Alternative Time-Average Covariance Matrix Estimation Procedures

Rebecca Kurtz-Garcia James Flegal

Department of Statistics, University of California Riverside

An Important Problem

Stationary random vector processes appear in time series, econometrics, spectral analysis, and Markov Chain Monte Carlo (MCMC) simulations. These fields all work with serially correlated multivariate data. With correlated data it is often difficult to create unbiased estimators for the time-average covariance matrix (TACM), Σ, even when the correlation structure is known. The spectral variance (SV) estimator is one of the most widely used,

$$\widehat{\Sigma}_b^{(SV)} = \sum_{h=-(T-1)}^{T-1} \kappa_b(h)\, \widehat{\Gamma}(h)$$

where T is the sample size, $\kappa_b(x)$ is the kernel function with bandwidth parameter b, and $\widehat{\Gamma}(h)$ is the estimated autocovariance function. The bandwidth parameter is the proportion of autocovariance matrices incorporated into the estimator. These SV estimators have two tuning parameters, the kernel function and bandwidth parameter. Research on these estimators is typically focused on optimizing b.

Standard inference procedures in these settings have three major limitations. First, these procedures utilize typical χ^2 critical values which rely on asymptotic assumptions that do not account for the variability incurred by estimating Σ. Secondly, these procedures rely on SV estimators which are subpar because they tends to underestimate the true TACM [2]. This bias is particularly problematic in data with moderate to high correlation. Lastly, SV estimators are typically optimized according to asymptotic mean squared error (AMSE) which indirectly measures the optimal goal of estimating Σ, which is to conduct inference procedures.

We use multiple tools to address these problems. In order to capture the variability of the estimators for Σ we consider the fixed-b asymptotic framework which has more realistic assumptions that match a finite setting. We also focus on the the lugsail estimator, instead of the SV estimator, which introduces two new tuning parameters (c, r) that induce a positive bias to offset that naturally occurring negative bias that occurs in the estimation process. Lastly, we select our bandwidth parameter b according to a testing optimal loss function which more directly measures the estimators ultimately utility. These tools used together can create an elegant solution to three pressing problems with testing procedures in settings with correlated stationary sequences.

Fixed-b Asymptotics

- Typical procedures rely on small-b asymptotics which assumes that $b \to 0$ as $T \to \infty$.
- We instead rely on the fixed-b asymptotics, which keeps the parameter b as a fixed value as $T \to \infty$.
- Under fixed-b asymptotics we obtain non-standard critical values for hypothesis testing, instead of the usual χ^2 values.
- These non-standard critical values account for the variability of the kernel function, the number of dimensions of testing restrictions (m), and the bandwidth parameter b, all of which are known to influence the variability of the test statistic.
- Figure 1a contains the $\alpha = 0.05$ critical values for a F type test statistic constructed using SV and lugsail estimators with a Bartlett kernel function.
- Figure 1b contains the distributions corresponding the test statistics in Figure 1a when $b = 0.07$.

