Comments on "HAR Inference: Recommendations for Practice"

By Lazarus, Lewis, Stock, and Watson

Discussant: Yixiao Sun, UC San Diego*

I would like to congratulate the authors (Lazarus, Lewis, Stock, and Watson, LLSW hereafter) for the timely and authoritative practical recommendations for heteroskedasticity and autocorrelation robust (HAR)[1] inference. Such recommendations are long overdue. There is substantial progress in the econometric theory literature on HAR inference in the last two decades, but the theoretical research has not influenced empirical practices as much as it should have, even though the new generation of HAR tests are more accurate than the classical HAR tests by Newey and West (1987) and Andrews (1991) and are as easy to use as the latter tests. LLSW's paper has the potential to change the status quo in applied research.

The literature on HAR inference has advanced on at least three fronts: fixed-smoothing asymptotics, testing-optimal smoothing parameter choice, and asymptotic t and F tests. LLSW provide practical recommendations on each of the three fronts. Below I discuss them in turn; the references provided are by no means complete. I also provide links to some Stata implementations of the new HAR tests.

1 Fixed-smoothing asymptotics

The first and most important advance is the development of the fixed-smoothing asymptotic theory, which includes as a special case the fixed-b asymptotic theory pioneered by Kiefer and Vogelsang (2002a,b, 2005) in the time series setting. The idea of holding the amount of smoothing fixed to develop new asymptotic approximations has been extended to accommodate spatial data, spatial and temporal data, panel data, and clustered data. From a broad perspective, the idea can be used in any setting where there is a need to estimate the asymptotic variance nonparametrically.

It is now well known that fixed-smoothing asymptotic approximations are more accurate than the normal and chi-square approximations. This has been confirmed by ample simulation evidence and supported by higher order asymptotic theory in Jansson (2004) and Sun, Phillips, and Jin (2008).

*I would like to thank Xiaohong Chen, David Kaplan, Cheng Liu, and Peter Phillips for helpful comments. The usual disclaimer applies.

[1] To the best of my knowledge, the acronym HAR was first introduced in Phillips (2005a) where he coined the term to refer collectively to a group of methods that attempts robustification of inference and includes all the methods that I discuss here.

LLSW focus on location models and regression models. I would like to add that fixed-smoothing approximations have been established for many more models and in more general frameworks including both the GMM framework and the empirical likelihood framework. In the former case, both the first-step GMM (e.g., (Kiefer and Vogelsang, 2005)) and the two-step GMM (e.g., (Sun, 2014b)) have been considered.

Regression models are the most important models in any discipline. LLSW's paper is, therefore, well positioned to advocate the use of fixed-smoothing approximations in applied research.

2 Testing-optimal smoothing parameter choice

The second important advance in HAR inference is the development of testing-optimal rules for selecting the smoothing parameter. These rules target the fundamental quantities in HAR testing and employ criteria that are different from the traditional mean squared error (MSE) criterion.

For hypothesis testing, the fundamental quantities are arguably the type I error and the type II error. In Sun, Phillips and Jin (2008), we suggest that one should choose the smoothing parameter to minimize a loss function that is a weighted average of the type I and type II errors with the weights reflecting the relative consequences of committing these two types of errors. I will refer to this rule as the SPJ rule hereafter. Equivalently, one could minimize one type of error subject to the control of the other type of error. See Sun (2014a) for details of such a rule. LLSW propose an alternative rule to smoothing parameter choice. Their rule is also oriented toward the testing problem at hand but strikes an optimal balance between the absolute size distortion and the size-adjusted power loss relative to an oracle who knows the true asymptotic variance.

LLSW's rule and SPJ's rule share some similarities. Both are oriented towards the fundamental objectives of interest in a testing problem. Both rules result in smoothing parameter choices that balance the asymptotic variance of the underlying LRV estimator with its asymptotic bias. This is in sharp contrast with the MSE-based rule that balances the asymptotic variance and the *squared* asymptotic bias. For this reason, when the moment processes are persistent, both rules call for a larger truncation lag in order to achieve more bias reduction than what is deemed optimal under the MSE criterion.

