^1\int_0^1 Q_h^*(r,s)d\mathbf{W}_p(r)d\mathbf{W}_p'(s) \end{aligned}$$

and $\mathbf{W}_p(r)$ 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, that is, the quantiles of $F_\infty(p,h)$ and $t_\infty(p,h)$, must 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

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

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

$$K = \max\left(\left\lceil \frac{1}{bc_2}\right\rceil, p\right) - p + 1 \tag{4}$$

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

$$\kappa = \frac{\exp\left[b\left\{c_1+(p-1)\,c_2\right\}\right] + \left[1+b\left\{c_1+(p-1)\,c_2\right\}\right]}{2} \tag{5}$$

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

$$\mathrm{t}_{\rm IV}^{1-\alpha} = \begin{cases} \sqrt{\kappa\mathcal{F}_{1,K}^{1-2\alpha}} & \alpha < 0.5 \\ -\sqrt{\kappa\mathcal{F}_{1,K}^{2\alpha-1}} & \alpha \geq 0.5 \end{cases} \tag{6}$$

To see this, consider the case where $\alpha < 0.5$. Because $t_\infty(p,h)$ 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 $\mathrm{t}_{\rm IV}^{1-\alpha} = \sqrt{\kappa\mathcal{F}_{1,K}^{1-2\alpha}}$ 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 $e_{\rm I}(b)$ and $e_{\rm II}(b)$, Sun (2014a) proposes selecting the bandwidth parameter b to solve the constrained minimization problem

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

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

Consider the local alternative $H_1(\delta) : \mathbf{R}\boldsymbol{\theta}_0 = \mathbf{r} + \boldsymbol{\Omega}^{1/2}\widetilde{\mathbf{c}}/\sqrt{T}$ for $\widetilde{\mathbf{c}}$ uniformly distributed on $\mathcal{S}_p\left(\delta^2\right) = (\widetilde{\mathbf{c}} \in \mathbb{R}^p : \|\widetilde{\mathbf{c}}\|^2 = \delta^2)$. Let $G'_{p,\delta^2}(\cdot)$ be the probability density function of the noncentral $\chi^2_p\left(\delta^2\right)$ distribution with degrees of freedom p and noncentrality parameter δ^2. Sun (2014a) shows 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

$$b_{\text{opt}} = \begin{cases} \left\{ \frac{2qG'_{p,\delta^2}\left(\mathcal{X}_p^{1-\alpha}\right)|\overline{\mathcal{B}}|}{\delta^2 G'_{(p+2),\delta^2}\left(\mathcal{X}_p^{1-\alpha}\right)c_2} \right\}^{\frac{1}{q+1}} T^{-\frac{q}{q+1}} & \overline{\mathcal{B}} > 0 \\ \left\{ \frac{G'_p\left(\mathcal{X}_p^{1-\alpha}\right)\mathcal{X}_p^{1-\alpha}|\overline{\mathcal{B}}|}{(\tau-1)\alpha} \right\}^{1/q} \frac{1}{T} & \overline{\mathcal{B}} \leq 0 \end{cases} \tag{7}$$

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 $\Pr\{\chi_p\left(\delta^2\right) > \mathcal{X}_p^{1-\alpha}\} = 75\%$,

$$\overline{\mathcal{B}} = tr(\mathbf{B}\boldsymbol{\Omega}^{-1})/p \quad \text{and} \quad \mathbf{B} = -\rho_q \sum_{h=-\infty}^{\infty} |h|^q \, E\mathbf{u}_t\mathbf{u}'_{t-h}$$

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 (7).

The parameter $\overline{\mathcal{B}}$ can be estimated by a standard vector autoregressive model of order 1 [VAR(1)] plugin procedure. This is what we opt for in the new commands. Plugging the estimate of $\overline{\mathcal{B}}$ into (7) yields $\widehat{b}_{\text{temp}}$. The data-driven choice of b_{opt} is then given by $\widehat{b}_{\text{opt}} = \min(\widehat{b}_{\text{temp}}, 0.5)$. We do not use a b larger than 0.5 to avoid large power loss.