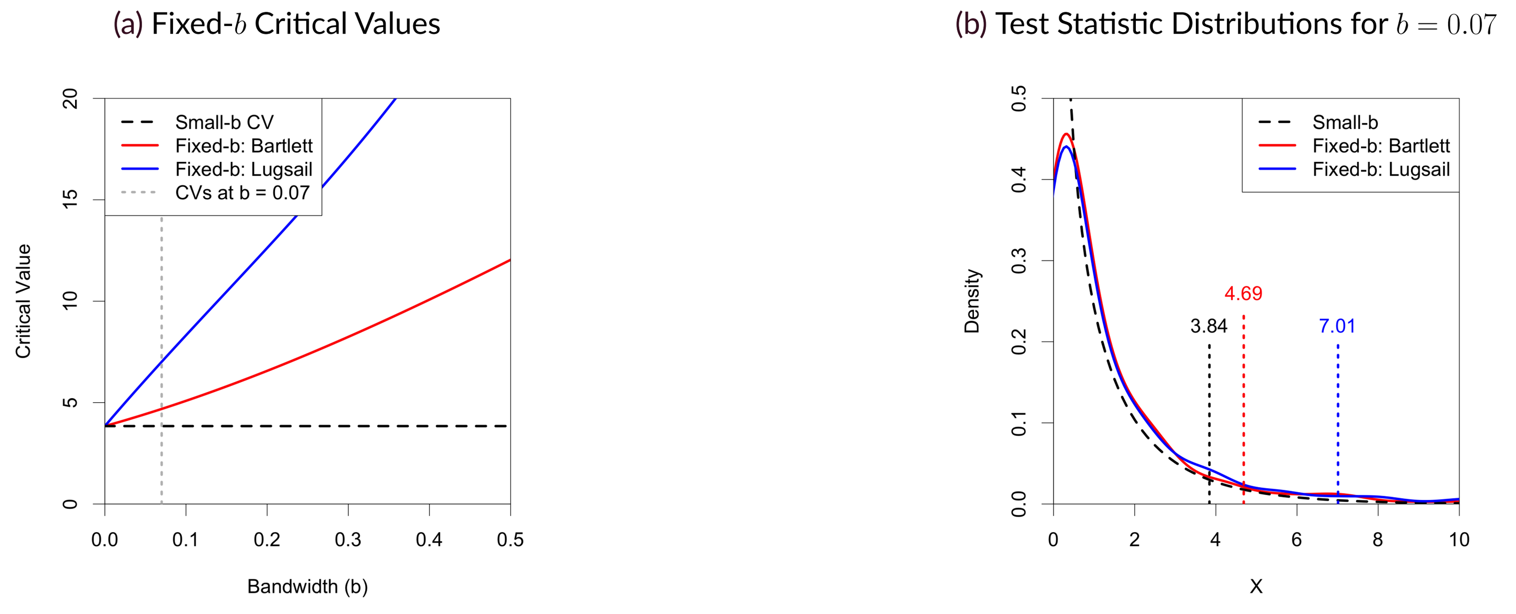

Figure 1. Critical Values as a Function of Bandwidth

Lugsail Estimators

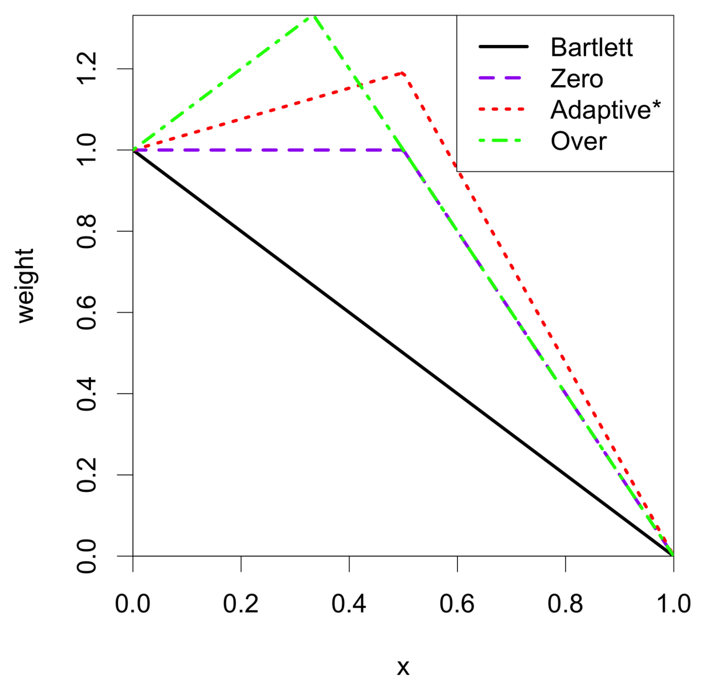

Figure 2. Lugsail Transformation of the Bartlett Kernel

The lugsail estimator is a linear combination of SV estimators,

$$\widehat{\Sigma}_b^{(L)} = \frac{1}{1-c}\widehat{\Sigma}_b^{(SV)} - \frac{c}{1-c}\widehat{\Sigma}_{\lfloor br \rfloor}^{(SV)}$$

where we have two new tuning parameters (c, r). It is recommended to pick (c, r) based on the amount of bias suspected. For example, if a sequence of random variables that has an underlying process similar to an AR(1) model with correlation coefficient $\rho \in \{[0, 0.7), [0.7, 0.95), [0.95, 1)\}$ then we classify the situation as moderate, high, or extreme, respectively. We then have guidelines for picking these new parameters in Table 1, where q is the Parzen characteristic component[6]. In Figure 2 we have the standard Bartlett kernel function with the three lugsail transformations for $T = 200$ and $b = 0.07$. This estimator essentially transforms the original kernel function to have larger weights, which induces a positive bias designed to offset the natural negative bias caused by the SV estimators.

