Heteroskedasticity Autocorrelation Robust Inference in Time Series Regressions with Missing Data

Seung-Hwa Rho — Louisiana State University[†]
Timothy J. Vogelsang[*] — Michigan State University[‡]

January 24, 2018

Abstract

In this paper, we investigate the properties of heteroskedasticity and autocorrelation robust (HAR) test statistics in time series regression settings when observations are missing. We primarily focus on the non-random missing process case where we treat the missing locations to be fixed as $T \to \infty$ by mapping the missing and observed cutoff dates into points on $[0,1]$ based on the proportion of time periods in the sample that occur up to those cutoff dates. We consider two models, the amplitude modulated series (Parzen (1963)) regression model, which amounts to plugging in zeros for missing observations, and the equal space regression model, which simply ignores the missing observations. When the amplitude modulated series regression model is used, the fixed-b limits of the HAR test statistics depend on the locations of missing observations but are otherwise pivotal. When the equal space regression model is used, the fixed-b limits of the HAR test statistics have the standard fixed-b limits as in Kiefer and Vogelsang (2005). We discuss methods for obtaining fixed-b critical values with a focus on bootstrap methods and find the naive *i.i.d.* bootstrap with missing dates fixed to be an effective and practical way to obtain the fixed-b critical values.

Keywords: Missing data; Unequally spaced; Irregularly observed; Serial correlation robust

[*]We are grateful for helpful suggestions and comments from the anonymous referees and seminar participants at Columbia University and the Joint Montreal Econometrics Seminar.
[†]Department of Economics, srho@lsu.edu
[‡]Department of Economics, tvj@msu.edu

1 INTRODUCTION

It is not unusual to encounter a time series data set with missing observations but in a relatively simple context of time series regression, there appears to be a sparsity of work related to missing data issues. Most of the times series literature on missing data focuses on the estimation of dynamic models where the goal is to forecast missing observations and little appears to be known about the impact of missing data on heteroskedasticity and autocorrelation robust (HAR) tests in regression settings. This paper attempts to fill this void by analyzing the impact of missing data on HAR tests based on nonparametric kernel estimators of long run variances (HAR variance estimators).

We consider the case where the missing process is exogenous to the underlying latent process and the estimators are identified by the observed data. Hence we are not proposing a new estimator. Our focus rather lies on the impact of missing data on the long run variance estimator and further deriving a reference distribution for HAR inference which captures the impact of missing observations. We consider two models, the amplitude modulated (AM) series and equal space (ES) regression models. The ES regression model, which is formally considered in Datta and Du (2012), simply ignores the missing observations and treats the observations as if they are equally spaced in time. The AM regression model, first proposed by Parzen (1963), characterizes the missing observations as being driven by a 0-1 binary missing process. In terms of a regression model this amounts to plugging in zeros for missing observations. The literature following Parzen (1963) mainly focused on investigating the impact of missing data on the consistent estimation of spectral density functions of the underlying latent series. See, for example, Scheinok (1965), Bloomfield (1970), Neave (1970), and Dunsmuir and Robinson (1981).[1] While HAR inference makes use of spectral estimation methods, with the exception of Datta and Du (2012), there appears to be no attempt in the literature to link this earlier literature on spectral density estimation to regression inference with missing data.

The AM approach appears to be intuitively more sensible than the ES approach because the time distances between observations become skewed for many pairs of time periods for the ES approach while the time distances are preserved for the AM approach. This would seem particularly relevant for testing based on HAR variance estimators (e.g.,Newey and West (1987), Andrews (1991)) given that those esti-

[1]Scheinok (1965) and Bloomfield (1970) consider estimating a spectral density function of the observed process (with missing data) with independent Bernoulli and dependent Bernoulli missing processes respectively. Neave (1970) estimates a spectral density function with initially scarce data. Later work by Dunsmuir and Robinson (1981) investigated consistent estimation of the spectral density of the underlying latent process.

mators employ quadratic forms with weights that depend on the time distances of pairs of observations. One might reasonably conjecture (and we also conjectured) that the ES approach would be problematic. In practice the AM approach is prominent, but the ES approach is still used. For example, in the statistical package Stata, the command *newey* with the 'force' option computes Newey-West standard errors using the AM approach whereas the command *newey2* with the 'force' option or the command *hacreg* computes Newey-West standard errors based on the ES approach. Surprisingly, we find that the ES approach can be justified theoretically with the fixed-b asymptotic framework and works better than one might expect.

Our work is most closely related to Datta and Du (2012) who also analyze the AM and ES approaches in time series regressions with missing data. Their results provide a good foundation for the traditional small bandwidth asymptotic theory on HAR tests which appeals to consistency of the HAR variance estimators. Our results, on the other hand, are based on fixed-b asymptotic framework as in Kiefer and Vogelsang (2005). The fixed-b results that we obtain can be viewed as useful refinements to the traditional theory because it is now well established that fixed-b theory provides improved approximations by capturing much of the impact of kernel and bandwidth choices on finite sample distributions of HAR tests (see e.g., Jansson (2002), Sun et al. (2008), Gonçalves and Vogelsang (2011)). In addition, we show that it is possible to capture the impact of missing data on HAR robust tests via the asymptotic reference distribution. This is done by mapping the missing locations into points on $[0, 1]$ as fixed proportions of the total time span. This mapping is conceptually the same as mapping a given number of break dates to corresponding break points in a model with structural change and obtaining asymptotic results holding the break points fixed as $T \to \infty$. Having critical values that depend on the missing dates improves the reliability of finite sample inference.

The following are the main theoretical findings in the paper. (1) When the missing process is non-random, the fixed-b limits of the HAR $Wald$ statistics computed from the AM approach depend on the locations of missing observations in addition to the kernel function and the bandwidth used but are otherwise pivotal. (2) When the missing data are simply ignored (ES approach), we find that the fixed-b limits of the HAR $Wald$ statistics are the standard fixed-b distribution as in Kiefer and Vogelsang (2005). (3) We also investigate methods for obtaining fixed-b critical values with a focus on bootstrap methods. We show that the results of Gonçalves and Vogelsang (2011) extend to both the AM and ES approaches. The naive *i.i.d.* bootstrap is a particularly good option for obtaining valid fixed-b critical values especially if the dates of the missing observations are taken as given when generating bootstrap samples.

It is important to draw similarities and differences between the results in this paper and recent work by Bester et al. (2016) who obtained fixed-b results in spatial settings. In spatial settings the sample region can have varied shapes and the shape can affect the sampling distribution of HAR tests. Bester et al. (2016) show that the shape of the sampling region is captured by fixed-b theory. This is similar to our result where the fixed-b limit depends on the dates (locations) of missing data when the missing process is non-random. One might wonder whether the results of Bester et al. (2016) can be directly applied to missing data in time series settings because Bester et al. (2016) obtain results using random field theory and a time series process is a one-dimensional random field. The answer is no because Bester et al. (2016) restrict attention to two-dimensional sampling regions with quadrant-wise monotonic boundaries which rules out holes in the sampling region. Missing observations generate holes in the sampling region and therefore the results in Bester et al. (2016) do not directly apply.

The rest of the paper is organized as follows. Section 2 defines the models, estimators, and relevant test statistics. Section 3 develops fixed-b asymptotic results for the estimators and the test statistics defined in Section 2. Section 4 discusses simulation of the asymptotic critical values with a focus on bootstrap methods. Finite sample performance is examined in Section 5 by Monte Carlo simulations. Attention is focused on the relative performance of simulated asymptotic critical values compared with bootstrap critical values and the relative performance of the AM approach compared with the ES approach. Practical recommendations are given. Section 6 concludes. Formal proofs for the fixed-b results are given in Appendices A-B for the AM and ES approaches respectively.

2 MODELS AND TEST STATISTICS

2.1 Models

Suppose that a researcher is interested in the regression model

$$y_t^* = x_t^{*'}\beta + u_t^*, \qquad (t = 1, 2, \ldots, T), \tag{1}$$

where y_t^* (scalar) and x_t^* ($k \times 1$ vector) are latent variables that are not fully observed in the missing data case. Suppose that the $k \times 1$ vector of parameters, β, is identified by the moment condition

$$E(m(w_t^*; \beta)) = 0, \tag{2}$$

where $w_t^{*\prime} = (y_t^*, x_t^{*'})$ and

$$m\left(w_t^*;\beta\right) = x_t^*(y_t^* - x_t^{*'}\beta) = x_t^* u_t^*.$$

Let $\{a_t\}$ be a *missing process* where

$$a_t = \begin{cases} 1 & \text{when all } g \text{ elements in } w_t^* \text{ are observed at time } t \\ 0 & \text{otherwise.} \end{cases}$$

If we assume that the mechanism generating the missing data does not generate an endogeneity problem, then the moment condition (2) above implies

$$E(a_t m\left(w_t^*;\beta\right)) = E\left(a_t x_t^*(y_t^* - x_t^{*'}\beta)\right) = E\left(a_t x_t^* u_t^*\right) = 0, \tag{3}$$

which puts zero weight on time period t whenever at least one element of w_t^* is not observed.[2] We maintain assumption (3) throughout the paper so that we can focus on the impact of missing data on HAR inference in settings where β is identified by the observed data.

We consider two regression models, the AM and ES regression models, which are naturally related to (3). The AM regression model, proposed by Parzen (1963),[3] is given by

$$y_t = x_t'\beta + u_t, \quad (t = 1, 2, \ldots, T), \tag{4}$$

where $y_t = a_t y_t^*$, $x_t = a_t x_t^*$, and $u_t = a_t u_t^*$. The ordinary least squares (OLS) moment condition, $E(x_t(y_t - x_t'\beta)) = 0$, is easily seen to be the same as (3) using the fact that $a_t^2 = a_t$. Practically speaking, zeros are plugged in for y_t^* and x_t^* for any time period that has a missing observation in y_t^* and/or any element of x_t^*. Because the zeros act as place holders for missing observations, the true time distances between observations are preserved. At a conceptual level this would appear important when using HAR variance estimators to obtain robust standard errors.

The ES regression model is obtained by ignoring time periods with missing data and re-indexing the

---

[2]As pointed out by the referees, applying the missing process to the moment condition rather than the regression model itself allows x_t to contain lags of y_t. However, we do not explicitly consider dynamic models in this paper because either dynamic completeness holds in which case the HAR variance estimator would not be required or dynamic completeness fails in which case (3) does not hold and β is not identified.

[3]Parzen (1963) adopted the label, *amplitude modulated series*, because the original time series are amplitude modulated by the missing process $\{a_t\}$.

observed time periods.[4] The observed time series processes for the ES regression model are $\{y_\tau^{ES} = y_t^*, x_\tau^{ES} = x_t^*; \quad \tau = \sum_{s=1}^{t} a_s\}_{\tau=1}^{T_{ES}}$, where $T_{ES} = \sum_{t=1}^{T} a_t$ is the number of time periods with no missing data (see Figure 1 below). This leads to the ES regression model

$$y_\tau^{ES} = x_\tau^{ES'}\beta + u_\tau^{ES}, \qquad (\tau = 1, 2, \ldots, T_{ES})\,. \tag{5}$$

Similar to the AM regression model, the OLS moment condition for (5) is equivalent to (3) with the missing time periods ignored and observed time periods reindexed. Because the ES approach changes the time index, time distances between observations become skewed for many pairs of time periods when computing an HAR variance estimator. For example, in Figure 1, w_5^{ES} and w_6^{ES}, which are 5 time periods apart, are treated as if they are a single time period away.

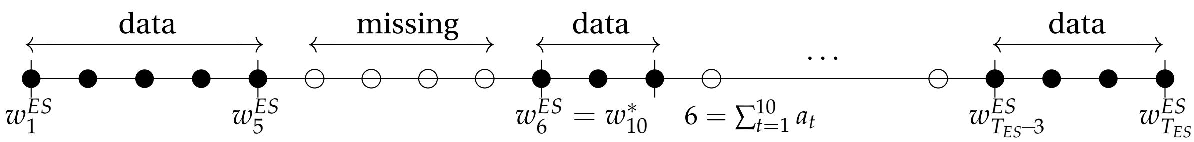


Figure 1: Equal Space Regression Model

In fact, in terms of estimating β, there is no difference between the AM and ES regression models. Let $\hat{\beta}$ be obtained by the sample analog to (3):

$$\frac{1}{T}\sum_{t=1}^{T} a_t x_t^*(y_t^* - x_t^{*'}\hat{\beta}) = 0.$$

Because $\sum a_t x_t^*(y_t^* - x_t^{*'}\hat{\beta}) = 0$ can be equivalently written as

$$\sum_{t=1}^{T} a_t x_t^*(y_t^* - x_t^{*'}\hat{\beta}) = \sum_{t=1}^{T} x_t(y_t - x_t'\hat{\beta}) = \sum_{\tau=1}^{T_{ES}} x_\tau^{ES}(y_\tau^{ES} - x_\tau^{ES'}\hat{\beta}) = 0,$$

it follows that $\hat{\beta}$ is the OLS estimator in both the AM and ES regression models.[5]

[4]The label 'equal space' was adopted by Datta and Du (2012) because the observations in the ES regression are adjacent to each other after reindexing time.