2.3 The OS case

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

$$F_{\text{IV}} \xrightarrow{d} \frac{K}{K-p+1} \times \mathfrak{F}_{p,K-p+1}$$

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 convenient result because the fixed-smoothing asymptotic approximation is a standard distribution. There is no need to simulate critical values. Let $\mathcal{F}^{1-\alpha}_{p,K-p+1}$ be the $1-\alpha$ quantile of the F distribution $F_{p,K-p+1}$; we can then use

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

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

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

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 in the kernel case. We can first compute the optimal b^* for the following configuration: $q=2$, $c_2=1$, $\rho_q=\pi^2/6$. These are characteristic values associated with the Daniell kernel, the equivalent kernel behind the OS HAR variance estimator using Fourier bases. We then take $K=\lceil 1/(bc_2)\rceil$. More specifically, we use the following K value:

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

As before, the parameter $\overline{\mathcal{B}}$ is estimated by a standard VAR(1) plugin procedure. Plugging the estimate of $\overline{\mathcal{B}}$ into K_{opt} yields $\widehat{K}_{\mathrm{temp}}$. We truncate $\widehat{K}_{\mathrm{temp}}$ to be between $p+4$ and T. 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 large. Finally, we round $\widetilde{K}_{\mathrm{temp}}$ to the greatest even number less than $\widetilde{K}_{\mathrm{temp}}$. We take this greatest even number, denoted by $\widehat{K}_{\mathrm{opt}}$, 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 $H_0\colon \mathbf{R}\boldsymbol{\theta}_0=\mathbf{r}$ and the significance level α.

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

$$\widehat{u}_t = \mathbf{R}\left(\mathbf{G}_T'\mathbf{W}_{0T}^{-1}\mathbf{G}_T\right)^{-1}\mathbf{G}_T'\mathbf{W}_{0T}^{-1}\mathbf{Z}_t'\left(\mathbf{Y}_t-\mathbf{X}_t\widehat{\boldsymbol{\theta}}_{\mathrm{IV}}\right)$$

3. Fit a VAR(1) model into $\{\widehat{\mathbf{u}}_t\}$, and obtain a plugin estimator $\overline{\mathcal{B}}^{\mathrm{est}}$. Compute $\widehat{b}_{\mathrm{opt}}$ or $\widehat{K}_{\mathrm{opt}}$ as described in the previous two subsections.

4. For the kernel case, plug $\widehat{b}_{\text{opt}}$ into (4) and (5) to obtain $\widehat{K}$ and $\widehat{\kappa}$, and compute $\widehat{\mathcal{F}}_{\text{IV}}^{1-\alpha} = \widehat{\kappa}\mathcal{F}_{p,\widehat{K}}^{1-\alpha}$. For the OS case, compute

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

5. Calculate

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

$$\widehat{\mathbf{\Omega}} = \frac{1}{\widehat{K}_{\text{opt}}}\sum_{j=1}^{\widehat{K}_{\text{opt}}}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_j\left(\frac{t}{T}\right)\widehat{\mathbf{u}}_t\right\}\left\{\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\phi_j\left(\frac{s}{T}\right)\widehat{\mathbf{u}}_s\right\}'$$

respectively for the kernel case and the OS case.

6. Construct the test statistic:

$$F_{\text{IV}} = \left\{\sqrt{T}\left(\mathbf{R}\widehat{\boldsymbol{\theta}}_{\text{IV}} - \mathbf{r}\right)\right\}'\widehat{\mathbf{\Omega}}^{-1}\left\{\sqrt{T}\left(\mathbf{R}\widehat{\boldsymbol{\theta}}_{\text{IV}} - \mathbf{r}\right)\right\}/p$$

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

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 $\mathbf{R}$ to be the selection vector. For example, to select the second element of $\boldsymbol{\theta}$, we can let $\mathbf{R} = (0, 1, 0, \ldots, 0)$. The $100\,(1-\alpha)\,\%$ confidence interval for $\mathbf{R}\boldsymbol{\theta}_0$ is

$$\left(\mathbf{R}\widehat{\boldsymbol{\theta}}_{\text{IV}} - \mathrm{t}_{\text{IV}}^{1-\alpha/2} \times \sqrt{\widehat{\Omega}_R/T}, \quad \mathbf{R}\widehat{\boldsymbol{\theta}}_{\text{IV}} + \mathrm{t}_{\text{IV}}^{1-\alpha/2} \times \sqrt{\widehat{\Omega}_R/T}\right)$$

where $\mathrm{t}_{\text{IV}}^{1-\alpha/2}$ is defined in (6). Here we have added a subscript 'R' to $\widehat{\mathbf{\Omega}}$ to indicate its dependence on the restriction matrix $\mathbf{R}$.

3 Fixed-smoothing asymptotics: The two-step GMM

When any element of $\mathbf{X}_t$ is endogenous and there are more instruments than the number of regressors, we have an overidentified model. In this case, for efficiency, we may use a two-step GMM estimator and conduct inferences based on this estimator.