Correlation	Lugsail Window	r	c
Moderate	Zero	2	r^{-q}
Moderate-High	Adaptive	2	$\frac{\log(T)-\log(bT)+1}{r^q(\log(T)-\log(bT))+1}$
High-Extreme	Over	3	$\frac{2}{(1+r^q)}$

Table 1. Recommended Lugsail Parameters

Lugsail estimators are intuitive and easy transformations of the familiar estimators that account for the finite sample bias. This positive bias is induced by allowing autocovariance weights to exceed 1, which is not commonly observed for typical kernel functions. Lugsail estimators inherit consistency from their SV counterparts, but the positive bias induces more variability, as we see in Figure 1.

Testing Optimal Loss Function

- We consider the following testing optimal loss function instead of using the traditional loss function, AMSE [4, 1],
$$Loss = (k)(|\Delta_s|)^2 + (1-k)(|\Delta_p^{max}|)^2$$
where
  - Δ_s size distortion, the deviation from the objective Type I error
  - Δ_p^{max} maximum size-adjusted power loss
  - $k \in [0, 1]$ is used to balance the weight of the two metrics.
- Recommended to use $k = 0.9$ in order to reflect the emphasis on size over power that is present in traditional inference procedures.
- Δ_p^{max} is a measure of the power loss compared to the oracle test of the same size, a metric for Type II error.
- This loss function results in an optimal bandwidth parameter of $b \sim O(T^{-q/(q+1)})$ instead of the usual $b \sim O(T^{-2q/(2q+1)})$ motivated by ASME [1].
- The loss function is a proxy measure for bias and variance, instead of bias *squared* and variance, which is measured by AMSE.
- The optimal bandwidth parameter is a function of the bias and variability of the estimator.
- Other testing optimal loss functions have been purposed which result in same rate of $b \sim O(T^{-2q/(2q+1)})$ [5, 3].

Performance

- Simulation study for an moderately correlated AR(1) data sequence ($\rho = 0.7$) where $T = 200$ using an F type statistic and $\alpha = 0.05$.
- We tried a varying number of hypothesis test restrictions ($m = 1, \dots, 12$) and bandwidth b values between (0, 1].
- The results below are for the case when $m = 1$ for the SV (red) and lugsail (blue) estimators with the Bartlett kernel.
- The solid lines correspond to the fixed-b CV, and the dashed lines the small-b CV.
- The power curves drawn in Figure 3b correspond to the optimal bandwidth parameter for the respective model, which is marked by a symbol in Figure 3a.