[5]The equivalence between the AM and ES approaches for parameter estimation could extend to models with more general moment conditions that are exactly identified. However, in the case of over-identified models, this equivalence will not hold in general. For example it will break down if the weighting matrix used for generalized method of moments (GMM) estimator is based on a nonparametric kernel covariance matrix estimator because the reindexing of time in the ES approach would make the covariance matrix estimators in the AM and ES approaches different. Using the fixed smoothing theory for GMM developed by Sun (2014a), it should be possible to extend the results we obtain here to a more general GMM setting, but such an extension is nontrivial and is beyond the current scope of this paper.

2.2 Test Statistics

Our focus is inference regarding β based on the common OLS estimator for the AM and ES regression models,

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} x_t y_t = \left(\sum_{\tau=1}^{T_{ES}} x_\tau^{ES} x_\tau^{ES\prime} \right)^{-1} \sum_{\tau=1}^{T_{ES}} x_\tau^{ES} y_\tau^{ES}.$$

Inference is carried out using nonparametric kernel covariance matrix estimators. Consider testing the null hypothesis, $H_0 : r(\beta) = 0$, against the alternative, $H_A : r(\beta) \neq 0$, where $r(\beta)$ is a $q \times 1$ vector $(q \leq k)$ of continuously differentiable functions with a first derivative matrix $R(\beta) = \partial r(\beta) / \partial \beta'$ of rank q.

The OLS residuals for the AM and ES regression models are given by $\hat{u}_t = y_t - x_t' \hat{\beta}$ and $\hat{u}_\tau^{ES} = y_\tau^{ES} - x_\tau^{ES\prime} \hat{\beta}$, respectively. The nonparametric kernel covariance matrix estimators for the AM and ES regressions are given by

$$\hat{\Omega} = \hat{\Gamma}_0 + \sum_{j=1}^{T-1} \mathcal{K}\left(\frac{j}{M} \right) \left(\hat{\Gamma}_j + \hat{\Gamma}_j' \right) = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathcal{K}\left(\frac{t-s}{M} \right) \hat{v}_t \hat{v}_s'$$

and

$$\hat{\Omega}_{ES} = \hat{\Gamma}_0^{ES} + \sum_{j=1}^{T_{ES}-1} \mathcal{K}\left(\frac{j}{M_{ES}} \right) \left(\hat{\Gamma}_j^{ES} + \hat{\Gamma}_j^{ES\prime} \right) = \frac{1}{T_{ES}} \sum_{\tau=1}^{T_{ES}} \sum_{\nu=1}^{T_{ES}} \mathcal{K}\left(\frac{\tau - \nu}{M_{ES}} \right) \hat{v}_\tau^{ES} \hat{v}_\nu^{ES\prime},$$

where $\hat{\Gamma}_j = T^{-1} \sum_{t=j+1}^{T} \hat{v}_t \hat{v}_{t-j}'$, $0 \leq j \leq T-1$, are the sample autocovariances of $\hat{v}_t = x_t \hat{u}_t$ and $\hat{\Gamma}_j^{ES} = T_{ES}^{-1} \sum_{\tau=j+1}^{T_{ES}} \hat{v}_\tau^{ES} \hat{v}_{\tau-j}^{ES\prime}$, $0 \leq j \leq T_{ES} - 1$, are the sample autocovariances of $\hat{v}_\tau^{ES} = x_\tau^{ES} \hat{u}_\tau^{ES}$. $\mathcal{K}(x)$ is a kernel function. M and M_{ES} are the bandwidths. We denote the bandwidth of the ES regression model by M_{ES} so as not to confuse with the bandwidth used in the AM case.

While the two variance estimators are different, $\hat{\Omega}_{ES}$ can be rewritten as a weighted sum of $\hat{v}_t \hat{v}_s'$ instead of $\hat{v}_\tau^{ES} \hat{v}_\nu^{ES\prime}$. First note that the elements of $\{\hat{v}_\tau^{ES} \hat{v}_\nu^{ES\prime}\}$ comprise the non-zero subset of elements of $\{\hat{v}_t \hat{v}_s'\}$ with the remaining elements of $\{\hat{v}_t \hat{v}_s'\}$ being zeros. Second, note that the distance between the t^{th} and s^{th} observations in the AM regression model converts to $|\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i|$ in the ES regression model which is the number of observed data between time periods t and s. Therefore $\hat{\Omega}_{ES}$ can be equivalently written as

$$\hat{\Omega}_{ES} = \frac{1}{T_{ES}} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathcal{K}\left(\frac{\sum_{i=1}^{t} a_i - \sum_{i=1}^{s} a_i}{M_{ES}} \right) \hat{v}_t \hat{v}_s'. \tag{6}$$

Using the OLS estimator and the robust variance estimators, the HAR *Wald* statistics are defined as

$$
\begin{aligned}
W_T &= Tr\left(\hat{\beta}\right)'\left[R\left(\hat{\beta}\right)\hat{Q}^{-1}\hat{\Omega}\hat{Q}^{-1}R\left(\hat{\beta}\right)'\right]^{-1}r\left(\hat{\beta}\right) \quad \text{and} \\
W_T^{ES} &= T_{ES}r\left(\hat{\beta}\right)'\left[R\left(\hat{\beta}\right)\hat{Q}_{ES}^{-1}\hat{\Omega}_{ES}\hat{Q}_{ES}^{-1}R\left(\hat{\beta}\right)'\right]^{-1}r\left(\hat{\beta}\right)
\end{aligned}
\tag{7}
$$

for the AM and ES regressions respectively, where $\hat{Q} = T^{-1}\sum_{t=1}^{T}x_t x_t'$ and $\hat{Q}_{ES} = T_{ES}^{-1}\sum_{\tau=1}^{T_{ES}}x_\tau^{ES}x_\tau^{ES\prime}$. For the single restriction ($q = 1$) case, the corresponding t-statistics are

$$
t_T = \frac{\sqrt{T}r\left(\hat{\beta}\right)}{\sqrt{R\left(\hat{\beta}\right)\hat{Q}^{-1}\hat{\Omega}\hat{Q}^{-1}R\left(\hat{\beta}\right)'}} \quad \text{and} \quad t_T^{ES} = \frac{\sqrt{T_{ES}}r\left(\hat{\beta}\right)}{\sqrt{R\left(\hat{\beta}\right)\hat{Q}_{ES}^{-1}\hat{\Omega}_{ES}\hat{Q}_{ES}^{-1}R\left(\hat{\beta}\right)'}}. \tag{8}
$$

To better compare W_T^{ES} with W_T, rewrite W_T^{ES} using $\hat{Q}_{ES} = \left(T/T_{ES}\right)\hat{Q}$ as

$$
W_T^{ES} = Tr\left(\hat{\beta}\right)'\left[R\left(\hat{\beta}\right)\hat{Q}^{-1}\left(\frac{T_{ES}}{T}\hat{\Omega}_{ES}\right)\hat{Q}^{-1}R\left(\hat{\beta}\right)'\right]^{-1}r\left(\hat{\beta}\right). \tag{9}
$$

It immediately follows that W_T^{ES} in (9) has the exact same form as W_T in (7) with $\left(T_{ES}/T\right)\hat{\Omega}_{ES}$ used in place of $\hat{\Omega}$. Furthermore, using (6), W_T^{ES} becomes

$$
\begin{aligned}
W_T^{ES} &= Tr\left(\hat{\beta}\right)'\left[R\left(\hat{\beta}\right)\hat{Q}^{-1}\hat{\Omega}_{\widetilde{\mathcal{K}}}\hat{Q}^{-1}R\left(\hat{\beta}\right)'\right]^{-1}r\left(\hat{\beta}\right), \\
\text{where} \quad \hat{\Omega}_{\widetilde{\mathcal{K}}} &\equiv \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathcal{K}\left(\frac{\sum_{i=1}^{t}a_i - \sum_{i=1}^{s}a_i}{M_{ES}}\right)\hat{v}_t\hat{v}_s'.
\end{aligned}
$$

Hence, in terms of the test statistics, choosing between the AM approach and the ES approach boils down to a choice of kernel function between $\mathcal{K}\left((t-s)/M\right)$ and $\widetilde{\mathcal{K}}(t,s,M_{ES}) \equiv \mathcal{K}\left(\left(\sum_{i=1}^{t}a_i - \sum_{i=1}^{s}a_i\right)/M_{ES}\right)$ when computing the HAR variance estimator. Suppose the same kernel and bandwidth are used for the AM and ES approaches with $M = M_{ES}$. Then

$$
\frac{\left|\sum_{i=1}^{t}a_i - \sum_{i=1}^{s}a_i\right|}{M_{ES}} \leq \frac{|t-s|}{M},
$$

in which case it follows that less down-weighting on $\hat{v}_t\hat{v}_s'$ is used in the ES approach because the ES regression model ignores the gaps created by the missing observations. The bottom line is that, for a given kernel and bandwidth, the ES *Wald* statistic has an equivalent AM *Wald* statistic where the HAR variance estimator is based on a weighting scheme that is a transformed version of the ES weighting scheme.

3 ASSUMPTIONS AND ASYMPTOTIC THEORY

In this section, we derive the asymptotic behavior of the OLS estimator, the HAR variance estimators, and the *Wald* statistics defined in Section 2. Our primary focus is on obtaining fixed-b asymptotic results. Throughout, the symbol "$\Rightarrow$" denotes weak convergence of a sequence of stochastic processes to a limiting stochastic process and $\mathcal{W}_q(r)$ denotes a $q \times 1$ vector of independent standard Wiener process. Following Sun (2014b) we make the following assumption on the kernel function.

Assumption $\mathcal{K}$. *The kernel function $\mathcal{K}(x) : \mathbb{R} \rightarrow [0, 1]$ is symmetric, piecewise smooth with $\mathcal{K}(0) = 1$ and $\int_0^\infty \mathcal{K}(x)xdx < \infty$.*

Assumption $\mathcal{K}$ imposes mild conditions on the kernel function and most of the kernel functions used in practice - including Bartlett, Parzen, and QS kernels - satisfy this assumption. With Assumption $\mathcal{K}$, $K_b(r, s) \equiv \mathcal{K}_b(r - s) = \mathcal{K}((r - s)/b)$ defined on $[0, 1] \times [0, 1]$ for every given b is symmetric and integrable in $L^2([0, 1] \times [0, 1])$ and we can expand $K_b(r, s)$ as $K_b(r, s) = \sum_{n=1}^{\infty} \nu_n f_n(r) f_n(s)$. Here ν_n is an eigenvalue of the kernel and $f_n(s)$ is the corresponding eigenfunction so that $\nu_n f_n(s) = \int_0^1 K_b(r, s) f_n(r)dr$. See Sun (2014b) for details.

3.1 Random Missing Process

In this section, we briefly discuss the random missing process case. We only sketch some results given that our main interest is the non-random missing process case. Suppose $\{a_t\}$ is a random process such that (3) holds. A sufficient condition for (3) is that $\{a_t\}$ is independent from the latent processes $\{(u_t^*, x_t^{*\prime})'\}$. The weaker condition such as $E(u_t^*|a_t, x_t^*) = 0$ with $E(u_t^*|x_t^*) = 0$ is also sufficient.

When the missing process is random, the asymptotic theory is driven by the observed AM series. Suppose the observed AM series satisfies the following for $r \in [0, 1]$: (a) $T^{-1} \sum_{t=1}^{[rT]} x_t x_t' \Rightarrow rQ$, uniformly in r with Q full rank, and (b) $T^{-1/2} \sum_{t=1}^{[rT]} v_t \Rightarrow \Lambda \mathcal{W}_k(r)$ where Λ is the matrix square root of the long run variance matrix of v_t. Then the HAR variance estimator and *Wald* statistics of the AM series have the usual fixed-b asymptotic limits as obtained by Kiefer and Vogelsang (2005). These conditions would be satisfied if the latent and missing processes are jointly weakly dependent with proper moment conditions. A set of primitive conditions on missing and latent processes that are sufficient for these high level assumptions are lengthy but standard. Interested readers can refer the Supplemental Appendix available on our website.[6]

For the ES approach, suppose the ES regression model satisfies the following for $r \in [0,1]$: (a) $T_{ES}^{-1}\sum_{\tau=1}^{[rT_{ES}]} x_\tau x_\tau' \Rightarrow rQ_{ES}$, uniformly in r with Q_{ES} full rank, and (b) $T_{ES}^{-1/2}\sum_{\tau=1}^{[rT_{ES}]} v_\tau^{ES} \Rightarrow \Lambda_{ES}\mathcal{W}_k(r)$ where Λ_{ES} is the matrix square root of the long run variance matrix of v_τ^{ES}. Then the HAR variance estimator and $Wald$ test statistics of the ES regression model have the usual fixed-b asymptotic limits as in Kiefer and Vogelsang (2005). Because the ES processes have less dependence than the latent processes, it is reasonable to conjecture that ES processes will satisfy conditions (a) and (b) above whenever the latent processes satisfy conditions analogous to (a) and (b). This is trivially true when the latent processes are white noise processes. In other words, if the latent processes satisfy conditions required for fixed-b theory then the asymptotic fixed-b distribution for the ES approach would be identical to the usual fixed-b distribution.