There are, however, some important differences between these two rules. Take the case with kernel HAR estimation as an example. Let F_T be the Wald statistic for testing m joint restrictions and $c_{m,b}^{\alpha}$ be the fixed-b critical value for a level-α test. Let q be the order of the kernel, and define

$$\omega^{(q)} = \operatorname{tr}\left(\frac{1}{m}\sum_{j=-\infty}^{\infty}|j|^q\,\Gamma_j\Omega^{-1}\right),$$

where Γ_j is the autocovariance of order j and Ω is the long run variance (LRV) of the underlying time series. Then it follows from Sun (2014a) that under the traditional small-b asymptotics:

$$\Pr_{H_0}(F_T > c_{m,b}^{\alpha}) \doteq \alpha + G_m'\left(\chi_m^{\alpha}\right)\chi_m^{\alpha}\cdot k^{(q)}\omega^{(q)}\cdot (bT)^{-q},$$

where higher order terms have been ignored. In the above, b is the ratio of the truncation lag to the sample size T, $k^{(q)}$ is its Parzen exponent, and χ_m^α is the $(1-\alpha)$-quantile of the chi-squared distribution with m degrees of freedom. The type I error and the size distortion of the fixed-b test are then approximately measured by

$$\begin{aligned} e_{\mathrm{I}}(b) &= \alpha + G_m'\left(\chi_m^\alpha\right)\chi_m^\alpha \cdot k^{(q)}\omega^{(q)} \cdot (bT)^{-q}, \\ \Delta_{\mathrm{S}}\left(b\right) &= G_m'\left(\chi_m^\alpha\right)\chi_m^\alpha \cdot k^{(q)}\omega^{(q)} \cdot (bT)^{-q}, \end{aligned}$$

respectively. Only the nonparametric bias of the LRV estimator is reflected in the above measures. The variance effect has been corrected by using the fixed-b critical value $c_{m,b}^\alpha$, which is larger than the chi-squared critical value χ_m^α. We note that SPJ use $e_{\mathrm{I}}(b)$ as a part of their loss function, and LLSW use $\Delta_{\mathrm{S}}\left(b\right)$ as the counterpart.

While $e_{\mathrm{I}}(b)$ and $\Delta_{\mathrm{S}}\left(b\right)$ differ only by a constant, i.e., the nominal level of the test, there is a substantial difference between the approximate type II error and the approximate size-adjusted power, which are employed respectively in SPJ's and LLSW's loss functions. In the location model with long run variance Ω and under the local alternative

$$H_1\left(\delta^2\right) : \beta = \beta_0 + \frac{\Omega^{1/2}c}{\sqrt{T}}$$

for $c \in \mathbb{R}^m$ that is uniformly distributed on a sphere with radius δ, the type II error can be approximated by

$$e_{\mathrm{II}}\left(\delta^2, b\right) = G_{m,\delta^2}\left(\chi_m^\alpha\right) - G_{m,\delta^2}'\left(\chi_m^\alpha\right)\chi_m^\alpha \cdot k^{(q)}\omega^{(q)} \cdot (bT)^{-q} + \frac{\delta^2}{2}G_{(m+2),\delta^2}'\left(\chi_m^\alpha\right)\chi_m^\alpha \cdot k_2 b,$$

where k_2 is the integral of the squared kernel function and G_{m,δ^2} is the CDF of a non-central chi-squared variate with noncentrality parameter δ^2. The target behind $e_{\mathrm{II}}\left(\delta^2, b\right)$ is $\Pr_{H_1}(F_T < c_{m,b}^\alpha)$, the finite-sample type II error. Note that $e_{\mathrm{II}}\left(\delta^2, b\right)$ is the first-order approximation to $\Pr_{H_1}(F_T < c_{m,b}^\alpha)$, that is

$$e_{\mathrm{II}}\left(\delta^2, b\right) \doteq \Pr_{H_1}(F_T < c_{m,b}^\alpha).$$

To calculate the size-adjusted power, LLSW first adjust the fixed-b critical value $c_{m,b}^\alpha$ to remove the bias effect in the null rejection probability so that the test has no first-order size distortion. This gives rise to the infeasible critical value

$$c_{m,b}^{\alpha,T} = c_{m,b}^\alpha\left(\left[1 + \omega^{(q)}k^{(q)}\left(bT\right)^{-q}\right]\right).$$