The two-step GMM estimator is given by

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}} &= \arg\min_{\theta\in\Theta} \mathbf{g}_T(\boldsymbol{\theta})' \left\{\mathbf{W}_T\left(\widehat{\boldsymbol{\theta}}_{\mathrm{IV}}\right)\right\}^{-1} \mathbf{g}_T(\boldsymbol{\theta}) \\ &= \left[\mathbf{S}'_{ZX}\left\{\mathbf{W}_T\left(\widehat{\boldsymbol{\theta}}_{\mathrm{IV}}\right)\right\}^{-1}\mathbf{S}_{ZX}\right]^{-1}\left[\mathbf{S}'_{ZX}\left\{\mathbf{W}_T\left(\widehat{\boldsymbol{\theta}}_{\mathrm{IV}}\right)\right\}^{-1}\mathbf{S}_{ZY}\right]\end{aligned}$$

where

$$\mathbf{W}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} Q_h\left(\frac{s}{T},\frac{t}{T}\right)\left\{v_t(\boldsymbol{\theta}) - \overline{v}(\boldsymbol{\theta})\right\}\left\{v_s(\boldsymbol{\theta}) - \overline{v}(\boldsymbol{\theta})\right\}' \tag{8}$$

$v_t(\boldsymbol{\theta}) = \mathbf{Z}'_t(\mathbf{Y}_t - \mathbf{X}_t\boldsymbol{\theta})$, and $\overline{v}(\boldsymbol{\theta}) = \sum_{t=1}^{T} v_t(\boldsymbol{\theta})/T$.

Note that $\mathbf{W}_T(\widehat{\boldsymbol{\theta}}_{\mathrm{IV}})$ is an estimator of the LRV of moment process $\{v_t(\boldsymbol{\theta}_0)\}$. It takes the same form as $\widehat{\boldsymbol{\Omega}}$ given in (3) but is based on the (estimated) moment process $\{v_t(\widehat{\boldsymbol{\theta}}_{\mathrm{IV}})\}$ instead of the (estimated) transformed moment process $\{\widehat{\mathbf{u}}_t\}$.

The Wald statistic is given by

$$\begin{aligned}F_{\mathrm{GMM}} = \sqrt{T}\left(\mathbf{R}\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}} - \mathbf{r}\right)'\left[\mathbf{R}\left\{\mathbf{G}'_T\mathbf{W}_T^{-1}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}\right)\mathbf{G}_T\right\}^{-1}\mathbf{R}'\right]^{-1} \\ \times\sqrt{T}\left(\mathbf{R}\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}} - \mathbf{r}\right)/p\end{aligned} \tag{9}$$

and the t statistic is given by

$$t_{\mathrm{GMM}} = \frac{\sqrt{T}\left(\mathbf{R}\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}} - \mathbf{r}\right)}{\sqrt{\mathbf{R}\left\{\mathbf{G}'_T\mathbf{W}_T^{-1}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}\right)\mathbf{G}_T\right\}^{-1}\mathbf{R}'}}$$

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

$$\begin{aligned}\mathbf{C}_{pp} &= \int_0^1\int_0^1 Q_h^*(r,s)dB_p(r)dB_p(s)' & \mathbf{C}_{pq} &= \int_0^1\int_0^1 Q_h^*(r,s)dB_p(r)dB_q(s)' \\ \mathbf{C}_{qq} &= \int_0^1\int_0^1 Q_h^*(r,s)dB_q(r)dB_q(s)' & \mathbf{D}_{pp} &= \mathbf{C}_{pp} - \mathbf{C}_{pq}\mathbf{C}_{qq}^{-1}\mathbf{C}'_{pq}\end{aligned}$$

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

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

The fixed-smoothing asymptotic distributions are nonstandard in both kernel and OS cases. For the OS case, Hwang and Sun (2017) show 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

$$F^c_{\text{GMM}} = \frac{K-p-q+1}{K}\frac{F_{\text{GMM}}}{1+\frac{1}{K}J_T}$$

$$t^c_{\text{GMM}} = \sqrt{\frac{K-q}{K}}\frac{t_{\text{GMM}}}{\sqrt{1+\frac{1}{K}J_T}}$$

where

$$J_T = T\mathbf{g}_T\left(\widehat{\boldsymbol{\theta}}_{\text{GMM}}\right)'\left\{\mathbf{W}_T\left(\widehat{\boldsymbol{\theta}}_{\text{GMM}}\right)\right\}^{-1}\mathbf{g}_T\left(\widehat{\boldsymbol{\theta}}_{\text{GMM}}\right)$$

is the usual J statistic for testing overidentification restrictions. It is shown in Hwang and Sun (2017) that

$$F^c_{\text{GMM}} \to^d F_{p,K-p-q+1} \quad \text{and} \quad t^c_{\text{GMM}} \to^d t_{K-q}$$