Hence, from a fixed-b perspective, the random missing process case is rather trivial when the missing processes are weakly dependent and satisfy the key assumptions for obtaining a fixed-b limit. The theory is more interesting when $\{a_t\}$ is viewed as a non-random process.

3.2 Non-random Missing Process

In this section, we analyze the asymptotic behavior of the OLS estimator and the HAR test statistics defined in Section 2 under the assumption that the missing process is non-random. The theory developed in this section is used to obtain an asymptotic reference distribution for the HAR test statistics that depends on the locations of the missing data in a given sample. We begin by describing the timing of the missing observations for an arbitrary data set.

Definition 1. ***[Timing of the missing observations]*** *We characterize an arbitrary data set with missing observations as follows. From $t = 1$ to $t = T_1$ we observe data, from $t = T_1 + 1$ to $t = T_2$ data are missing, from $t = T_2 + 1$ to $t = T_3$ we observe data and so forth. Let the number of missing clusters be C. For simplicity, we assume that data are observed at $t = 1$ and $t = T$.[7] Thus, in general, from $t = T_{2n-1} + 1$ to*

[6] http://msu.edu/∼ tjv/working.html

[7] This assumption is only for notational simplicity. The results of this paper go through without this assumption.

$t = T_{2n}$ data are missing whereas from $t = T_{2n} + 1$ to $t = T_{2n+1}$ data are observed, $n = 1, \ldots, C$ (see Figure 2). For notational purposes, let $T_0 = 0$ and $T_{2C+1} = T$.

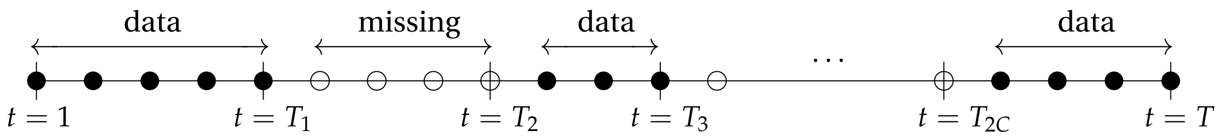


Figure 2: Timing of the data with missing observations

When the missing process is non-random, missing locations are non-random, and hence the finite sample behavior of the estimators and statistics will depend on the locations of missing observations. To capture this dependence on missing locations in the asymptotic theory, we map the missing and observed cutoff dates into points on $[0, 1]$ based on the proportion of time periods in the sample that occur up to a particular cutoff date, T_i, by defining $\lambda_i = T_i/T$ for $i = 0, 1, \ldots, 2C + 1$. By definition, $T_0 = 0$ and $T_{2C+1} = T$ and it follows that $\lambda_0 = 0$ and $\lambda_{2C+1} = 1$.

Once a researcher obtains the data, the number of missing clusters and the locations of the missing observations are known. Since these locations are defined within the obtained sample with time span T, they are fully characterized by the set of ratios $\{\lambda_1, \ldots, \lambda_{2C}\}$ which we denote as $\{\lambda_i\}$. In order to obtain useful asymptotic results that capture the locations of the cutoff dates, we treat $\{\lambda_i\}$ and C fixed as $T \to \infty$, which is stated below.

Assumption fixed-λ. *Let the missing and observed cutoff dates be $\{T_i\}$ and C be the number of missing clusters as defined in Definition 1. When the missing process is non-random, we let*

$$\lambda_i \equiv \frac{T_i}{T}, \quad i = 0, 1, \ldots, 2C + 1,$$

and $C < \infty$ be fixed as $T \to \infty$.

Intuitively, we are preserving the relative finite sample locations of the missing clusters when obtaining our asymptotic results. It is important to keep in mind that this asymptotic nesting is only used to obtain reference distributions for generating critical values for test statistics and is *not* meant to be a description of the way the data is gathered or generated. Following the suggestion of a referee, we label this asymptotic nesting the 'fixed-λ device'.[8] In some missing data settings, such as missing weekends for

[8]Conceptually, the fixed-λ device is the same as mapping a given number of break dates to corresponding breakpoints in a model with structural change and obtaining asymptotic results holding the breakpoints fixed as $T \to \infty$.

daily data, the assumption that C is fixed as $T \to \infty$ may not match the way data is gathered. It is important to keep in mind that what ultimately matters is the usefulness of the approximation delivered by a particular asymptotic theory. A good example is continuous records asymptotics. In particular, continuous record asymptotics uses an asymptotic nesting where the frequency of observation increases with the span of the data held fixed. This nesting may not match the way data is gathered if the frequency of observation cannot be increased and new data is only generated as the future evolves. Nonetheless, continuous records asymptotics can deliver useful reference distributions in this situation as shown by Phillips (1987, Section 5) and Perron (1991) in certain unit root settings where initial conditions matter in finite samples. Continuous records asymptotics can capture the impact of initial conditions whereas conventional asymptotics does not.

Suppose we allowed $C \to \infty$ as $T \to \infty$. For cases like missing weekends in daily data, by construction the collection of $\{T_i\}$ would be fixed as $T \to \infty$ in which case $\lambda_i \to 0$ for all i. While we have not worked out the details in this case, we conjecture that the fixed-b limit of the test statistics would either be degenerate or the same as the standard fixed-b limit. In either case, the asymptotics would not capture the impact of missing data on the finite sample distribution of the test statistic. Alternatively, one could allow $C \to \infty$ but retain the assumption that $T_i/T \to \lambda_i$ as $T \to \infty$. Because the proportion of missing data must be in the range $(0,1)$, the sequence $\{\lambda_i\}_{i=1}^{\infty}$ would need to converge to zero. This would imply that missing data becomes less frequent as T grows. That alone would not be a problem, per se, if the resulting limits for the test statistics deliver a useful reference distribution. However, the limits would depend on the infinite sequence $\{\lambda_i\}_{i=1}^{\infty}$ which would be unknown making the simulation of asymptotic critical values infeasible. In contrast, by treating C fixed in the asymptotics, we obtain a limit that depends on the observed finite number of missing dates that can easily be simulated.

Under the fixed-λ device, the asymptotic theory is driven by the latent process, and what matters is whether the latent process satisfies conditions required for fixed-b asymptotic theory. For this purpose the following high-level assumptions on the latent process are sufficient where we define $v_t^* = x_t^* u_t^*$.

Assumption LP.

1. *$T^{-1} \sum_{t=1}^{[rT]} x_t^* x_t^{*\prime} \Rightarrow rQ^*$, uniformly in $r \in [0,1]$, where $Q^* = E(x_t^* x_t^{*\prime})$ is full rank.*

2. *$T^{-1/2} \sum_{t=1}^{[rT]} v_t^* \Rightarrow \Lambda^* \mathcal{W}_k(r)$, $\forall r \in [0,1]$, where $\Lambda^* \Lambda^{*\prime} = \Omega^* = \Gamma_0^* + \sum_{j=1}^{\infty} (\Gamma_j^* + \Gamma_j^{*\prime})$, $\Gamma_j^* = E(v_t^* v_{t-j}^{*\prime})$.*

Assumption LP.1 states that a uniform (in r) law of large numbers (LLN) holds for $\{x_t^* x_t^{*\prime}\}$. Assumption LP.2 states that a functional central limit theorem (FCLT) holds for the scaled partial sums of $\{v_t^*\}$. (See Kiefer and Vogelsang (2005) for details. For primitive conditions that are sufficient for Assumption LP, see e.g., Gonçalves and Vogelsang (2011).)

Our theoretical results for the non-random missing process case are given below. Because the fixed-b asymptotic distributions depend on the kernels used to compute the HAR variance estimators, the random matrix that appears in the asymptotic results needs to be defined.

Definition 2. *Let the $h \times h$ random matrix, $P_{\mathcal{K}}(b, B_h)$, be defined as*

$$P_{\mathcal{K}}(b, B_h) \equiv \int_0^1 \int_0^1 \mathcal{K}\left(\frac{r-s}{b}\right) dB_h(r)\, dB_h(s)'.$$

We first state our results for the AM regression model followed by the ES regression model. The proof of Theorem 1 is provided in Appendix A. We use the following notation. Let $\phi(s) = \sum_{n=0}^{C} \mathbb{1}\{\lambda_{2n} < s \leq \lambda_{2n+1}\}$ and $r \wedge s = \min(r,s)$. Also λ denotes the proportion of the time periods in the sample where data are observed, i.e., $\lambda = \int_0^1 \phi(s) ds = {}^1/_T \sum_{t=1}^{T} a_t$.

Theorem 1. ***[Results for the AM series]***
Let $\breve{B}_k(r, \{\lambda_i\}) \equiv \int_0^r d\breve{B}_k(s, \{\lambda_i\})$, $r \in [0,1]$, where $d\breve{B}_k(s, \{\lambda_i\}) = \phi(s)(d\mathcal{W}_k(s) - ds\lambda^{-1}\overline{\mathcal{W}}_k)$ and $\overline{\mathcal{W}}_k \equiv \int_0^1 \phi(s) d\mathcal{W}_k(s)$. Under Assumptions $\mathcal{K}$, fixed-λ, and LP, the following hold as $T \to \infty$.

(a) [*Asymptotic Behavior of OLS*]

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \Rightarrow (\lambda Q^*)^{-1} \Lambda^* \overline{\mathcal{W}}_k = N\left(0, \lambda^{-1} Q^{*-1} \Omega^* Q^{*-1}\right)$$

(b) [*Fixed-b asymptotic limit of the HAR variance estimator*] *Assume $M = bT$ where $b \in (0,1]$ is fixed. Then,*

$$\hat{\Omega} \Rightarrow \Lambda^* P_{\mathcal{K}}\left(b, \breve{B}_k(\{\lambda_i\})\right) \Lambda^{*\prime},$$

where $P_{\mathcal{K}}(\cdot)$ is given by Definition 2.

(c) [*Fixed-b asymptotic distribution of W_T*] *Under H_0 and $M = bT$ where $b \in (0,1]$ is fixed,*

$$W_T \Rightarrow \overline{\mathcal{W}}_q' \left[P_{\mathcal{K}}\left(b, \breve{B}_q(\{\lambda_i\})\right)\right]^{-1} \overline{\mathcal{W}}_q,$$

and when $q = 1$,

$$t_T \Rightarrow \frac{\overline{\mathcal{W}}_1}{\sqrt{P_{\mathcal{K}}\left(b, \breve{B}_1(\{\lambda_i\})\right)}}.$$

□

According to Theorem 1(a) the asymptotic variance of the OLS estimator depends on Ω^*, the long run variance of the latent process, even though W_T is constructed using the AM series long run variance estimator $\hat{\Omega}$. This poses no problems for inference because the fixed-b limit of $\hat{\Omega}$ is proportional to Ω^* through Λ^* and $\Lambda^{*\prime}$ and Ω^* cancels from the limiting distributions of the test statistics. Theorem 1(b) shows that the fixed-b limit of $\hat{\Omega}$ depends on $P_{\mathcal{K}}(\cdot)$ which takes the same form as in the usual fixed-b limits in models with no missing data (Kiefer and Vogelsang (2005)). This is not surprising since test statistics based on the AM series preserve the true time distances between observations and the kernel weights stay the same even if there are missing observations. The key difference between the fixed-b limit of $\hat{\Omega}$ in Theorem 1(b) compared to that without missing observations is the form of $\breve{B}_k(r, \{\lambda_i\})$, the asymptotic limit of the partial sum $T^{-1/2}\sum_{t=1}^{[rT]}\hat{v}_t$. When there are missing observations, $\breve{B}_k(r, \{\lambda_i\})$ depends on the locations of the missing observations via the fixed-λ device and is not a Brownian bridge.[9] Hence the critical values generated by the reference distributions given by Theorem 1(c) depend on the locations of the missing data and are different from the standard fixed-b critical values.[10]

Next, we state our result for the ES regression model. The proof for Theorem 2 is given in Appendix B.

Theorem 2. ***[Results for the ES regression model]***
Define the $k \times 1$ *stochastic process* $\dddot{\mathcal{W}}_k(r) = \lambda^{-1/2}\int_0^{(r\lambda+\sum_{k=1}^{2n}(-1)^k\lambda_k)}\phi(u)d\mathcal{W}_k(u)$ *for* $n \in \{0, \ldots, C\}$ *such that* $r\lambda + \sum_{k=1}^{2n}(-1)^k\lambda_k \in (\lambda_{2n}, \lambda_{2n+1}]$ *and let* $\widetilde{\dddot{\mathcal{W}}}_k(r) = \dddot{\mathcal{W}}_k(r) - r\dddot{\mathcal{W}}_k(1)$, $r \in [0,1]$. *Under Assumptions* $\mathcal{K}$, *fixed-*λ, *and LP, the following hold as* $T \to \infty$.

