UNIVERSITY OF CALIFORNIA
RIVERSIDE

# Testing Optimal Bandwidth for Zero Lugsail Estimators

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Rebecca Kurtz-Garcia

September 2023

Dissertation Committee:

Dr. James Flegal, Chairperson
Dr. Xinping Cui
Dr. Gloria Gonzalez-Rivera

The Dissertation of Rebecca Kurtz-Garcia is approved:

---

---

---

Committee Chairperson

University of California, Riverside

## Acknowledgments

It is not possible to thank everyone who helped me get to where I am today. I am fortunate to have had wonderful teachers, friends, and family to support me. I am forever grateful and indebted to all of you.

To my teachers and mentors, I thank you for your encouragement and guidance. In particular, Nathan Nichols and Rebecca Pierce for their support and mentorship. My dissertation and oral qualifying exam committee members Gloria Gonzalez-Rivera, Xinping Cui, Esra Kurum, Weixin Yao, and Rami Zwick for their flexibility, questions, support, and suggestions. My dissertation advisor James Flegal who has supported me as a student in the classroom, doctoral candidate, new instructor, and future faculty member. Thank you for your patience, guidance, and advice.

I would like to thank my friends and the UCR statistics department. In particular, Lauren Daly, Luke Klein, Christian Michael, Abigail Modell, and Lauren Parker-Cappiello. Thank you for your friendship, sharing your stories, encouragement, and the coffee. You are all immensely appreciated and treasured.

To my family, immediate and beyond, thank you for your continuous love and support. Specifically my siblings, parents and in-laws who checked in with me and reminded of me of my life outside of work. In particular, my dad who bought me my first calculator. Finally, my partner Jacob Garcia, for all our discussions mathematical and otherwise, and the continuous stream of positivity.

# ABSTRACT OF THE DISSERTATION

Testing Optimal Bandwidth for Zero Lugsail Estimators

by

Rebecca Kurtz-Garcia

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, September 2023
Dr. James Flegal, Chairperson

Test statistics, confidence intervals, and p-values all typically rely on an estimate for variance. For data sets that are not independent and identically distributed (iid) caution must be used when selecting a variance estimator. If the dependence structure is unknown but stationary, a robust long run variance (LRV) estimator can be used which can handle a wide variety of scenarios. Estimation of the LRV is of interest in various fields such as time series, econometrics, spectral analysis, and Markov chain Monte Carlo simulations.

Spectral variance (SV) estimators are one of the most common LRV estimation methods, but they suffer from a negative bias in the presence of positive correlation. An alternative zero lugsail estimator has been proposed to combat this issue which has a zero asymptotic bias regardless of correlation. The optimal bias properties come at the expense of increased variability which causes testing error rates to suffer. Further advancements have been made regarding nonstandard limiting distributions that better incorporate the variability of LRV estimators. The zero lugsail estimator and the nonstandard limiting distribution address different issues, the former being bias and the latter variability. In

conjunction the two mechanics yield a marked improvement for inference procedures that rely on LRV estimators.

Both SV and zero lugsail estimators rely on a bandwidth parameter, a critical component for the estimation process. For SV estimators bandwidth selection typically revolves around the bias and variance of a LRV estimator. Most SV bandwidth rules do not apply to the zero lugsail estimator because of its optimal bias properties. Currently no guidelines exist for selecting a bandwidth for the zero lugsail estimator. We propose a testing optimal bandwidth rule for zero lugsail estimators when relying on nonstandard limiting distributions. With this procedure we can greatly improve bias, account for variability, and obtain an estimator optimized for inference.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Random vector processes appear in time series, econometrics, spectral analysis, and steady state simulation. Although the practical applications, set ups, assumptions, and priorities of these realms differ, the problem space is notably similar, where each field works with serially correlated multivariate data. With correlated data it is difficult to estimate the long run variance (LRV), which we denote as $\Omega$. The LRV is the product of the sample size and asymptotic variance of the mean vector, and is a critical component for inference procedures. Estimating $\Omega$ is especially difficult when the error structure is unknown and the data is positively correlated. Andrews [1], Newey & West [41], and Priestley [49] are prominent works that investigate this issue.