So we can use

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

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

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

as the critical value for the t statistic $|t_{\text{GMM}}|$. 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 an asymptotically F distribution:

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

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 (2008) for a suggestion for the smoothing parameter choice that is oriented toward interval estimation. For practical implementations, Hwang and Sun (2017) suggest selecting K based on the conventional average mean squared error (AMSE) criterion implemented by using the VAR(1) plugin procedure. More specifically,

$$\widehat{K}_{\text{tmp}} = \left\lceil\left(\frac{tr\left\{\left(\mathbf{I}_{m^2}+\mathbf{K}_{mm}\right)\left(\widehat{\boldsymbol{\Omega}}_v\otimes\widehat{\boldsymbol{\Omega}}_v\right)\right\}}{4vec\left(\widehat{\mathbf{B}}_v\right)'vec\left(\widehat{\mathbf{B}}_v\right)}\right)^{1/5}T^{4/5}\right\rceil$$

where $\mathbf{K}_{mm}$ is the $m^2 \times m^2$ commutation matrix and $\mathbf{I}_{m^2}$ is the $m^2 \times m^2$ identity matrix. In the above, $\widehat{\mathbf{B}}_v$ is the plugin estimator of

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

and $\widehat{\boldsymbol{\Omega}}_v$ is the plugin estimator of the LRV of $\{\mathbf{v}_t\}$. The formula for $\widehat{\boldsymbol{\Omega}}_v$ and $\widehat{\mathbf{B}}_v$ in terms of the estimated VAR(1) matrix and the error variance are available from Andrews (1991). We then obtain $\widetilde{K}_{\text{tmp}}$ by truncating $\widehat{K}_{\text{tmp}}$ to be between $p+q+4$ and T. Finally, we round $\widetilde{K}_{\text{tmp}}$ to $\widehat{K}_{\text{MSE}}$, the greatest even number less than $\widetilde{K}_{\text{tmp}}$ and use $\widehat{K}_{\text{MSE}}$ 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 $H_0 \colon \mathbf{R}\boldsymbol{\theta}_0 = \mathbf{r}$ and the significance level α.
2. Estimate $\boldsymbol{\theta}_0$ by the IV estimator, and construct $\widehat{\mathbf{v}}_t = \mathbf{Z}_t'(\mathbf{Y}_t - \mathbf{X}_t \widehat{\boldsymbol{\theta}}_{\text{IV}})$.
3. Fit a VAR(1) model into $\{\widehat{\mathbf{v}}_t\}$, and compute the data-driven choice $\widehat{K}_{\text{MSE}}$.
4. On the basis of $\widehat{K}_{\text{MSE}}$, construct the weighting matrix $\widehat{\mathbf{W}}_T = \mathbf{W}_T(\widehat{\boldsymbol{\theta}}_{\text{IV}})$ in (8).
5. Estimate $\boldsymbol{\theta}_0$ by

$$\widehat{\boldsymbol{\theta}}_{\text{GMM}} = \left(\mathbf{S}_{ZX}' \widehat{\mathbf{W}}_T^{-1} \mathbf{S}_{ZX}\right)^{-1} \left(\mathbf{S}_{ZX}' \widehat{\mathbf{W}}_T^{-1} \mathbf{S}_{ZY}\right)$$

6. Calculate the test statistic F_{GMM} defined in (9) and the critical value

$$\widehat{\mathcal{F}}_{\text{GMM}}^{1-\alpha} = \left\{1 + \frac{1}{\widehat{K}_{\text{MSE}}} J_T\left(\widehat{\boldsymbol{\theta}}_{\text{GMM}}\right)\right\} \left(\frac{\widehat{K}_{\text{MSE}}}{\widehat{K}_{\text{MSE}} - p - q + 1}\right) \mathcal{F}_{p, \widehat{K}_{\text{MSE}} - p - q + 1}^{1-\alpha}$$

7. If $F_{\text{GMM}} > \widehat{\mathcal{F}}_{\text{GMM}}^{1-\alpha}$, then we reject the null. Otherwise, we fail to reject the null.