(a) [*Asymptotic Behavior of OLS*]

$$\sqrt{T_{ES}}\left(\hat{\beta} - \beta\right) \Rightarrow Q^{*-1}\Lambda^*\dddot{\mathcal{W}}_k(1) = N\left(0, Q^{*-1}\Omega^* Q^{*-1}\right)$$

[9] This is shown formally in Appendix A.
[10] Similar to Kiefer and Vogelsang (2005), $\breve{B}_k(r, \{\lambda_i\})$ is independent of $\overline{\mathcal{W}}_k$. See Appendix A.

(b) $[$*Fixed-b asymptotic limit of* $\hat{\Omega}_{ES}$ $]$ *Let* $M_{ES} = bT_{ES}$ *where* $b \in (0,1]$ *is fixed. Then,*

$$\hat{\Omega}_{ES} \Rightarrow \Lambda^* P_{\mathcal{K}}\left(b, \widetilde{\dddot{\mathcal{W}}}_k\right)\Lambda^{*\prime},$$

where $P_{\mathcal{K}}(\cdot)$ *is defined via Definition 2.*

(c) $[$*Fixed-b asymptotic distribution of* W_T^{ES} $]$ *Let* $M_{ES} = bT_{ES}$ *where* $b \in (0,1]$ *is fixed. Under* H_0*,*

$$W_T^{ES} \Rightarrow \dddot{\mathcal{W}}_q(1)'\left[P_{\mathcal{K}}\left(b, \widetilde{\dddot{\mathcal{W}}}_q\right)\right]^{-1}\dddot{\mathcal{W}}_q(1),$$

and when $q = 1$*,*

$$t_T^{ES} \Rightarrow \frac{\dddot{\mathcal{W}}_1(1)}{\sqrt{P_{\mathcal{K}}\left(b, \widetilde{\dddot{\mathcal{W}}}_1\right)}}.$$

□

The asymptotic result for the OLS estimator given by Theorem 2(a) is identical to the result in Theorem 1(a) because the AM and ES regression models have the same OLS estimator. Here λ^{-1} is gone because the normalizing factor is $\sqrt{T_{ES}}$ instead of $\sqrt{T}$. What is not obvious is that the asymptotic distributions in Theorem 2(b) and (c) are in fact equivalent to the standard fixed-b asymptotic distributions as in Kiefer and Vogelsang (2005) with $b = {}^{M_{ES}}/_{T_{ES}}$. Thus, we obtain the surprising result that when the missing process is non-random, the HAR *Wald* test of the ES regression model has the usual fixed-b limit that does not depend on the missing locations. This result holds because it can be shown that $\dddot{\mathcal{W}}_k(r)$ is a vector of standard Wiener process (see Appendix B).

We can provide some intuition as to why the HAR statistics have the usual fixed-b limits in the ES regression even though the ES approach contaminates the timing of the dates where data is observed. Suppose the latent processes are *i.i.d.*. When the missing dates are ignored and dropped, we are left with a regression model with T_{ES} observations and no serial correlation in the data. Therefore, W_T^{ES}, the usual HAR *Wald* statistic computed with T_{ES} observations, has the usual fixed-b limit as in Kiefer and Vogelsang (2005) with $b = M_{ES}/T_{ES}$. Because fixed-b limits of HAR statistics naturally extend from the *i.i.d.* case to cases with dependence due to the proportionality of the numerator and denominator to the long run variance, the results obtained in Theorem 2(b) and (c) are less surprising given that they trivially hold in the *i.i.d.* case.

4 BOOTSTRAP CRITICAL VALUES

In Section 3, we showed that when the missing process is non-random and the AM regression is used, the HAR *Wald* statistic has a fixed-b asymptotic distribution that depends on the missing locations through $\{\lambda_i\}$ ('λ,fixed-b' distribution henceforth). We also showed that when the ES regression model is used, the HAR *Wald* test statistic has the usual fixed-b asymptotic distribution. Both the usual fixed-b and 'λ,fixed-b' distributions are non-standard but it is relatively straightforward to obtain the critical values by directly simulating from the asymptotic distributions because they are functions of Brownian motions.[11] A potentially more user-friendly method, especially for the 'λ,fixed-b' distribution, is to obtain critical values using the bootstrap.

In the fixed-b literature, Gonçalves and Vogelsang (2011) showed that the naive moving block bootstrap has the same limiting distribution as the fixed-b asymptotic distribution under suitable regularity conditions. Their result hinges on the idea that if the bootstrap samples satisfy the LLN and FCLT that are required for fixed-b theory to go through with the probability measure induced by the bootstrap resampling conditional on a realization of the original time series (denoted as $p^\bullet$ henceforth), then because the fixed-b asymptotic distribution depends on the kernel function and $b = M/T$ but is otherwise pivotal, the bootstrap statistic will have the same limiting distribution as the original test statistic. The regularity conditions are those such that bootstrap samples satisfy the LLN and FCLT in the bootstrap world.

When observations are missing, the fixed-b asymptotic distribution of the HAR *Wald* statistics depends on the kernel function and b. For the case of the 'λ,fixed-b' distribution, it also depends on $\{\lambda_i\}$. However, the asymptotic distribution is otherwise pivotal. Hence it is expected that similar results as in Gonçalves and Vogelsang (2011) would go through if the bootstrap samples for the AM series and the ES regression models satisfy Assumptions fixed-λ and LP under the probability measure $p^\bullet$. Assumption fixed-λ treats the locations of missing data as fixed within the sample, so we need a bootstrap resampling scheme that preserves the missing locations. This means that the moving block bootstrap with block length greater than one is not practical because blocks will shuffle the locations of the missing data upon resampling. Instead, the *i.i.d.* bootstrap is appropriate where bootstrap samples are created

[11] For the usual fixed-b distribution, Vogelsang (2012) has developed a numerical method for the easy computation of standard fixed-b critical values and p-values for any significance level when the Bartlett kernel is used. Kiefer and Vogelsang (2005) provide critical value functions for popular significance levels for a set of kernels.

by first resampling with replacement from the observed data and creating a bootstrap sample with the same missing locations as the original data.

Define $\omega_t = (y_t, x_t')'$, $t = 1, \ldots, T$, that collects the dependent and independent variables of the AM series. Among those T observations define the subset of only the observed data, which we denote $\widetilde{\omega}_t = (\widetilde{y}_t, \widetilde{x}_t')'$, $t = 1, \ldots, \widetilde{T}$, $\widetilde{T} = \sum_{t=1}^{T} a_t$. Using block length $l = 1$, resample $\widetilde{T}$ observations with replacement from $\widetilde{\omega}_t$ and obtain a bootstrap sample for the ES regression model, which we denote as $\widetilde{\omega}_t^\bullet = (\widetilde{y}_t^\bullet, \widetilde{x}_t^{\bullet\prime})'$, $t = 1, \ldots, \widetilde{T}$. Using $\widetilde{\omega}_t^\bullet$, construct the bootstrap sample for the AM series denoted as $\omega_t^\bullet = (y_t^\bullet, x_t^{\bullet\prime})'$, $t = 1, \ldots, T$, by filling in the observed locations with resampled data $\widetilde{\omega}_t^\bullet$ and the missing locations with zeros. This way we construct an *i.i.d.* bootstrap sample with missing locations that match those in the observed data. The naive bootstrap test statistic for the AM regression, $W_T^\bullet$, is computed with $\omega_t^\bullet$ as

$$W_T^\bullet = T \left(r(\hat{\beta}^\bullet) - r(\hat{\beta}) \right)' [R(\hat{\beta}^\bullet) \hat{Q}^{\bullet -1} \hat{\Omega}^\bullet \hat{Q}^{\bullet -1} R(\hat{\beta}^\bullet)']^{-1} \left(r(\hat{\beta}^\bullet) - r(\hat{\beta}) \right), \tag{10}$$

and the naive bootstrap test statistic for the ES regression model, $W_T^{ES\bullet}$, is computed with $\widetilde{\omega}_t^\bullet$ as

$$W_T^{ES\bullet} = \widetilde{T} \left(r(\hat{\beta}^\bullet) - r(\hat{\beta}) \right)' [R(\hat{\beta}^\bullet) \hat{Q}_{ES}^{\bullet -1} \hat{\Omega}_{ES}^\bullet \hat{Q}_{ES}^{\bullet -1} R(\hat{\beta}^\bullet)']^{-1} \left(r(\hat{\beta}^\bullet) - r(\hat{\beta}) \right), \tag{11}$$

where $\hat{\beta}^\bullet$ is the OLS estimator from the regression of $y_t^\bullet$ on $x_t^\bullet$, $\hat{Q}^\bullet = {}^1\!/_T \sum_{t=1}^{T} x_t^\bullet x_t^{\bullet\prime}$, $\hat{Q}_{ES}^\bullet = {}^1\!/_{\widetilde{T}} \sum_{\tau=1}^{\widetilde{T}} \widetilde{x}_\tau^\bullet \widetilde{x}_\tau^{\bullet\prime}$, $\hat{\Omega}^\bullet = {}^1\!/_T \sum_{t=1}^{T} \sum_{t=1}^{T} \mathcal{K}\left({}^{(t-s)}\!/_M \right) \hat{v}_t^\bullet \hat{v}_s^{\bullet\prime}$ where $\hat{v}_t^\bullet = x_t^\bullet (y_t^\bullet - x_t^{\bullet\prime} \hat{\beta}^\bullet)$, and $\hat{\Omega}_{ES}^\bullet = {}^1\!/_{\widetilde{T}} \sum_{\tau=1}^{\widetilde{T}} \sum_{\nu=1}^{\widetilde{T}} \mathcal{K}({}^{(\tau-\nu)}\!/_{M_{ES}}) \widehat{\widetilde{v}}_\tau^\bullet \widehat{\widetilde{v}}_\nu^{\bullet\prime}$ where $\widehat{\widetilde{v}}_\tau^\bullet = \widetilde{x}_\tau^\bullet (\widetilde{y}_\tau^\bullet - \widetilde{x}_\tau^{\bullet\prime} \hat{\beta}^\bullet)$. Note that bootstrap statistics use the same formula as W_T and W_T^{ES}, which is why this approach is called the *naive* bootstrap.

Because we resample from observed time periods only, this resampling can be thought of as resampling from the latent process $\omega_t^* \equiv (y_t^*, x_t^{*\prime})'$. We do not know the value of ω_t^* when $a_t = 0$ and thus we are resampling from $\widetilde{T}$ observations not the full number of time periods T. However, because the resampling is based on *i.i.d.* draws, this bootstrap sample has essentially the same properties as an *i.i.d.* bootstrap sample of the fully observed latent process. We could take another $T - \widetilde{T}$ independent draws from $\widetilde{\omega}_t$ and fill in the missing locations of $\widetilde{\omega}_t^\bullet$. Call this "filled-in" bootstrap sample $\omega_t^{*\bullet}$. Then by construction $\omega_t^\bullet = a_t \omega_t^{*\bullet}$ where $\omega_t^{*\bullet}$ can be viewed as a bootstrap sample from the latent process given the *i.i.d.* resampling. If the bootstrap sample, $\omega_t^{*\bullet}$, satisfies

$$\begin{aligned}
&\text{(a) } T^{-1}\sum_{t=1}^{[rT]} x_t^{*\bullet} x_t^{*\bullet\prime} \overset{p^\bullet}{\Rightarrow} rQ^{*\bullet} \quad \text{and} \\
&\text{(b) } T^{-1/2}\sum_{t=1}^{[rT]} v_t^{*\bullet} \overset{p^\bullet}{\Rightarrow} \Lambda^{*\bullet}\mathcal{W}_k(r)
\end{aligned} \tag{12}$$

for some $Q^{*\bullet}$ and $\Lambda^{*\bullet}$, Assumption LP is satisfied for the bootstrap samples with the probability measure $p^\bullet$. This implies that $W_T^{ES\bullet}$ has the usual fixed-b asymptotic distribution with the probability measure $p^\bullet$ according to Theorem 2 (c), and by Theorem 1 (c),

$$W_T^\bullet \overset{p^\bullet}{\Rightarrow} \overline{\mathcal{W}}_q^{\bullet\prime} P_{\mathcal{K}}(b, \breve{B}_q(\{\lambda_i^\bullet\}))\overline{\mathcal{W}}_q^\bullet,$$

with $\overline{\mathcal{W}}_q^\bullet = \int_0^1 \sum_{n=0}^{C^\bullet} \mathbb{1}\left\{\lambda_{2n}^\bullet < s \leq \lambda_{2n+1}^\bullet\right\} d\mathcal{W}_q(s)$ where $\{\lambda_m^\bullet\}_{m=0}^{2C^\bullet+1}$ are the missing locations in the bootstrap sample and $C^\bullet$ is the number of missing clusters in the bootstrap sample. Because the missing locations of the bootstrap samples are configured to be identical to the missing locations of the data, it follows that $\lambda_j^\bullet = \lambda_j$ and $C^\bullet = C$. Therefore,

$$W_T^\bullet \overset{p^\bullet}{\Rightarrow} \overline{\mathcal{W}}_q^{\prime} P_{\mathcal{K}}(b, \breve{B}_q(\{\lambda_i\}))\overline{\mathcal{W}}_q,$$

which is the same 'λ,fixed-b' limit as in Theorem 1(c).