Such a critical value is not supposed to be used in practice, as $\omega^{(q)}$ is not known. It is a theoretical construct for the purpose of calculating the size-adjusted power. The size-adjusted power loss is based on the expansion

$$\Pr_{H_1(\delta^2)}(F_T > c_{m,b}^{\alpha,T}) \doteq 1 - G_{m,\delta^2}\left(\chi_m^\alpha\right) - \frac{\delta^2}{2}G_{(m+2),\delta^2}'\left(\chi_m^\alpha\right)\chi_m^\alpha \cdot k_2 b$$

and is given by

$$\Delta_{\mathrm{P}}\left(\delta^2, b\right) = \frac{\delta^2}{2}G_{(m+2),\delta^2}'\left(\chi_m^\alpha\right)\chi_m^\alpha \cdot k_2 b.$$

The target behind $\Delta_{\mathrm{P}}\left(\delta^{2},b\right)$ is the difference between $1-G_{m,\delta^{2}}\left(\chi_{m}^{\alpha}\right)$, the oracle power, and the actual finite-sample size-adjusted power $\Pr_{H_1}(F_T > c_{m,b}^{\alpha,T})$. That is

$$\begin{aligned}\Delta_{\mathrm{P}}\left(\delta^{2},b\right) &\doteq 1-G_{m,\delta^{2}}\left(\chi_{m}^{\alpha}\right)-\Pr_{H_1}(F_T > c_{m,b}^{\alpha,T}) \\ &= \Pr_{H_1}(F_T < c_{m,b}^{\alpha,T})-G_{m,\delta^{2}}\left(\chi_{m}^{\alpha}\right).\end{aligned}$$

Comparing the targets behind $e_{\mathrm{II}}\left(\delta^{2},b\right)$ and $\Delta_{\mathrm{P}}\left(\delta^{2},b\right)$, we can see that there are two major differences. First, different critical values are used. While $e_{\mathrm{II}}\left(\delta^{2},b\right)$ is based on the fixed-b critical value $c_{m,b}^{\alpha}$, which is feasible, $\Delta_{\mathrm{P}}\left(\delta^{2},b\right)$ is based on the adjusted fixed-b critical value $c_{m,b}^{\alpha,T}$, which is not feasible. Second, there is an additional constant in $\Delta_{\mathrm{P}}\left(\delta^{2},b\right)$, namely, $-G_{m,\delta^{2}}\left(\chi_{m}^{\alpha}\right)$. The presence of such a constant does have an effect on LLSW's rule.

The testing-optimal b of SPJ minimizes a weighted average of the type I and type II errors:

$$b_{\mathrm{SPJ}}=\arg\min_{b}\left\{ae_{\mathrm{I}}\left(b\right)+(1-a)e_{\mathrm{II}}\left(\delta^{2},b\right)\right\},$$

where a is the relative loss under the two types of errors. The loss function has both economic and statistical interpretations. From an economic perspective, the loss function characterizes the expected loss for the decision-making rule $1\left\{F_T > c_{m,b}^{\alpha}\right\}$. To see this, let the losses from committing type I error and type II error be $2a$ and $2\left(1-a\right)$, respectively and assume that H_0 and $H_1\left(\delta^{2}\right)$ are equally likely. Then the expected loss from the decision rule $1\left\{F_T > c_{m,b}^{\alpha}\right\}$ is approximately $ae_{\mathrm{I}}\left(b\right)+(1-a)e_{\mathrm{II}}\left(\delta^{2},b\right)$, which is the same as the loss behind the SPJ rule. From a statistical perspective, the above choice b_{SPJ} can be equivalently formulated as minimizing the type II error (or maximizing the raw power) subject to the control of the type I error. So, the rule is in exactly the same spirit as the classical Neyman–Pearson lemma.

The loss function that LLSW employ is a weighted average of the *squared* size distortion and the *squared* size-adjusted power loss. Their choice of b is given by

$$b_{\mathrm{LLSW}}=\arg\min_{b}\left\{\kappa\left[\Delta_{\mathrm{S}}\left(b\right)\right]^{2}+(1-\kappa)\left[\Delta_{\mathrm{P}}\left(\delta^{2},b\right)\right]^{2}\right\}, \tag{1}$$

where κ is a constant. LLSW consider using the worst-case power loss, and they replace $\Delta_{\mathrm{P}}\left(\delta^{2},b\right)$ by $\max_{\delta^{2}}\Delta_{\mathrm{P}}\left(\delta^{2},b\right)$. However, to facilitate the comparison, I will use the version in (1).