It is well known that under reasonable assumptions that the asymptotic LRV is equal to an infinite sum of autocovariance matrices, $\Omega = \sum_{s=-\infty}^{\infty} \Gamma(s)$ [11, 22, 67]. One may have a natural instinct to estimate the LRV by using a sum of estimated autocovariance matrices that spans the range of the data, i.e. $\hat{\Omega}_T^{(N)} = \sum_{s=-(T-1)}^{T-1} \hat{\Gamma}(s)$. However,

this estimator is problematic for four reasons. First, in a finite sample setting estimated autocovariance matrices with large lags suffer from boundary effects and behave sporadically. Second, the estimator is not consistent, increasing the sample size does not increase stability [49, Section 6.2.3]. Third, test statistics depending on LRV estimators typically do not converge quickly to a standard limiting distribution due to the estimation of the LRV [1, 25]. This problem is not specific to the $\hat{\Omega}_T^{(N)}$, but to LRV estimators generally. Lastly, while $\Omega$ is asymptotically positive semi-definite, $\hat{\Omega}_T^{(N)}$ may not be. Lack of positive semi-definiteness is typically the easiest issue to address [35, 36, 37, 61].

To combat these issues several alternative estimators to $\hat{\Omega}_T^{(N)}$ have been proposed, e.g. spectral variance (SV) [1, 41, 49], steep origin [46], flat top [47, 48], and lugsail estimators [61]. These methods apply some sort of weighting scheme to the autocovariance matrices to control boundary effects and the general estimation process. SV estimators are perhaps the most common and consist of two tuning parameters, a kernel function $\kappa^*$ and a bandwidth parameter $b \in [0, 1]$. Define the SV estimator as

$$\hat{\Omega}_T = \sum_{h=-(T-1)}^{T-1} \kappa^* \left( \frac{h}{bT} \right) \hat{\Gamma}(h). \tag{1.1}$$

The kernel function generates weights and is selected from a class of functions $\mathscr{K}_1$ with various restrictions discussed in Chapter 2. The bandwidth is the proportion of autocovariance matrices that are given a non-zero weight. Typically the only autocovariance matrix given a unit weight is at lag zero, and the rest of the weights decrease as the lags increase [49]. The SV estimator can be re-parameterized using a truncation parameter $M = bT$, the lag of the largest non-zero autocovariance matrix. However, we avoid this construction to keep the notation consistent throughout. A chief benefit of the SV estimator

is that with the kernel function and tuning parameters we can control the general variability of $\hat{\Omega}_T$. However, it comes at a cost of having a negatively biased estimator in the presence of positive correlation [1, 14].

When assuming that $bT \to \infty$ but at a slower rate than $T \to \infty$, the SV estimator is consistent. We refer to the results generated under this construction as small-$b$ asymptotic theory and to SV estimators as being heteroskedastic and autocorrelation consistent (HAC) [41, 65]. Under this framework standard limiting distributions can be used; however, they are known to not represent finite data sets well, especially in the presence of high correlation [1, 17, 40]. To combat this issue research has been dedicated to optimizing bandwidth and kernel selection. The foremost rule for classical SV estimators under this framework is based on asymptotic mean squared error (AMSE), but the discrepancy still persists [1].

Keifer & Vogelsang [24, 25] considered foregoing the assumption that $bT \to \infty$ and instead assumed that $b$ is a fixed number. We refer to the results generated under this construction as fixed-$b$ asymptotic theory. In this context the SV estimator is no longer consistent and we instead refer to it as being heteroskedastic and autocorrelation robust (HAR) [14, 39]. Standard limiting distributions are no long valid, but instead nonstandard distributions which better capture the distributional properties of estimating the LRV can be used [14, 24, 25]. The nonstandard limiting theory opens different avenues for accounting for the variability of estimating $\Omega$ and optimizing the tuning parameters.

The steep origin, flat top, and lugsail estimators can all be considered as transformations or generalizations of the traditional SV estimator. In addition to the bandwidth and kernel, these estimators rely on at least one other tuning parameter. The steep origin

uses a parameter that exponentiates the weights generated by kernels in $\mathscr{K}_1$ [46]. In contrast, the flat top and lugsail estimators introduce parameters that inflate weights generated by kernels in $\mathscr{K}_1$, and have the unique property that multiple lags near the origin can be given a unit weight [48, 61]. The lugsail estimator goes a step further and allows some weights to exceed one to offset the negative bias inherent to these estimators [61]. The inflated weights of the lugsail estimator greatly decrease the bias while still controlling for boundary effects.