These asymptotic equivalences are mainly due to the fact that the limiting distributions in Theorems 1(c) and 2(c) are pivotal with respect to Λ^* and Q^* so that $W_T^\bullet$ and $W_T^{ES\bullet}$ have asymptotic distributions equivalent to those of W_T and W_T^{ES} respectively even though $\Lambda^{*\bullet}$ and $Q^{*\bullet}$ are potentially different from Λ^* and Q^*. Hence conditions (a) and (b) in (12) are crucial for the asymptotic equivalence of the naive *i.i.d.* bootstrap and fixed-b asymptotic distribution for both the AM series and the ES regression models. For primitive conditions that are sufficient for conditions (a) and (b) in (12), see Gonçalves and Vogelsang (2011). Here the results of Gonçalves and Vogelsang (2011) directly apply as long as the block length is one because these assumptions are made about the latent process which has nothing to do with the missing process. We do not restate the conditions here.[12]

[12] Gonçalves and Vogelsang (2011) showed that the naive moving block bootstrap has the same limiting distribution as the fixed-b asymptotic distribution under regularity conditions which include that the block length $l = o(\sqrt{T})$. Therefore, the results of Gonçalves and Vogelsang (2011) indicate that the conditions (a) and (b) in (12) are satisfied with *i.i.d.* bootstrap method.

Finally, we briefly consider the bootstrap critical values for the random missing process case. When the missing process is random, the HAR *Wald* statistics for both the AM series and ES regression models have the usual fixed-b asymptotic distribution as sketched in Section 3.1. If (12) is satisfied for the ES regression model, and (a) $T^{-1}\sum_{t=1}^{[rT]} x_t^{\bullet} x_t^{\bullet\prime} \overset{p^{\bullet}}{\Rightarrow} rQ^{\bullet}$ and (b) $T^{-1/2}\sum_{t=1}^{[rT]} v_t^{\bullet} \overset{p^{\bullet}}{\Rightarrow} \Lambda^{\bullet}\mathcal{W}_k(r)$ are satisfied for the AM regression model, then the moving block bootstrap would have the same limiting distribution as the usual fixed-b distribution. Hence the bootstrap critical values can be obtained by computing the bootstrap test statistics as in (10) and (11) with resamples from the observed AM series and ES regression models respectively. This in fact is identical to how one would normally proceed either with the AM regression model or the ES regression model as if no observations are missing. Because the missing locations are not treated as fixed for the random missing process case, the locations of missing observations need not be preserved for the AM series bootstrap samples and also the block length need not be one for both the AM series and ES regression models. The primitive conditions are lengthy but standard. Interested readers can refer to the Supplemental Appendix which is available on our website.

5 FINITE SAMPLE PERFORMANCE

In this section we use Monte Carlo simulations to evaluate the finite sample performance of the asymptotic approximations of the HAR *Wald* tests for the AM series and ES regression models. Our main focus is on the non-random missing process case, but we also provide some results for the random missing process cases.

5.1 Data Generating Process (DGP)

We consider a simple linear regression model for the latent processes given by,

$$
\begin{aligned}
&y_t^* = \alpha + \beta x_t^* + u_t^*,\\
&x_t^* = \rho x_{t-1}^* + \sqrt{1-\rho^2}\varepsilon_t^{x*},\\
&u_t^* = \rho u_{t-1}^* + \sqrt{1-\rho^2}\varepsilon_t^{u*},\\
&\varepsilon_t^{x*} \sim i.i.d.N\left(0,1\right),\ \varepsilon_t^{u*} \sim i.i.d.N\left(0,1\right),\\
&\qquad \{\varepsilon_t^{x*}\} \text{ independent of } \{\varepsilon_t^{u*}\},\\
&u_1^* = 0,\ x_1^* = 0,
\end{aligned}
$$

with $t = 1, 2, ..., T$ so that T is the time span. We set $\alpha = \beta = 0$ and $\rho \in \{0, 0.3, 0.6, 0.9\}$. For the random missing process case, we model $\{a_t\}$ as a Bernoulli(p) process, i.e. $P(a_t = 1) = p$, with $p \in \{0.3, 0.7\}$. Hence Bernoulli(0.7) implies that 30% of the data are missing. We provide results for the time spans $T \in \{50, 100\}$. For the non-random missing process case, we consider cases where data are missing in two clusters ($C = 2$) configured to match the durations of World War I (from 1914 to 1918) and World War II (from 1939 to 1945). We generate data both yearly and quarterly where the time span in years is from 1911 to 1946. For yearly data, this means that 12 observations are missing out of $T = 36$ time periods. For quarterly data, 48 observations are missing out of $T = 144$ time periods. Therefore, the missing process is configured as $a_{[rT]} = 0$ when $r \in (\lambda_1, \lambda_2] \cup (\lambda_3, \lambda_4]$ and $a_{[rT]} = 1$ otherwise, with $\lambda_1 = 1/12$, $\lambda_2 = 2/9$, $\lambda_3 = 7/9$, and $\lambda_4 = 35/36$ for both yearly and quarterly data. See Figure 3 for the yearly data case. For practical comparisons between quarterly and yearly data, we generate yearly data by first generating quarterly data and taking every fourth observation. Hence, using ρ at the quarterly frequency implies that the correlation between yearly observations is ρ^4.

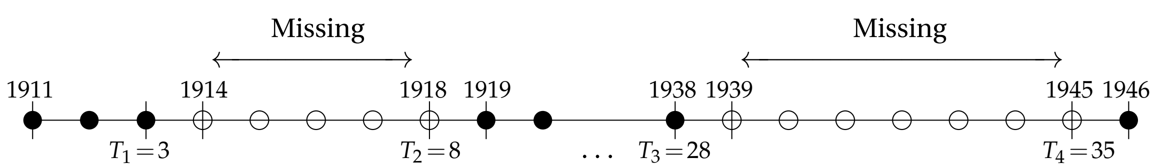


Figure 3: Missing due to World War I and World War II : Yearly data

With the latent and the missing processes generated above we construct the AM regression model

$$y_t = a_t y_t^* = \alpha a_t + \beta x_t + u_t, \quad x_t = a_t x_t^*, \; u_t = a_t u_t^* \quad (t = 1, \ldots, T),$$

and the ES regression model

$$y_\tau^{ES} = \alpha + \beta x_\tau^{ES} + u_\tau^{ES} \quad (\tau = 1, \ldots, T_{ES}), \quad T_{ES} = \sum_{t=1}^{T} a_t.$$

5.2 Test Statistics and Critical Values

We consider testing the null hypothesis, $\beta = 0$, against the alternative, $\beta \neq 0$. Let $\overline{y_{ES}} = \sum_{\tau=1}^{T_{ES}} y_\tau^{ES} / T_{ES}$ and $\overline{x_{ES}} = \sum_{\tau=1}^{T_{ES}} x_\tau^{ES} / T_{ES}$. Define $\ddot{y}_t = y_t - \overline{y_{ES}} a_t$, $\ddot{x}_t = x_t - \overline{x_{ES}} a_t$, $\ddot{y}_\tau^{ES} = y_\tau^{ES} - \overline{y_{ES}}$, and $\ddot{x}_\tau^{ES} = x_\tau^{ES} - \overline{x_{ES}}$.

Following (8) in Section 2, the HAR t-statistics for β are

$$t_T = \frac{\hat{\beta}}{\sqrt{T\left(\sum_{t=1}^{T} \ddot{x}_t^2\right)^{-2} \widehat{\Omega}}} \quad \text{and} \quad t_T^{ES} = \frac{\hat{\beta}}{\sqrt{T_{ES}\left(\sum_{\tau=1}^{T_{ES}} \ddot{x}_\tau^{ES\,2}\right)^{-2} \widehat{\Omega}_{ES}}},$$

where

$$\hat{\beta} = \left(\sum_{t=1}^{T} \ddot{x}_t^2\right)^{-1} \sum_{t=1}^{T} \ddot{x}_t y_t = \left(\sum_{\tau=1}^{T_{ES}} \ddot{x}_\tau^{ES\,2}\right)^{-1} \sum_{\tau=1}^{T_{ES}} \ddot{x}_\tau^{ES} y_\tau^{ES},$$

$$\hat{\Omega} = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T} \mathcal{K}\left(\frac{|i-j|}{M}\right) \hat{v}_i \hat{v}_j, \quad \hat{v}_t = \ddot{x}_t\left(\ddot{y}_t - \hat{\beta}\ddot{x}_t\right), \quad \text{and}$$

$$\hat{\Omega}_{ES} = \frac{1}{T_{ES}}\sum_{i=1}^{T_{ES}}\sum_{j=1}^{T_{ES}} \mathcal{K}\left(\frac{|i-j|}{M_{ES}}\right) \hat{v}_i^{ES} \hat{v}_j^{ES}, \quad \hat{v}_\tau^{ES} = \ddot{x}_\tau^{ES}(\ddot{y}_\tau^{ES} - \hat{\beta}\ddot{x}_\tau^{ES}).$$

For the World War missing case, we use bandwidth sample size ratios $b \in \{1/12, 2/12, \ldots, 11/12, 1\}$ which corresponds to bandwidths $M \in \{3, 6, 9, \ldots, 36\}$ and $M_{ES} \in \{2, 4, 6, \ldots, 24\}$ for the yearly data and $M \in \{12, 24, 36, \ldots, 144\}$ and $M_{ES} \in \{8, 16, 24, \ldots, 96\}$ for the quarterly data. For the Bernoulli missing case, we use $b \in \{0.02, 0.1, 0.2, \ldots, 0.9, 1\}$ giving $M \in \{1, 5, 10, \ldots, 45, 50\}$ for $T = 50$ and $M \in \{2, 10, 20, \ldots, 90, 100\}$ for $T = 100$. For the ES regression model we use $M_{ES} = [bT_{ES}]$. Because T_{ES} changes from one iteration to another, M_{ES} changes from one iteration to another given the same b. We use the Bartlett and quadratic spectral (QS) kernels.

We provide results using simulated asymptotic critical values and bootstrap critical values. The labels *AM: fixed-b* and *ES:fixed-b* denote the usual fixed-b asymptotic critical values whereas *AM:λ,fixed-b* denotes asymptotic critical values based on the reference distribution given by Theorem 1. The AM regression bootstrap critical values are labeled as *AM: λ,bootstrap* for the *i.i.d.* bootstrap that preserves the locations of missing data and *AM: bootstrap* for the *i.i.d.* bootstrap that resamples from the AM series that have zeros in the missing locations. For the ES regression, *i.i.d.* bootstrap critical values are labeled *ES: bootstrap*. In all cases, 999 bootstrap resamples are used.

5.3 Finite Sample Results

Using 10,000 replications, we compute empirical null rejection probabilities and power. The nominal level is 0.05 in all cases. Here we mainly report results for the Bartlett kernel. The results for the QS kernel are very similar and available on our website. We consider the non-random missing process case first. Figure 4 shows empirical rejection probabilities for the World War missing processes. When the AM approach is used, the first thing to notice is that the *AM:λ,fixed-b* critical values give more accurate inference than the *AM: fixed-b* critical values regardless of the serial correlation and the time span. Figure 4 shows that *AM: λ,fixed-b* rejection rates are always closer to 0.05 than *AM: fixed-b* rejection rates. Similar patterns hold for the bootstrap critical values: *AM: λ,bootstrap* rejection rates are always closer to 0.05 than *AM: bootstrap* rejection rates. These results are as expected given that the missing process is non-random and both the *AM:λ,fixed-b* and *AM: λ,bootstrap* critical values capture the locations of the missing observations.

These simulation results also suggest that the bootstrap is able to correct some of the size distortions present when using the *AM: λ,fixed-b* asymptotic critical values. When $T = 36$, *AM: λ,bootstrap* rejection rates are closer to 0.05 than *AM: λ,fixed-b* rejection rates. Similarly, *AM: bootstrap* rejection rates are closer to 0.05 than *AM: fixed-b* rejection rates. This improvement shrinks when the time span increases to $T = 144$. There is an intuitive explanation for why the bootstrap is performing better than the asymptotic critical values when T is small. While the regressor and regression error are Gaussian, the product, $x_t u_t$, is not Gaussian and has a *product normal* distribution. It is the distribution of $x_t u_t$ that matters for the accuracy of the FCLT when T is small. The product normal distribution has a density that is very different from a Gaussian density and the bootstrap is picking this up to some extent. As T increases and the FCLT becomes more accurate, we see smaller differences between the use of bootstrap critical values and asymptotic critical values.