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

Following the same procedure as in the IV case, we can use the two-step GMM estimator and construct the associated $100\,(1-\alpha)\,\%$ confidence interval for $\mathbf{R}\boldsymbol{\theta}_0$ as

$$\mathbf{R}\widehat{\boldsymbol{\theta}}_{\text{GMM}} \pm \mathrm{t}_{\text{GMM}}^{1-\alpha/2} \times \sqrt{\mathbf{R}\left\{\mathbf{G}_T' \mathbf{W}_T^{-1}\left(\widehat{\boldsymbol{\theta}}_{\text{GMM}}\right) \mathbf{G}_T\right\}^{-1} \mathbf{R}'/T}$$

where

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

4 The har command

4.1 Syntax

`har` *depvar* [*varlist1*] `(`*varlist2* `=` *varlist_iv*`)` [*if*] [*in*] `, kernel(`*string*`)`
[`noconstant level(`#`)`]

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

Time-series operators are allowed.

4.2 Options

`kernel(`*string*`)` sets the type of kernel. For the Bartlett kernel, any of the four usages—`kernel(bartlett)`, `kernel(BARTLETT)`, `kernel(B)`, or `kernel(b)`—produce the same results. Similarly, for the Parzen, QS, and OS 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.

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

4.3 Stored results

The `har` command uses `ivregress` to get the estimates of the model parameters. In addition to the standard stored results from `ivregress`, `har` 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(depvar)`	name of dependent variable
`e(title)`	title in the estimation output
`e(vcetype)`	title used to label Std. Err.
`e(carg)`	`nocons` or `""` if specified
`e(varline)`	variable line as typed
`e(kerneltype)`	kernel in the estimation

Matrices

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

Functions

`e(sample)`	marks estimation sample

NOTES: ∗ for OS; † for Bartlett, Parzen, and QS kernels.

We use the time-series data downloaded from the Stata Press website http://www.stata-press.com/data/r15/idle2.dta to illustrate the use of `har` by analyzing the influence of `idle` and `wio` on `usr`. The data are time series of 30 observations covering the periods from 08:20 to 18:00. We must `tsset` the dataset before using `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 titles of the table 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`, which use a single standard normal distribution because the testing-optimal smoothing parameter b depends on the null restriction vector $\mathbf{R}$. Each model parameter corresponds to a different vector $\mathbf{R}$ 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 QS 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
```

usr	Coef.	HAR 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 OS 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 Syntax

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

You must **tsset** your data before using **gmmhar**; see [TS] **tsset**.

Time-series operators are allowed.

5.2 Options

`noconstant` suppresses the constant term.

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

5.3 Stored results

The `gmmhar` command uses `ivregress` to get the column name 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 overidentification
Macros	
`e(cmd)`	`gmmhar`
`e(cmdline)`	command as typed
`e(depvar)`	name of the dependent variable
`e(title)`	title in estimation output
`e(vcetype)`	`orthonormal series`
`e(varline)`	variable line as typed
`e(carg)`	`nocons` or `""` if specified
`e(exog)`	exogenous variables
`e(endog)`	endogenous variables
`e(inst)`	instrument variables
Matrices	
`e(betahat)`	two-step GMM coefficient vector
`e(sstderr)`	adjusted standard error for each individual coefficient
`e(sdf)`	degrees of freedom of the t statistic
`e(st)`	t statistic
Functions	
`e(sample)`	marks estimation sample

5.4 Examples

To illustrate the use of `gmmhar` in the two-step GMM framework, we fit 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 gross domestic product 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 OS 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

------------------------------------------------------------------------------
             |                HAR
       D.inf |      Coef.   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 (UR) 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 OS approach, `noconstant`, confidence level 99%, AMSE automatic bandwidth selection:

```
. use http://fmwww.bc.edu/ec-p/data/stockwatson/macrodat, clear
. 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

------------------------------------------------------------------------------
             |                HAR
       D.inf |      Coef.   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 (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

Data-generating process for har

We consider the data-generating process

$$y_t = x_{0,t}\gamma + x_{1,t}\beta_1 + x_{2,t}\beta_2 + \varepsilon_t \tag{10}$$

where $x_{0,t} \equiv 1$ and $x_{1,t}$, $x_{2,t}$, and ε_t follow independent first-order autoregressive [AR(1)] processes,

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

or first-order moving-average [MA(1)] processes,

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

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

Data-generating process for gmmhar

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

$$y_t = x_{0,t}\gamma + x_{1,t}\beta_1 + x_{2,t}\beta_2 + \varepsilon_{y,t} \tag{11}$$

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

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

We consider the AR design here. Ye and Sun (2018), the working paper version of this article, contains more designs. In the AR design, each $z_{i,t}$ follows an AR(1) process of the form