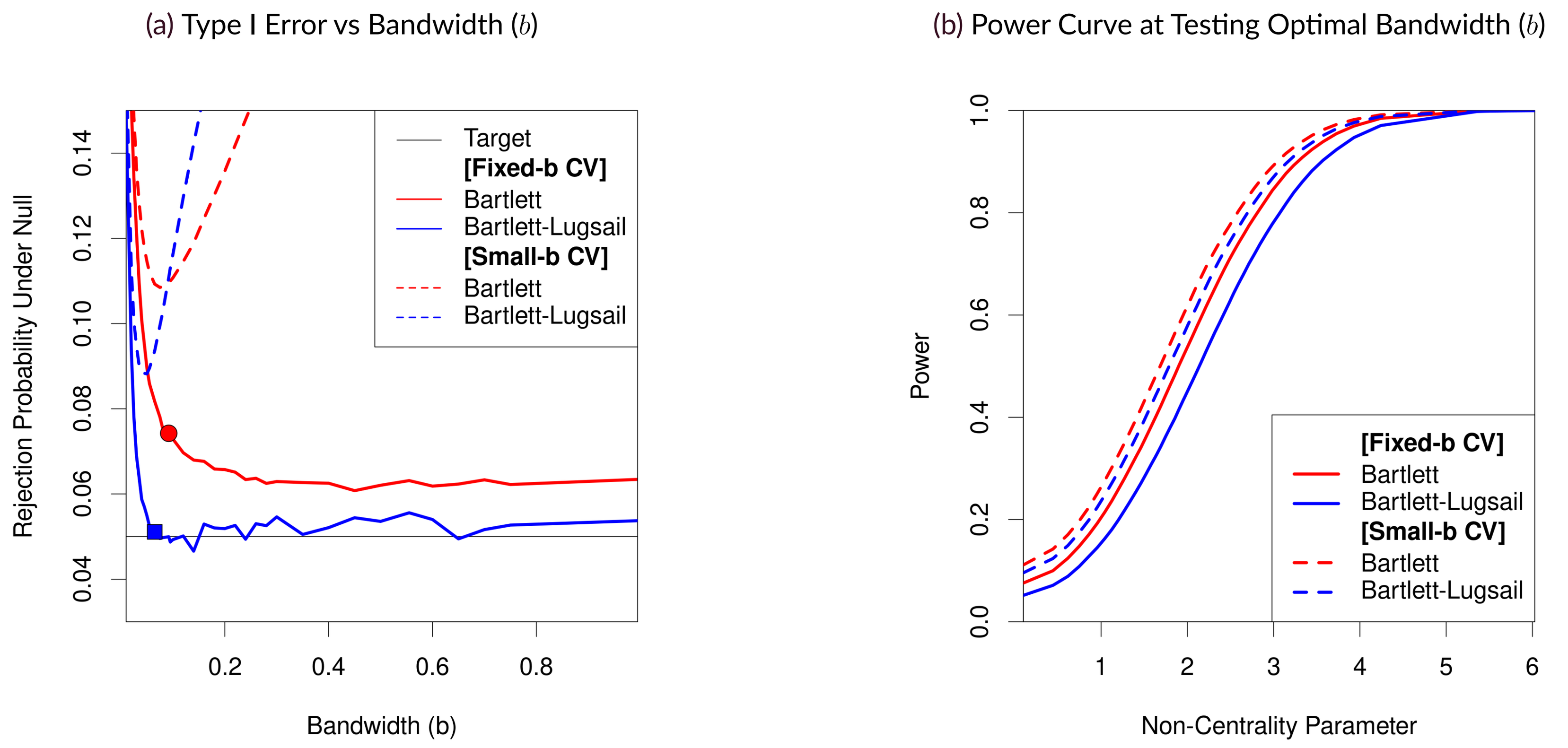

Figure 3. Selected Simulation Results

Remarks

- Smaller b values cause a higher bias for the estimate of Σ, which causes a higher size distortion. Larger b values result in smaller size distortions.
- The optimal b for the lugsail estimator is smaller than the optimal b for the SV estimator, but the size distortion is consistently smaller.
- The power between the two tests for their respective optimal bandwidth is negligible.
- The fixed-b CVs noticeably out preform the standard critical values when looking at Type I error, and the difference between the power curves between the two types of CVs is negligible.
- Not shown here is that smaller b values have lower variability, resulting in higher power. As the b increases power of the test tends to suffer.
- The choice of b reflects a trade off between bias and variability, and Type I and Type II errors.
- The combination of fixed-b asymptotic theory, lugsail estimators, and testing optimal loss functions result in an estimator with noticeably better size distortion, while keeping power in an acceptable range.

References

[1] Donald WK Andrews. "Heteroskedasticity and autocorrelation consistent covariance matrix estimation". In: *Econometrica: Journal of the Econometric Society* (1991), pp. 817–858.

[2] Wouter J Den Haan and Andrew T Levin. "12 A practitioner's guide to robust covariance matrix estimation". In: *Handbook of statistics* 15 (1997), pp. 299–342.

[3] Eben Lazarus, Daniel J Lewis, and James H Stock. "The Size-Power Tradeoff in HAR Inference". In: *Econometrica* 89.5 (2021), pp. 2497–2516.

[4] Eben Lazarus et al. "HAR inference: Recommendations for practice". In: *Journal of Business & Economic Statistics* 36.4 (2018), pp. 541–559.

[5] Yixiao Sun and Peter CB Phillips. "Optimal bandwidth choice for interval estimation in GMM regression". In: *Cowles Foundation Discussion Paper* (2008).

[6] Dootika Vats and James M Flegal. "Lugsail lag windows for estimating time-average covariance matrices". In: *Biometrika* (2021).