As suggested by Theorem 2, the ES regression approach works well when the missing process is non-random.[13] For example, consider the yearly data case (Figure 4, top two graphs). When $\rho = 0.3$, there are only very mild over-rejection problems. When $\rho = 0.9$, the over-rejection problem becomes severe but this over-rejection tendency when the data is highly correlated is something that is routinely found even when no observations are missing. Given that the effective sample size is only 24 for the World War missing process with $T = 36$, the ES regression model works surprisingly well. In fact the performance of the ES regression model with the *ES: bootstrap* critical values is similar to the AM approach with the *AM: λ,bootstrap* critical values, which has the least size distortion for the AM regression.

Because the ES regression approach with the *ES: bootstrap* critical values and the AM approach with

[13] For the ES regression model, the differences between *ES:fixed-b* rejection rates and *ES: bootstrap* rejection rates are small in general. Hence only the *ES: bootstrap* rejection rates are reported.

the *AM: λ,bootstrap* critical values work similarly well under the null, a researcher might consider choosing between the two approaches based on power performance. We compared the power of these two approaches (not size adjusted) for both the Bartlett and QS kernels by generating a latent process at the quarterly frequency with $\rho = 0.6$. We found the general tendencies that, as b increases, power tends to decrease and that the Bartlett kernel has better power than the QS kernel. However when comparing the AM approach with the ES approach, there was little difference in power performance. Therefore, the two approaches have similar performance in terms of both size and power.[14]

Next, we consider the random missing process case. Figure 5 shows empirical rejection probabilities for the Bernoulli missing processes. In this case, the HAR t-statistics for both the AM and ES approaches follow the usual fixed-b limit. Thus, *AM: fixed-b*, *AM: bootstrap*, and *ES: bootstrap* critical values are relevant. Similar to the previous findings in the non-random missing process case, the bootstrap critical values perform better than the asymptotic critical values. When $T = 50$ and $p = 0.3$ (Figure 5, top-left graph), *AM: bootstrap* rejection rates are closer to 0.05 than *AM: fixed-b* rejection rates. The differences between these two rejection rates are larger when ρ is higher. However, the ES regression model does not work as well as in the non-random missing process case. When $T = 50$ and 70% of the observations are missing, over rejection problems for the ES regression approach are severe. Thus, *AM: bootstrap* rejection rates show the least size distortion in these DGPs. As the time span (T) increases or the proportion of the missing observations shrinks, *AM: fixed-b*, *AM: bootstrap*, and *ES: bootstrap* rejection rates converge.

Interestingly, simulation results for the random missing process case suggest that using bootstrap critical values that treat the missing locations as fixed performs no worse than the bootstrap that correctly treats the missing locations as random. There exists negligible difference between *AM: bootstrap* rejection rates and *AM: λ,bootstrap* rejection rates when $T = 50$ and 70% of the data are missing. These differences go away if either the sample size increases or the proportion of missing observations decreases. Hence it appears that even if the missing process is random, there is little harm in using bootstrap critical values that treat the missing locations as given. This finding can be explained using the following thought experiment. Suppose we take the limiting distributions in 1(c) and ask what happens to these fixed-λ limiting random variable as the number of missing clusters grows ($C \to \infty$) and each missing cluster shrinks ($\lambda_{2n} - \lambda_{2n-1} \to 0$) such that the proportion of observed data in any closed inter-

[14]Interested readers can refer the Supplemental Appendix available on our website, http://msu.edu/∼ tjv/working.html, where we report the power performance for $b = \{0.17, 0.5, 0.75, 0.83\}$.

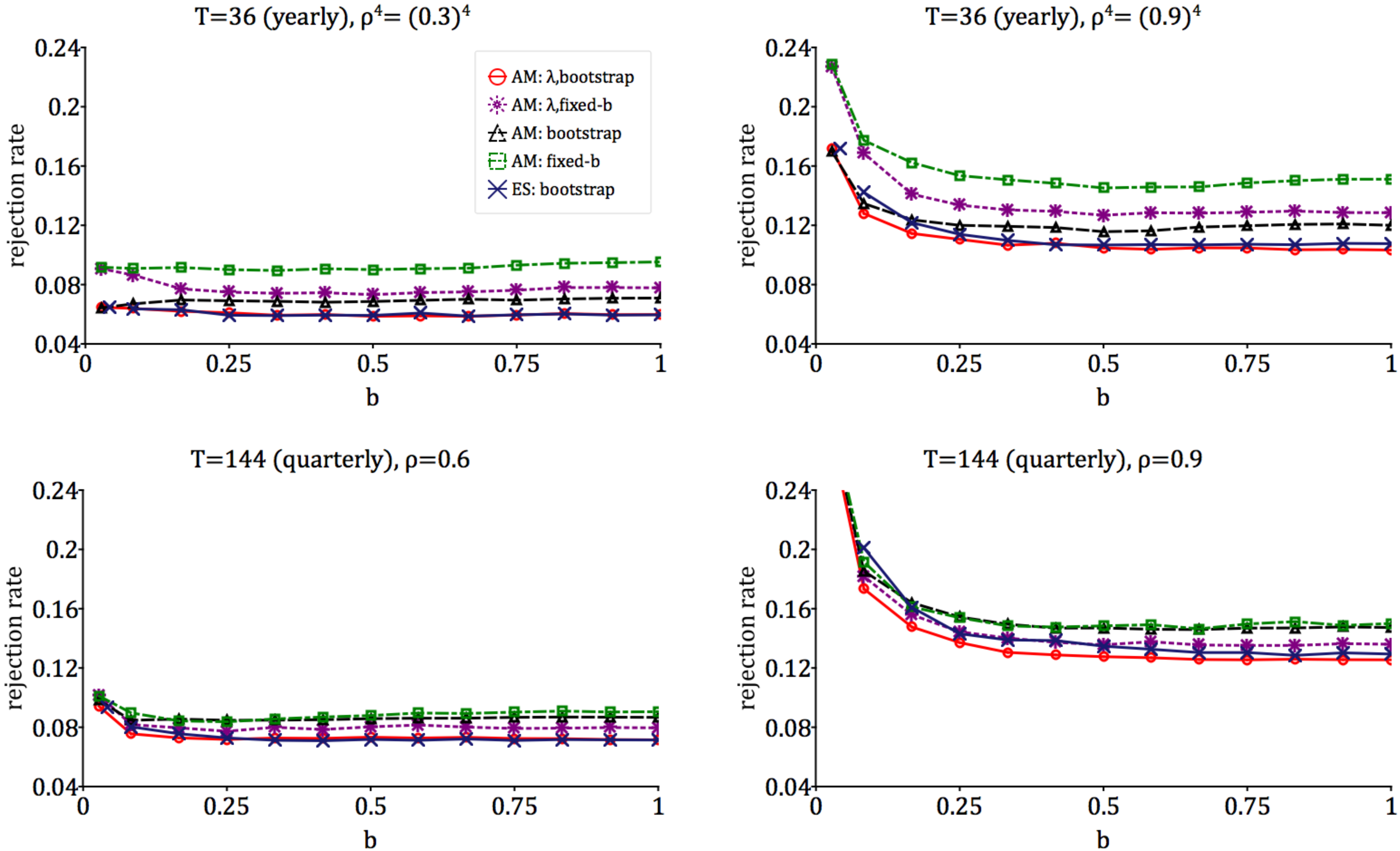
val of $[0,1]$ remains a constant (as is the case for the Bernoulli missing process). In Appendix A we show that the fixed-λ limiting random variable becomes the usual fixed-b distribution.

In summary, the simulation results suggest the AM approach with the *AM:* λ*,bootstrap* critical values as a generally reliable approach for inference in regression models with missing data because (1) the *AM:* λ*,bootstrap* critical values work particularly well whether the missing process is random or non-random and (2) while the ES approach can work well for the non-random missing process case it still does not dominate the AM approach either in terms of size distortions or power.

6 CONCLUSION

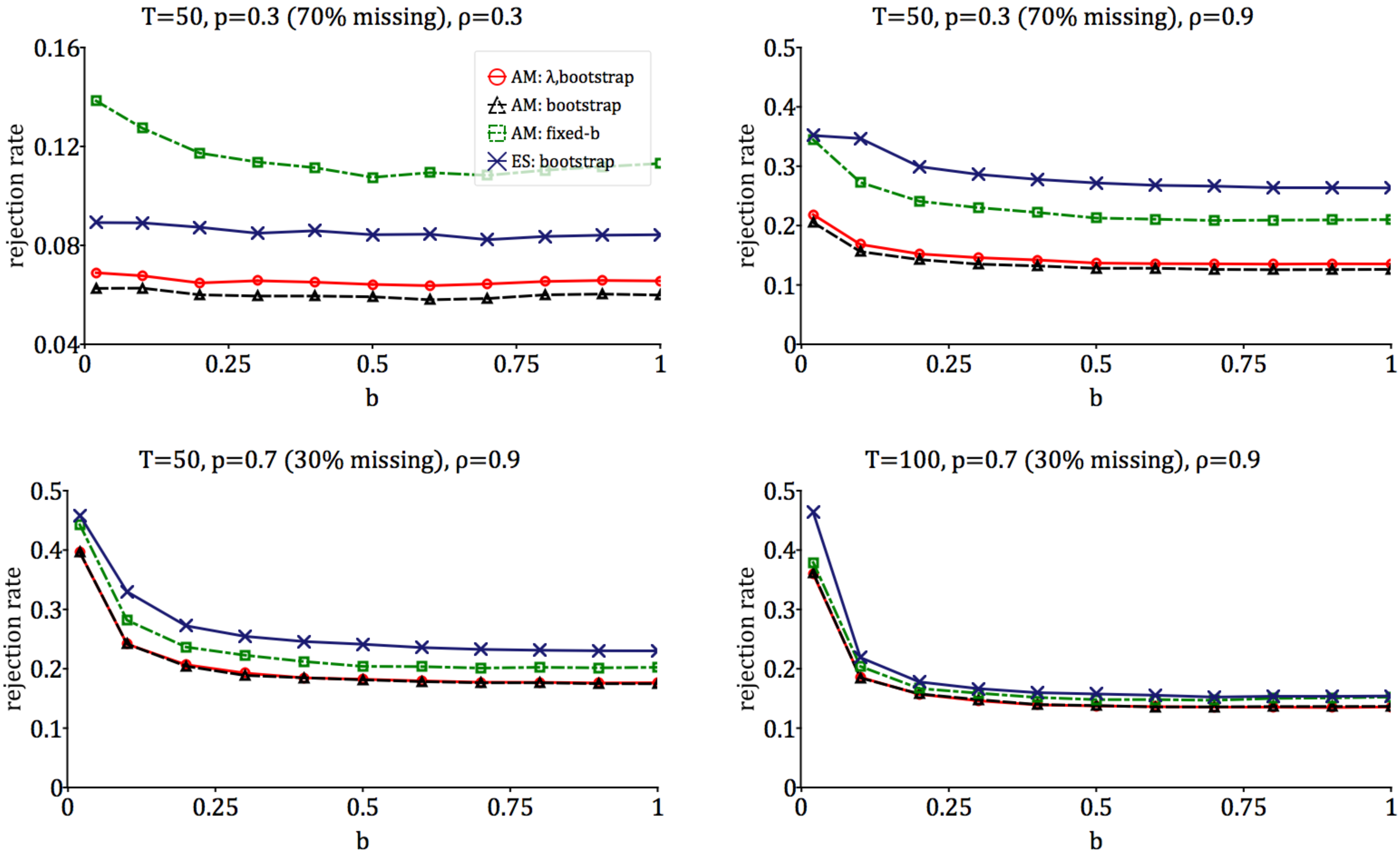
In this paper, we analyze the impact of missing observations on the HAR inference under the fixed-b asymptotic framework. Our theoretical analysis and finite sample simulations suggest the following advice for practitioners especially in situations where it is not clear whether it is more reasonable to view the missing data as fixed and exogenous or as generated by a random missing process. First, the missing observations for all variables should be replaced with zeros to preserve the timing of the observed data. Second, fixed-b critical values that treat the locations of the missing observations as fixed and non-random should be used for HAR robust tests. Third, the naive $i.i.d.$ bootstrap with missing locations matching those in the sample is a convenient way to validly generate these fixed-b critical values.

Figure 4: Empirical null rejection probabilities: World War missing process case



Notes: This figure provides empirical null rejection probabilities across different bandwidth sample size ratios (b) for the World War missing process case with nominal level 0.05. For the AM series, the rejection probabilities are computed based on the λ, fixed-b and the usual fixed-b distributions with and without bootstrapping. For the ES regression model, the rejection probabilities are computed from the usual fixed-b distribution with bootstrapping. Yearly data are generated taking every fourth observation of the quarterly data so ρ at the quarterly frequency implies that the correlation between yearly observations is ρ^4.