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

where $e_{z_i,t} = \left(e_{zt}^i + e_{zt}^0\right)/\sqrt{2}$ and $\mathbf{e}_t = (e_{zt}^0, e_{zt}^1, \ldots, e_{zt}^{m-1})' \sim$ i.i.d.$N(\mathbf{0}, \mathbf{I}_m)$. By construction, each nonconstant z_{it} has unit variance and the correlation coefficient between the nonconstant $z_{i,t}$ and $z_{j,t}$ for $i \neq j$ is 0.5. The gross domestic product for $\boldsymbol{\varepsilon}_t = (\varepsilon_{y,t}, \varepsilon_{x_1,t}, \varepsilon_{x_2,t})$ is the same as that for $(z_{1,t}, \ldots, z_{m-1,t})$ except for the dimensional difference. We take $\rho = -0.5, 0.5, 0.8$. We let $\gamma = 1$, $\beta_1 = 3$, and $\beta_2 = 2$ without loss of generality.

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 use HAR variance estimators based on the Bartlett, Parzen, and QS kernels, as well as the orthonormal Fourier series. The number of simulation replications is 1,000.

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 (10). 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.

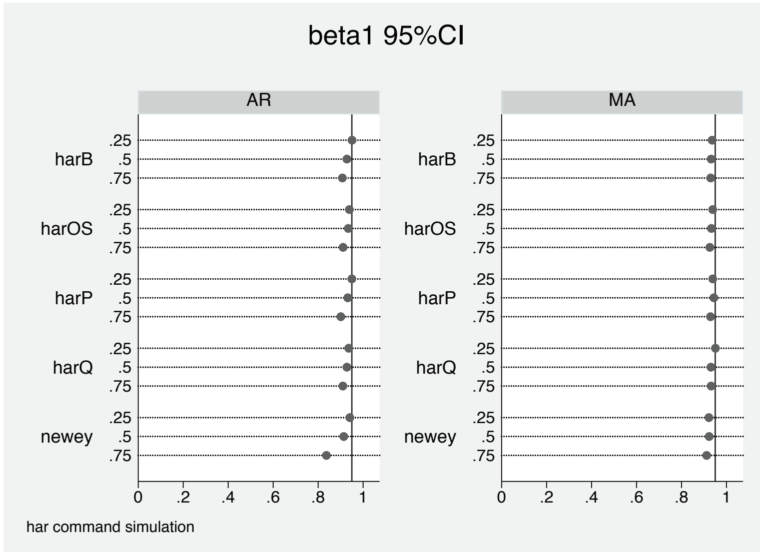

Figure 1. Empirical coverage rates of 95% confidence intervals of β_1 in model (10): the y labels 0.25, 0.5, and 0.75 indicate the AR or MA parameter

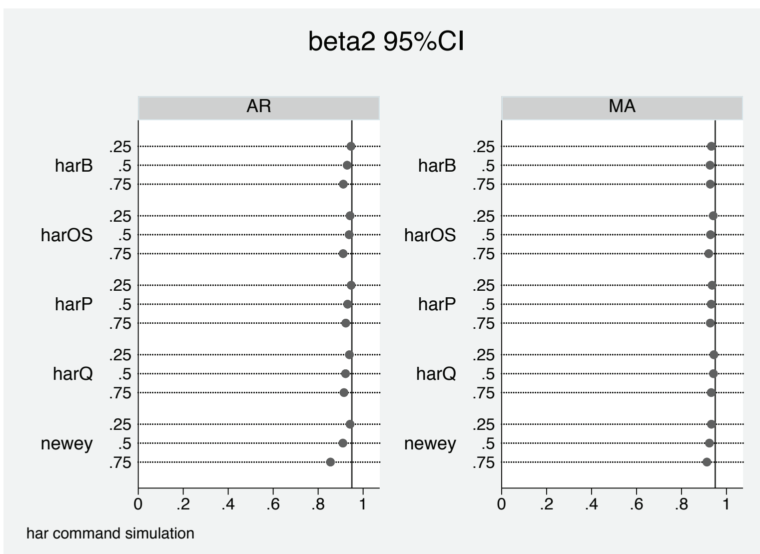

Figure 2. Empirical coverage rates of 95% confidence intervals of β_2 in model (10): the y labels 0.25, 0.5, and 0.75 denote the AR or MA parameter

Figures 3 and 4 report the simulation results based on `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 heteroskedasticity- and autocorrelation-consistent estimator using the data-driven truncation lag proposed by Newey and West (1994).