To see the difference between these two rules, we can examine their respective exact finite-sample versions, namely

$$b_{\mathrm{SPJ}}^{*}=\arg\min_{b}\left\{a\Pr_{H_0}(F_T > c_{m,b}^{\alpha})+(1-a)\Pr_{H_1(\delta^{2})}(F_T > c_{m,b}^{\alpha})\right\},$$

$$b_{\mathrm{LLSW}}^{*}=\arg\min_{b}\left\{\kappa\left[\Pr_{H_0}(F_T > c_{m,b}^{\alpha})-\alpha\right]^{2}+(1-\kappa)\left[\Pr_{H_1(\delta^{2})}(F_T < c_{m,b}^{\alpha,T})-G_{m,\delta^{2}}\left(\chi_{m}^{\alpha}\right)\right]^{2}\right\}.$$

The LLSW rule insists that the type I error of the fixed-b test should be as close to α as possible. Because of the quadratic form it uses, the rule treats over-rejection and under-rejection symmetrically. Given that over-rejection corresponds to a larger type I error and under-rejection corresponds to a smaller type I error, the loss function designed by LLSW appears to be incompatible with the preference for a smaller type I error. Fundamentally, the LLSW loss function takes the absolute size distortion as an ingredient. As a result, their optimal choice of b does not respect the direction of the asymptotic bias of the LRV estimator. This aspect of smoothing parameter choice resembles the MSE-based rule. In contrast, the SPJ rule allows positive and negative biases to have different effects on the loss function. The difference between the SPJ rule and the LLSW rule may not be too large if the underlying moment processes have positive autocorrelation. In this case, $\omega^{(q)} > 0$, so $e_{\mathrm{I}}(b)$ and $[\Delta_{\mathrm{S}}(b)]^2$ move in the same direction as b changes. Similarly, $e_{\mathrm{II}}\left(\delta^2, b\right)$ and $\left[\Delta_{\mathrm{P}}\left(\delta^2, b\right)\right]^2$ move in the same direction as b changes. However, in a regression setting or a more general GMM setting, the moment processes are not necessarily positively autocorrelated, and the difference between the SPJ rule and the LLSW rule can be large.

To a great extent, both SPJ and LLSW attempt to solve a multi-objective optimization problem, but different sets of objectives are chosen. SPJ employ the type I and type II errors as the objectives with zero as the ideal target for each of the two types of errors. LLSW employ the absolute size distortion and the size-adjusted power as the objectives with the exact nominal level α and the oracle power $1 - G_{m,\delta^2}\left(\chi_m^\alpha\right)$ as the ideal and implicit targets. Both rules involve scalarizing their respective multi-objective optimization problem into a single objective. While we often examine the size-adjusted power alone in a simulation study, I feel that it is not so natural to combine the size distortion and the size-adjusted power in a simple loss function as they are computed under different critical values. There may not exist any test that has the same size distortion with the unadjusted raw power equal to the best size-adjusted power.

To sum up, testing-oriented rules to smoothing parameter choice should be used in HAR testing, but there is room for debating which testing-oriented rule we should use. The SPJ and LLSW rules share some similarities, but there are some important differences, reflecting different preferences and loss functions.

3 Asymptotic t and F tests

The third recent advance in the HAR inference is the development of asymptotic t and F tests. These tests are based on the orthonormal series (OS) HAR variance estimator. Such an estimator has a long history. An early example is the average periodograms estimator that involves taking a simple average of the first few periodograms. More recently, starting from Phillips (2005b), Sun (2006), and Müller (2007), there is a renewed interest in this type of estimator in econometrics. In the traditional increasing-smoothing framework, there is no advantage of using the OS HAR estimator. In fact, it is suboptimal compared to the kernel HAR estimator using the quadratic spectral kernel. However, in the fixed-smoothing asymptotic framework, the asymptotic distributions for the t statistic and the Wald statistics based on OS HAR variance estimators turn out to be the standard t and F distributions, which are as convenient to use as the standard normal and chi-squared approximations.

One of the reasons that the fixed-b asymptotic approximations have not been widely used is that they are nonstandard, and critical values have to be simulated. In principle, critical values can be simulated and tabulated. However, it is encouraging to see that there is an alternative class of HAR estimators that can deliver convenient, competitive, and reliable tests.