Figure 5: Empirical null rejection probabilities: Bernoulli missing process case



Notes: This figure provides empirical null rejection probabilities across different bandwidth sample size ratios (b) for the Bernoulli (p) missing process case with nominal level 0.05. For the AM series, the rejection probabilities are computed based on the λ, fixed-b distribution with bootstrapping and the usual fixed-b distribution with and without bootstrapping. For the ES regression model, the rejection probabilities are computed from the usual fixed-b distribution with bootstrapping.

Appendix A: Proofs for AM Statistics with Non-Random Missing Process

This section contains proofs for Theorem 1. Notation in this section are defined in Sections 2 and 3. This section is organized as follows. First two lemmas establish the limits of scaled partial sums of the AM series, $T^{-1}\sum_{t=1}^{[rT]} x_t x_t'$ and $T^{-1/2}\sum_{t=1}^{[rT]} v_t$. These two lemmas then lead to the asymptotic result for the OLS estimator (Theorem 1 (a)). Then we derive the limit of the partial sum, $T^{-1/2}\hat{S}_{[rT]} = T^{-1/2}\sum_{t=1}^{[rT]} \hat{v}_t$ in Lemma A3 which allows us to obtain the asymptotic limit of the HAR variance estimator of the AM series, $\hat{\Omega}$ (Theorem 1 (b)). Finally using Theorem 1 (a) and 1 (b), we establish the limit of the HAR *Wald* test (Theorem 1 (c)).

Lemma A1. *Under Assumptions fixed-λ and LP, for all $r \in [0,1]$,*

$$\frac{1}{T}\sum_{t=1}^{[rT]} x_t x_t' \Rightarrow \int_0^r \phi(s) ds Q^*.$$

Proof: Recall that data are observed at $t = [rT]$ whenever $r \in (\lambda_{2n}, \lambda_{2n+1}]$, $n = 0, \ldots, C$ (see Definition 1). Hence, we can write

$$a_t = \sum_{n=0}^{C} \mathbb{1}\left\{\lambda_{2n} T < t \leq \lambda_{2n+1} T\right\} = \phi\left(\frac{t}{T}\right). \tag{A.1}$$

Then,

$$\frac{1}{T}\sum_{t=1}^{[rT]} x_t x_t' = \frac{1}{T}\sum_{t=1}^{[rT]} a_t x_t^* x_t^{*\prime} = \frac{1}{T}\sum_{t=1}^{[rT]} \phi\left(\frac{t}{T}\right) x_t^* x_t^{*\prime} \Rightarrow \int_0^r \phi(s) ds Q^*$$

The weak convergence follows from Assumption LP.1. □

Lemma A2. *Under Assumptions fixed-λ and LP, for all $r \in [0,1]$,*

$$T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \Lambda^* \int_0^r \phi(s) d\mathcal{W}_k(s).$$

Proof:

$$T^{-1/2}\sum_{t=1}^{[rT]} v_t = T^{-1/2}\sum_{t=1}^{[rT]} a_t v_t^* = T^{-1/2}\sum_{t=1}^{[rT]} \phi\left(\frac{t}{T}\right) v_t^* \Rightarrow \Lambda^* \int_0^r \phi(s) d\mathcal{W}_k(s)$$

The second equality comes from (A.1) and the weak convergence immediately follows from Assumption LP.2. Note that $\int_0^r \phi(s) d\mathcal{W}_k(s)$ is not a Wiener process because for $r, s \in [0,1]$,

$$E\left(\int_0^r \phi(u) d\mathcal{W}_k(u) \int_0^s \phi(v) d\mathcal{W}_k(v)'\right) = \int_0^{r\wedge s} \phi(u)^2 du I_k = \int_0^{r\wedge s} \phi(u) du I_k \neq (r \wedge s) I_k, \tag{A.2}$$

which depends on $\{\lambda_i\}$ through $\phi(u)$. Here, $\phi(u)^2 = \phi(u)$ since $\phi(u)$ is an indicator function. □

Proof of Theorem 1 (a): Using Lemmas A1 and A2, it follows that

$$\sqrt{T}\left(\hat{\beta} - \beta\right) = \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} T^{-1/2}\sum_{t=1}^{T} v_t \Rightarrow Q^{*-1}\Lambda^* \frac{\int_0^1 \phi(s) d\mathcal{W}_k(s)}{\int_0^1 \phi(s) ds} \equiv Q^{*-1}\Lambda^* \frac{\overline{\mathcal{W}}_k}{\lambda}.$$

□

Lemma A3. *Let* $T^{-1/2}\hat{S}_{[rT]} = T^{-1/2}\sum_{t=1}^{[rT]}\hat{v}_t$. *Then, under Assumptions fixed-*λ *and LP, as* $T \to \infty$,

$$T^{-1/2}\hat{S}_{[rT]} \Rightarrow \Lambda^* \int_0^r d\breve{B}_k(s,\{\lambda_i\}) \equiv \Lambda^* \breve{B}_k(r,\{\lambda_i\}),\ r \in [0,1],$$

$$\text{where } d\breve{B}_k(s,\{\lambda_i\}) = \phi(s)\left(d\mathcal{W}_k(s) - ds\frac{\int_0^1 \phi(s)d\mathcal{W}_k(s)}{\int_0^1 \phi(s)ds}\right),\ s \in [0,1].$$

Proof: For $r \in [0,1]$, we can write

$$\begin{aligned}
T^{-1/2}\sum_{t=1}^{[rT]}\hat{v}_t &= T^{-1/2}\sum_{t=1}^{[rT]} v_t - \frac{1}{T}\sum_{t=1}^{[rT]} x_t x_t' \sqrt{T}\left(\hat{\beta}-\beta\right) \\
&\Rightarrow \Lambda^* \int_0^r \left(\phi(s)d\mathcal{W}_k(s) - \phi(s)ds\frac{\int_0^1 \phi(u)d\mathcal{W}_k(u)}{\int_0^1 \phi(u)du}\right)
\end{aligned}$$

where the weak convergence is straightforward from Lemmas A1-A3. □

Proof of Theorem 1(b): Define $\mathcal{K}_b(x) \equiv \mathcal{K}\left(\frac{x}{b}\right)$ and $K_b(x,y) = \mathcal{K}_b(x-y)$. With Assumption $\mathcal{K}$, $K_b(x,y)$ can be represented as

$$K_b(x,y) = \sum_{j=1}^{\infty} \nu_j f_j(x) f_j(y), \tag{A.3}$$

where $\{\nu_j\}$ is a sequence of eigenvalues and $\{f_j(t)\}$ is an orthonormal sequence of eigenfunctions corresponding to eigenvalues ν_j, because the series on the right side converges in $L^2([0,1]\times[0,1])$ to $K_b(x,y)$. See Sun (2014b) for details.

Using (A.3), we can write

$$\begin{aligned}
\hat{\Omega} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} K_b\left(\frac{t}{T},\frac{\tau}{T}\right)\hat{v}_t\hat{v}_\tau' \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\sum_{j=1}^{\infty}\nu_j f_j\left(\frac{t}{T}\right) f_j\left(\frac{\tau}{T}\right)\hat{v}_t\hat{v}_\tau' \\
&\Rightarrow \Lambda^* \int_0^1\int_0^1 \sum_{j=1}^{\infty}\nu_j f_j(r) f_j(s)\, d\breve{B}_k(r,\{\lambda_i\}) d\breve{B}_k(s,\{\lambda_i\})'\Lambda^{*\prime} \\
&= \Lambda^* \int_0^1\int_0^1 K_b(r,s) d\breve{B}_k(r,\{\lambda_i\}) d\breve{B}_k(s,\{\lambda_i\})'\Lambda^{*\prime} \\
&= \Lambda^* P_{\mathcal{K}}\left(b, \breve{B}_k(\{\lambda_i\})\right)\Lambda^{*\prime}.
\end{aligned}$$

The weak convergence follows from Lemma A3. □

Proof of Theorem 1(c): Applying the delta method and using Theorem 1(a), it is straightforward to show that

$$\sqrt{T}r(\hat{\beta}) \Rightarrow R(\beta_0)(\lambda Q^*)^{-1}\Lambda^*\overline{\mathcal{W}}_k,$$

where $R(\beta_0) = {}^{\partial r(\beta)}/_{\partial\beta'}\big|_{\beta=\beta_0}$. Note that the limit is q linear combinations of k independent Wiener processes. Because Wiener processes are Gaussian, linear combinations of Wiener processes are also Gaussian. Thus, we can rewrite the q linear combinations of k independent Wiener processes as q linear combinations of q independent Wiener processes. Define the $q \times q$ matrix Δ^* such that

$$\Delta^*\Delta^{*\prime} = R(\beta_0)(\lambda Q^*)^{-1}\Omega^*(\lambda Q^*)^{-1}R(\beta_0)'.$$

An equivalent representation for $R\left(\beta_0\right)\left(\lambda Q^*\right)^{-1}\Lambda^*\overline{\mathcal{W}}_k$ is given by

$$R\left(\beta_0\right)\left(\lambda Q^*\right)^{-1}\Lambda^*\overline{\mathcal{W}}_k \equiv \Delta^*\overline{\mathcal{W}}_q. \tag{A.4}$$

Similarly,

$$R\left(\beta_0\right)\left(\lambda Q^*\right)^{-1}\Lambda^*\mathcal{W}_k(r) \equiv \Delta^*\mathcal{W}_q(r). \tag{A.5}$$

We can write

$$\begin{aligned}
W_T &= \sqrt{T}r(\hat{\beta})'\left[R(\hat{\beta})\left(T^{-1}\sum_{t=1}^{T}x_t x_t'\right)^{-1}\hat{\Omega}\left(T^{-1}\sum_{t=1}^{T}x_t x_t'\right)^{-1}R(\hat{\beta})'\right]^{-1}\sqrt{T}r(\hat{\beta}) \\
&\Rightarrow \left[R(\beta_0)\left(\lambda Q^*\right)^{-1}\Lambda^*\overline{\mathcal{W}}_k\right]' \\
&\times\left[R\left(\beta_0\right)\left(\lambda Q^*\right)^{-1}\Lambda^* P_{\mathcal{K}}\left(b,\breve{B}_k\left(\{\lambda_i\}\right)\right)\Lambda^{*\prime}\left(\lambda Q^*\right)^{-1}R\left(\beta_0\right)'\right]^{-1}\left[R(\beta_0)\left(\lambda Q^*\right)^{-1}\Lambda^*\overline{\mathcal{W}}_k\right] \\
&= \left(\Delta^*\overline{\mathcal{W}}_q\right)'\left[\Delta^* P_{\mathcal{K}}\left(b,\breve{B}_q\left(\{\lambda_i\}\right)\right)\Delta^{*\prime}\right]^{-1}\Delta^*\overline{\mathcal{W}}_q \\
&= \overline{\mathcal{W}}_q' P_{\mathcal{K}}\left(b,\breve{B}_q\left(\{\lambda_i\}\right)\right)^{-1}\overline{\mathcal{W}}_q.
\end{aligned}$$

The weak convergence follows from Lemma A1 and Theorem 1(a)- 1(b). The second equality follows from (A.4) and (A.5). The third equality is straightforward because Δ^* is invertible and cancels completing the proof. For the single restriction case ($q = 1$), it follows that

$$t_T \Rightarrow \frac{\overline{\mathcal{W}}_1}{\sqrt{P_{\mathcal{K}}\left(b,\breve{B}_1\left(\{\lambda_i\}\right)\right)}}.$$

Here, $\overline{\mathcal{W}}_k$ is independent of $P_{\mathcal{K}}\left(b,\breve{B}_k\left(\{\lambda_i\}\right)\right)$ because

$$\begin{aligned}
E\left[\overline{\mathcal{W}}_k\breve{B}_k\left(r,\{\lambda_i\}\right)'\right] &= E\left[\int_0^1\phi(u)d\mathcal{W}_k(u)\int_0^r\left(\phi(v)d\mathcal{W}_k(v)' - \phi(v)dv\frac{\int_0^1\phi(s)d\mathcal{W}_k(s)'}{\int_0^1\phi(s)ds}\right)\right] \\
&= \int_0^r\phi(v)dvI_k - \int_0^r\phi(v)dv\frac{\int_0^1\phi(s)^2dsI_k}{\int_0^1\phi(s)ds} = 0_{k\times k}
\end{aligned}$$

and both $\breve{B}_k\left(r,\{\lambda_i\}\right)$ and $\overline{\mathcal{W}}_k$ are Gaussian. □

Suppose we now take the limiting distributions in Theorem 1(c) and ask what happens to these fixed-λ limiting random variables as the number of missing clusters grows ($C\to\infty$) and each missing cluster shrinks ($\lambda_{2n} - \lambda_{2n-1}\to 0$) such that the proportion of observed data in any closed interval of $[0,1]$ remains a constant $\eta = \Sigma a_t/T$ (as is the case for the Bernoulli missing process). It follows that $\int_0^{r\wedge s}\phi(u)du \to (r\wedge s)\,\eta$ as $C\to\infty$ and we have

$$E\left(\int_0^r\phi(u)d\mathcal{W}_k(u)\int_0^s\phi(v)d\mathcal{W}_k(v)'\right) = \int_0^{r\wedge s}\phi(u)duI_k \to (r\wedge s)\,\eta I_k,$$

which in turn implies that $\int_0^r\phi(u)d\mathcal{W}_k(u)/\sqrt{\eta}$ is now a $k\times 1$ standard Wiener process. Similarly,

$$\int_0^r\phi(u)duQ^* \to r\eta Q^* \quad \because \text{Lemma } A1$$

as $C\to\infty$. The dependence on η scales out of the test statistics and the standard fixed-b random variables are obtained.