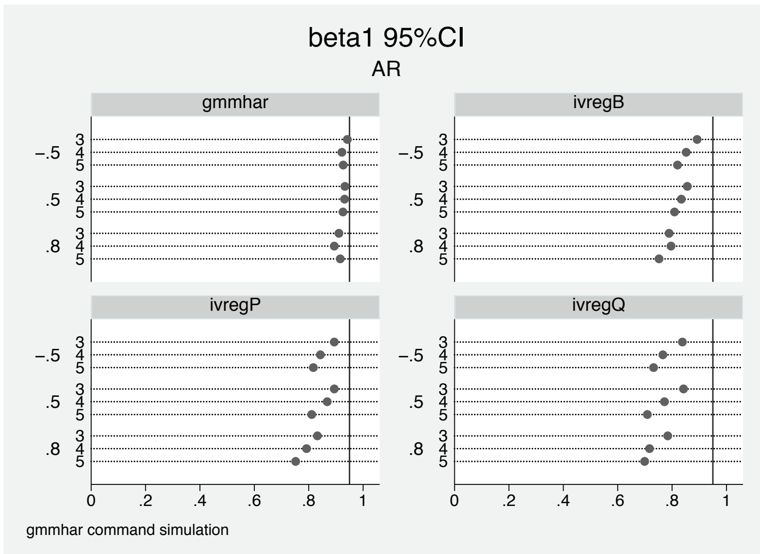

Figure 3. Empirical coverage rates of 95% confidence intervals of β_1 in model (11): the y labels −0.5, 0.5, and 0.8 indicate the values of the AR parameter, and the y sublabels 3, 4, and 5 indicate the number of instruments used (P = Parzen, B = Bartlett, and QS = kernels)

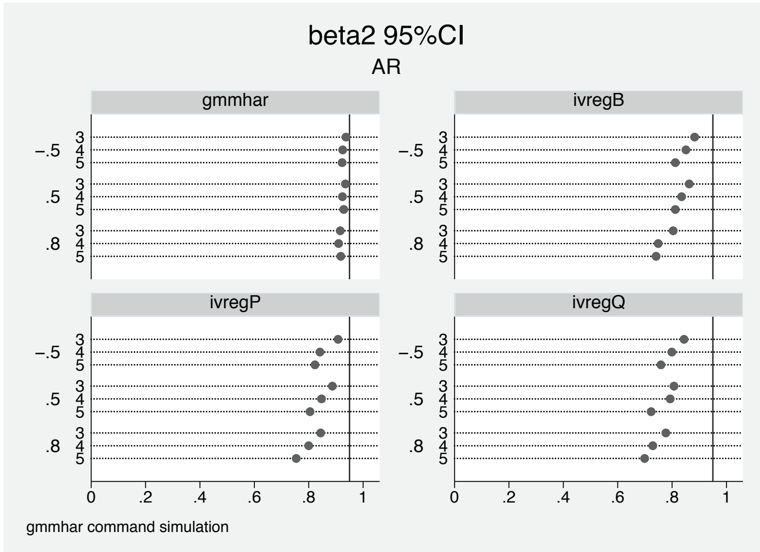

Figure 4. Empirical coverage rates of 95% confidence intervals of β_2 in model (11): the y labels -0.5, 0.5,, and 0.8 indicate the values of the AR parameter, and the y sublabels 3, 4, and 5 indicate the number of instruments used (P = Parzen, B = Bartlett, and QS = kernels)

As demonstrated by the two figures, the confidence intervals based on `gmmhar`, which uses the fixed-smoothing t approximations, are more accurate than those based on `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. In contrast, the coverage accuracy of the confidence intervals produced by `gmmhar` does not appear to be affected by the number of instruments.

7 The hart and gmmhart commands

`hart` and `gmmhart` are the postestimation commands that should be used immediately after the respective estimation commands `har` and `gmmhar.` These two commands perform the Wald type of tests but use more accurate fixed-smoothing critical values. The test statistics are given in (2) and (9), respectively.

7.1 Syntax

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

Tests that the listed coefficient are jointly 0:

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

`gmmhart` *coeflist* [, `accumulate`]

Tests a single or multiple linear restrictions:

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

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

`hart` implements the test described in section 2 for testing the null $H_0 : \mathbf{R}\boldsymbol{\theta}_0 = \mathbf{r}$ against the alternative $H_1 : \mathbf{R}\boldsymbol{\theta}_0 \neq \mathbf{r}$. The options `kernel(`*string*`)` and `level(`#`)` in `hart` must be consistent with those in `har`.

`gmmhart` implements the test described in section 3 for the same null and alternative hypotheses.