LLSW include two OS HAR variance estimators as serious contenders in their extensive simulation study. They find that the asymptotic t and F tests based on type II cosine series can deliver superior performance, especially for testing a single restriction. This echoes the results in Lazarus et al. (2017). LLSW's paper may start a new tradition of using OS HAR variance estimators and the associated t and F tests in applied research.

Here I would like to add that asymptotic t and F tests are nearly universally applicable. In some situations, carefully crafted basis functions are needed to develop the asymptotic t and F theory. As a partial list, the theory has been established in the following time series settings:

1. Trend regression (Sun, 2011)
2. First-step GMM (Müller, 2007; Sun, 2013, 2014a; Lazarus et al., 2017)
3. Two-step efficient GMM (Sun and Phillips, 2009; Sun, 2014b; Hwang and Sun, 2017)
4. Overidentification testing (Sun and Kim, 2012; Chen and Liao, 2015)
5. Nonparametric and semiparametric models (Chen et al., 2014; Kim et al., 2017)

Asymptotic t and F tests that accommodate panel data and spatial data include Kim and Sun (2013) and Sun and Kim (2015), among others. Liu and Sun (2018) develop an asymptotic t test in a difference-in-differences regression.

From a more broad perspective, the idea of using orthonormal basis functions in HAR inference is closely related to the use of trend sinusoidal regressors/instruments in extracting long run variation of a single or multiple time series. To a great extent, the literature on OS HAR inference is inspired from Phillips (1998, 2005b, 2014). See Müller and Watson (2016) for further discussion on basis-function transformations and their applications in econometrics.

4 Conclusion

Recent research on HAR inference after the seminal contributions by Newey and West (1987) and Andrews (1991) has developed more accurate approximations. These approximations, combined with the smoothing parameter choice that targets the testing problem at hand, lead to more trustworthy inferences. Through extensive simulations, especially those with DGP's estimated from macroeconomic data, LLSW provide compelling evidence that the new generation of the HAR tests should become the standard practice in empirical applications. Depending on the ultimate objectives, one can use the smoothing parameter choice rule introduced by LLSW or a rule in the existing literature such as the one by SPJ.

I would like to take the opportunity to advertise some implementations of the fixed-smoothing tests. For a Stata implementation of the fixed-b tests, readers are referred to Tim Vogelsang's web page at https://msu.edu/ tjv/working.html. For a Stata implementation of the fixed-smoothing asymptotic t and F tests, readers are referred to my web page at http://www.econ.ucsd.edu/ yisun/har.html. The latter implementation includes the testing-optimal smoothing parameter choice that balances type I and type II errors, and we plan to incorporate the rule suggested by LLSW in a future edition. See Ye and Sun (2018) for more details.

References

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

Chen, X. and Liao, Z. (2015). Sieve semiparametric two-step GMM under weak dependence. *Journal of Econometrics*, 189(1):163 – 186.

Chen, X., Liao, Z., and Sun, Y. (2014). Sieve inference on possibly misspecified semi-nonparametric time series models. *Journal of Econometrics*, 178:639–658.

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

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

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

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

Kim, M. S., Sun, Y., and Yang, J. (2017). A fixed-bandwidth view of the pre-asymptotic inference for kernel smoothing with time series data. *Journal of Econometrics*, 197(2):298–322.

Lazarus, E., Lewis, D. J., and Stock, J. H. (2017). The size-power tradeoff in HAR inference. Working Paper, Harvard University.

Liu, C. and Sun, Y. (2018). Simple and trustworthy asymptotic t tests in difference-in-differences regressions. Working Paper, UC San Diego.

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

Müller, U. K. and Watson, M. W. (2016). Low frequency econometrics. Working Paper, Department of Economics, Princeton University.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Phillips, P. C. B. (1998). New tools for understanding spurious regressions. *Econometrica*, 66(6):1299–1325.

Phillips, P. C. B. (2005a). Automated discovery in econometrics. *Econometric Theory*, 21(1):3–20.

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

Phillips, P. C. B. (2014). Optimal estimation of cointegrated systems with irrelevant instruments. *Journal of Econometrics*, 178:210–224.

Sun, Y. (2006). Best quadratic unbiased estimators of integrated variances in the presence of market microstructure noise. Working Paper, Department of Economics, UC San Diego.

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

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

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

Ye, X. and Sun, Y. (2018). Heteroscedasticity and autocorrelation robust F and t tests in Stata. Working Paper, UC San Diego.