Appendix B. Proofs for ES Statistic with Non-Random Missing Process

Appendix B contains proofs for Theorem 2. First two lemmas show that $T_{ES}^{-1/2}\sum_{t=1}^{[rT_{ES}]} x_t^{ES}u_t^{ES} \Rightarrow \Lambda^*\dddot{\mathcal{W}}_k(r)$ where $\dddot{\mathcal{W}}_k(r)$ is a $k\times 1$ standard Wiener process and $T_{ES}^{-1}\sum_{\tau=1}^{[rT_{ES}]} x_\tau^{ES}x_\tau^{ES\prime} \Rightarrow rQ^*$ when the missing process is non-random, i.e.,$\{\lambda_i\}$ are treated as fixed. Then by the usual fixed-b arguments as in Kiefer and Vogelsang (2005), Theorem 2 follows. Notation in this section are defined in Sections 2 and 3. Throughout this section we assume that $M_{ES} = bT_{ES}$ where $b \in (0,1]$ is fixed. We define the summation be zero whenever the starting value is larger than the final value. For example, for a sequence $\{a_k\}$, we have $\sum_{k=1}^{0} a_k = 0$.

Lemma 1. *Under Assumptions fixed-λ and LP, for $r \in [0,1]$,*

$$T_{ES}^{-1/2}\sum_{\tau=1}^{[rT_{ES}]} x_\tau^{ES}u_\tau^{ES} \Rightarrow \Lambda^*\dddot{\mathcal{W}}_k(r),$$

where $\dddot{\mathcal{W}}_k(r) = \lambda^{-1/2}\int_0^{\left(r\lambda+\sum_{k=1}^{2n}(-1)^k\lambda_k\right)} \phi(u)d\mathcal{W}_k(u)$ is a standard $k\times 1$ Wiener process.

Proof: Let $\tau = [rT_{ES}]$ be the time index for the ES regression model. Then there is a corresponding time index $t = [sT]$ for the AM series such that $\tau = [rT_{ES}] \equiv \sum_{i=1}^{[sT]} a_i$. Because we are matching the ES regression model to the AM series, $t = [sT]$ should be a time period where the data is observed. By definition this implies that $s \in (\lambda_{2n}, \lambda_{2n+1}]$, for some $n = 0,\ldots,C$ and it follows that $\sum_{i=1}^{[sT]} a_i = [sT] - \sum_{k=1}^{2n}(-1)^k\lambda_k T$. Combining this expression with $[rT_{ES}] = \sum_{i=1}^{[sT]} a_i$, we have $[sT] = [rT_{ES}] + \sum_{k=1}^{2n}(-1)^k\lambda_k T = [r\lambda T] + \sum_{k=1}^{2n}(-1)^k\lambda_k T$. Hence for every r, there exists $s \in (\lambda_{2n}, \lambda_{2n+1}]$ such that

$$[sT] = \left[\left(r\lambda + \sum_{k=1}^{2n}(-1)^k\lambda_k\right)T\right] \tag{A.1}$$

for some $n = 0,\ldots,C$. Hence for $r\in[0,1]$, $\exists s \in (\lambda_{2n}, \lambda_{2n+1}]$, such that

$$\begin{aligned}\frac{1}{\sqrt{T_{ES}}}\sum_{\tau=1}^{[rT_{ES}]} v_\tau^{ES} &= \frac{1}{\sqrt{\lambda T}}\sum_{t=1}^{[sT]} v_t \\ &\Rightarrow \lambda^{-1/2}\Lambda^*\int_0^{\left(r\lambda+\sum\limits_{k=1}^{2n}(-1)^k\lambda_k\right)} \phi(u)d\mathcal{W}_k(u) \quad \because \text{Lemma } A2 \text{ and (A.1)} \\ &\equiv \Lambda^*\dddot{\mathcal{W}}_k(r).\end{aligned}$$

Next, we show that $\dddot{\mathcal{W}}_k(r)$ is a standard Wiener process. Suppose $l \le r$ and $m \le n$. Then,

$$\begin{aligned}E\left(\dddot{\mathcal{W}}_k(r)\dddot{\mathcal{W}}_k(l)'\right) &= E\left(\lambda^{-1}\int_0^{\left(r\lambda+\sum_{k=1}^{2n}(-1)^k\lambda_k\right)}\phi(u)d\mathcal{W}_k(u)\int_0^{\left(l\lambda+\sum_{k=1}^{2m}(-1)^k\lambda_k\right)}\phi(v)d\mathcal{W}_k(v)'\right) \\ &= \lambda^{-1}\int_0^{\left(l\lambda+\sum_{k=1}^{2m}(-1)^k\lambda_k\right)}\phi(v)dvI_k \\ &= \lambda^{-1}\left(\left(l\lambda+\sum_{k=1}^{2m}(-1)^k\lambda_k\right) - \lambda_{2m} + \lambda_{2m-1} - \lambda_{2m-2} + \cdots + \lambda_1\right)I_k = lI_k\end{aligned}$$

The second equality follows from $l\lambda + \sum_{k=1}^{2m}(-1)^k\lambda_k \le r\lambda + \sum_{k=1}^{2n}(-1)^k\lambda_k$. By definition, $l\lambda + \sum_{k=1}^{2m}(-1)^k\lambda_k \in (\lambda_{2m}, \lambda_{2m+1}]$ and the third equality follows immediately. By the same reasoning, when $r < l$, $E\left(\dddot{\mathcal{W}}_k(r)\dddot{\mathcal{W}}_k(l)'\right) = rI_k$. Hence, for $r, l \in [0,1]$,

$$E\left(\dddot{\mathcal{W}}_k(r)\dddot{\mathcal{W}}_k(l)'\right) = (r\wedge l)\, I_k.$$

Since $\ddot{\mathcal{W}}_k(u)$ is a zero-mean Gaussian process with $cov\left(\ddot{\mathcal{W}}_k(r), \ddot{\mathcal{W}}_k(u)\right) = (r \wedge u)\, I_k$, it is a standard Wiener process. □

Alternatively, $\ddot{\mathcal{W}}_k(u)$ being a standard Wiener process can be shown through the time-changed Brownian motion. Let $\alpha(t) = \int_0^t \lambda^{-1}\phi(u)^2 du$ and $\mathcal{B}(t) = \int_0^t \lambda^{-1/2}\phi(u) d\mathcal{W}_k(u)$. Then $\mathcal{B}(t)$ is a Gaussian process with independent increments where $var\left(\mathcal{B}(t) - \mathcal{B}(s)\right) = \left(\alpha(t) - \alpha(s)\right) I_k$, $t > s$. Hence $\mathcal{B}(t)$ is a time-changed Brownian motion with a time index $\alpha(t)$: $\mathcal{B}(t) \overset{d}{=} \mathcal{W}_k\left(\alpha(t)\right)$.

Note that

$$\ddot{\mathcal{W}}_k(r) = \mathcal{B}\left(r\lambda + \sum_{k=1}^{2n}(-1)^k\lambda_k\right) \overset{d}{=} \mathcal{W}_k\left(\alpha\left(r\lambda + \sum_{k=1}^{2n}(-1)^k\lambda_k\right)\right).$$

In addition, since $\phi(s)$ is an indicator function, $\alpha(t) = \int_0^t \lambda^{-1}\phi(u)^2 du = \int_0^t \lambda^{-1}\phi(u) du$. It follows that

$$\alpha\left(\left(r\lambda + \sum_{k=1}^{2n}(-1)^k\lambda_k\right)\right) = \int_0^{\left(r\lambda + \sum_{k=1}^{2n}(-1)^k\lambda_k\right)} \lambda^{-1}\phi(u) du = \lambda^{-1} r\lambda = r.$$

Therefore, $\ddot{\mathcal{W}}_k(r) = \mathcal{W}_k(r)$ and $\ddot{\mathcal{W}}_k(r)$ is a standard Brownian motion.

Lemma 2. *Under Assumptions fixed-λ and LP, for $r \in [0,1]$,*

$$\frac{1}{T_{ES}} \sum_{\tau=1}^{[rT_{ES}]} x_\tau^{ES} x_\tau^{ES\prime} \Rightarrow rQ^*.$$

Proof: Recall from (A.1) that for each time index of the ES regression model $\tau = [rT_{ES}]$, there is a corresponding time index $t = [sT]$ of the AM series such that $s \in (\lambda_{2n}, \lambda_{2n+1}]$ and $s = r\lambda + \sum_{k=1}^{2n}(-1)^k\lambda_k$. Hence we can write

$$\begin{aligned}
\frac{1}{T_{ES}} \sum_{\tau=1}^{[rT_{ES}]} x_\tau^{ES} x_\tau^{ES\prime} &= \frac{1}{\lambda T} \sum_{t=1}^{[sT]} x_t x_t' \\
&\Rightarrow \lambda^{-1} \int_0^s \phi(u) du Q^* \quad \because \text{Lemma } A1 \\
&= \lambda^{-1}\left(s - \lambda_{2n} + \lambda_{2n-1} - \ldots + \lambda_1\right) Q^* \quad \because s \in (\lambda_{2n}, \lambda_{2n+1}] \\
&= \lambda^{-1}\left(\left(r\lambda + \sum_{k=1}^{2n}(-1)^k\lambda_k\right) - \lambda_{2n} + \lambda_{2n-1} - \ldots + \lambda_1\right) Q^* \\
&= rQ^*.
\end{aligned}$$

Proof of Theorem 2(b) and (c): We proved that $T_{ES}^{-1/2}\sum_{\tau=1}^{[rT_{ES}]} v_\tau^{ES} \Rightarrow \Lambda^* \ddot{\mathcal{W}}_k(r)$ where $\ddot{\mathcal{W}}_k(r)$ is a standard Wiener process in Lemma 1 and $T_{ES}^{-1}\sum_{\tau=1}^{[rT_{ES}]} x_\tau^{ES} x_\tau^{ES\prime} \Rightarrow rQ^*$ in Lemma 2. Hence

$$\begin{aligned}
T_{ES}^{-1/2} \sum_{\tau=1}^{[rT_{ES}]} \hat{v}_\tau^{ES} &= T_{ES}^{-1/2} \sum_{\tau=1}^{[rT_{ES}]} v_\tau^{ES} - \frac{1}{T_{ES}} \sum_{\tau=1}^{[rT_{ES}]} x_\tau^{ES} x_\tau^{ES\prime} \sqrt{T_{ES}}\left(\hat{\beta} - \beta\right) \\
&\Rightarrow \Lambda^* \ddot{\mathcal{W}}_k(r) - rQ^* \Lambda^* Q^{*-1} \ddot{\mathcal{W}}_k(1) \\
&= \Lambda^*\left(\ddot{\mathcal{W}}_k(r) - r\ddot{\mathcal{W}}_k(1)\right) \equiv \Lambda^* \widetilde{\ddot{\mathcal{W}}}_k(r),
\end{aligned}$$

where $\widetilde{\ddot{\mathcal{W}}}_k(r)$ is a Brownian bridge. Hence Theorem 2(b) and 2(c) follow from the usual fixed-b arguments (see Kiefer and Vogelsang (2005)). □

References

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

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

Bloomfield, P. (1970). Spectral analysis with randomly missing observations. *Journal of the Royal Statistical Society*, 32:369–380.

Datta, D. D. and Du, W. (2012). Nonparametric HAC estimation for time series data with missing observations. *Working Paper*.

Dunsmuir, W. and Robinson, P. (1981). Asymptotic theory for time series containing missing and amplitude modulated observations. *The Indian Journal of Statistics*, 43:260–281.

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

Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory*, 18:1449–1459.

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

Neave, H. R. (1970). Spectral analysis of a stationary time series using initially scarce data. *Biometrika*, 57:111–122.

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

Parzen, E. (1963). On spectral analysis with missing observations and amplitude modulation. *The Indian Journal of Statistics*, 25.

Perron, P. (1991). A continuous time approximation to the unstable first-order autoregressive process: The case without an intercept. *Econometrica*, 59(1):211–236.

Phillips, P. C. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3):535–47.

Scheinok, P. A. (1965). Spectral analysis with randomly missed observations: The binomial case. *The Annals of Mathematical Statistics*, 36(3):971–977.

Sun, Y. (2014a). Fixed-smoothing asymptotics in a two-step generalized method of moments framework. *Econometrica*, 82(6):2327–2370.

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

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

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