7.2 Options

`kernel(`*string*`)` sets the type of kernel. For the Bartlett kernel, any of the four usages—`kernel(bartlett)`, `kernel(BARTLETT)`, `kernel(B)`, or `kernel(b)`—produce the same results. Similarly, for the Parzen, QS, and OS 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.

`accumulate` tests the hypothesis jointly with previously tested hypotheses.

`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%.

7.3 Stored results

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

Scalars

†∗ `r(firdf)`	first degrees of freedom
†∗ `r(secdf)`	second degrees of freedom
†∗ `r(kopt)`	data-driven optimal K
∗ `r(lag)`	data-driven optimal truncation lag
†∗ `r(F)`	adjusted F statistic

Matrices

∗ `r(thetaiv)`	IV coefficient vector
† `r(thetagmm)`	two-step GMM coefficient vector

NOTES: ∗ 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 the 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) =   47.6645
            Prob > F =    0.0000

. quietly hart idle=0, kernel(bartlett)

. hart idle=wio, kernel(bartlett) acc
       F(  2,    17) =   47.6645
            Prob > F =    0.0000

. hart 1.168*idle=wio, kernel(bartlett)
       F(  1,    14) =    0.0000
            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 use the OS 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) =   43.1681
            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) =   11.9994
            Prob > F =    0.0123
```

Case 11: We use `gmmhart` to test three hypotheses based on the two-step GMM estimator with the (inverse) weighting matrix estimated by the OS 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
(10 missing values generated)
       F(  1,    43) =    2.05
            Prob > F =    0.1597
. gmmhart UR
(10 missing values generated)
       F(  1,    43) =    2.05
            Prob > F =    0.1597
. gmmhart UR=-0.09715
(10 missing values generated)
       F(  1,    43) =    0.00
            Prob > F =    1.0000
```

8 Conclusion

In this article, we presented the new estimation command **har** and the new postestimation test command **hart** in Stata. These commands extend the existing commands for linear regression models with time-series data. We used 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 **har** and **hart** and the tests based on the Stata commands **newey** and **test**. First, the bandwidth parameter is selected differently. While **newey** and **test** use a single data-driven smoothing parameter for all tests, **har** and **hart** use different smoothing parameters for different tests. The smoothing parameter behind **har** and **hart** is tailored toward 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 restriction, **newey** and **test** use a chi-squared approximation, while **har** and **hart** use an F approximation.

We also introduced another pair of commands, `gmmhar` and `gmmhart`, to be used in an overidentified linear IV regression. In this case, the efficient estimator minimizes a GMM criterion function that uses an LRV estimator as the weighting matrix. Thus, 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 `har` and `gmmhar` are consistent with those produced by the authors using Matlab.

9 Acknowledgments

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. W. K. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59: 817–858.

Bester, C. A., T. G. Conley, C. B. Hansen, and T. J. Vogelsang. 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 T. J. Vogelsang. 2011. Block bootstrap HAC robust tests: The sophistication of the naive bootstrap. *Econometric Theory* 27: 745–791.

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

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

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

———. 2002b. Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica* 70: 2093–2095.

———. 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21: 1130–1164.

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

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson. 2016. HAR inference: Kernel choice, size distortions, and power losses. Working Paper, Harvard University. https://scholar.harvard.edu/files/elazarus/files/har_inference.pdf.

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

Newey, W. K., and K. D. West. 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61: 631–653.

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

Sun, Y. 2006. Best quadratic unbiased estimators of integrated variance in the presence of market microstructure noise. Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1714751.

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

———. 2013. A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator. *Econometrics Journal* 16: 1–26.

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

———. 2014b. Fixed-smoothing asymptotics in a two-step generalized method of moments framework. *Econometrica* 82: 2327–2370.

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

———. 2015. Asymptotic F-test in a GMM framework with cross-sectional dependence. *Review of Economics and Statistics* 97: 210–223.

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

Sun, Y., and P. C. B. Phillips. 2008. Optimal bandwidth choice for interval estimation in GMM regression. Discussion Paper No. 1661, Cowles Foundation. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1136111.

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

Ye, X., and Y. Sun. 2018. Heteroscedasticity and autocorrelation robust F and t tests in Stata. Working Paper, Department of Economics, University of California, San Diego.

About the authors

Xiaoqing Ye is an associate professor at the 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 the Department of Economics at the University of California, San Diego. His research interests focus on topics in theoretical econometrics, including HAR inference in the presence of temporal, spatial, and temporal-spatial dependence, as well as inference in nonparametric and semiparametric models.