Weighted orthonormal series (WOS) and series estimators are yet another alternative. The WOS estimator is similar to a SV estimator but it approximates the weights generated by $\kappa^* \in \mathscr{K}_1$ using a series of $K$ orthonormal basis functions [29, 53]. The series estimator is the same as the WOS estimator except that it approximates the weights of a steep origin estimator instead [52].

In addition, batch estimators have also been proposed as LRV estimators and are especially popular when working with highly correlated data sets [38]. In particular, the batch means estimator [7, 32] is used in Markov Chain Monte Carlo (MCMC) settings, and a comparable univariate version is used in econometric settings, which we will refer to as a batch betas estimator [17, 40]. The idea of a batch estimator is to separate the data into a smaller subsets, or batches. Batch statistics then act as data points which should be less correlated then the original data, and the batch statistics are then combined via a weighting scheme. This construction decreases discrepancy from the exact test statistic distribution and limiting distribution [14, 40]. Some connections have been drawn between the batch beta estimator to the $t$-distribution [14, 40], and the fixed-$b$ distribution through

a connection to WOS estimators under specific settings [14, 27]. However, to the best of our knowledge batch estimators have not been connected to the fixed-$b$ distribution generally.

When data sets have a high positive correlation the problems inherent to LRV estimation are pronounced [1, 40]. To avoid these problems practitioners typically filter observations as a preprocessing step to decorrelate data. Parameter selection for these LRV estimators is typically driven by the bias and the variability of the estimator. Data filtering is used to mitigate bias issues in the presence of high positive correlation, and down weighing autocovariance matrices is a key component that controls variability. The lugsail estimator defies common practice by inflating the weights to decrease the bias at the cost of variability. Although lugsail estimators have high variability, they free practitioners to collect data at a higher frequency in the presence of high positive correlation. Lugsail estimators are expected to perform well because the inflated weights decrease the bias and the ability to record data at a higher sampling frequency improves power [1, 2, 43].

Inference motivated loss functions using testing error rates, power, and coverage probability error have been proposed for the SV [54, 56], series, steep origin [53], and WOS estimators [29, 27] using fixed-$b$ limiting theory. Most rules rely on bias properties that lugsail estimators do not possess, and hence do not apply. To the best of our knowledge, lugsail estimators do not currently have guidelines for finding an optimal bandwidth in either the small-$b$ or fixed-$b$ setting. The lugsail estimator using 'zero' lugsail settings, i.e. zero lugsail estimator, is the most elusive as the asymptotic bias of the estimator is zero [61]. It also the most widely applicable of the lugsail settings because of this property.

The LRV estimators discussed are not distinct, and all overlap with at least one other type of estimator under specific settings. For example, there are strong connections between the lugsail and SV estimators, and connections between the lugsail, flat top, and batch means estimators under specific settings. We focus on the zero lugsail estimator due to its flexibility, versatility, relationship to other estimators, and optimal bias properties.

The focus of this document is to address issues inherent to LRV estimation and inference, and to obtain an optimal inference procedure. Specifically we concentrate on the zero lugsail estimator and use fixed-$b$ limiting theory to find a testing optimal bandwidth rule that accounts for the estimator's ideal bias characteristics. We assume the data is positively correlated unless otherwise specified. In addition, we also focus on highly correlated data and the effects thereof. Chapter 2 formerly establishes assumptions and the problem space. Chapters 3 and 4 verify and derive properties related to the test statistic using infinite and finite samples, respectively. Chapter 5 derives expressions for the testing error rates and establishes an optimal bandwidth rule. Chapter 6 discusses practical considerations. Lastly, Chapter 7 concludes with a simulation study, application, and brief discussion. Our results largely follow Sun [54], with the major contributions being verifying and extending fixed-$b$ results to include lugsail estimators, incorporating more information related to the bias, and deriving an optimal bandwidth rule for the zero lugsail estimator.

# Chapter 2

# Autocorrelation Robust Testing and Long Run Variance Estimators

## 2.1 Set Up and Assumptions

We use a standard generalized method of moments (GMM) framework which is valid for a wide variety of applications [12, 24, 25, 54]. Consider a vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^p$. Let $\theta_0$ be the true value of $\theta$, and $v_t$ denote a vector of observed values at time $t$. Assume the following moment conditions,

$$E[f(v_t; \theta_0)] = 0$$

where $t = (1, \ldots, T)$, $f(v_t; \theta)$ is a $m \times 1$ vector of twice differentiable continuous real valued functions, $m \geq p$, and $E[f(v_t; \theta)]$ has rank $p$, i.e. $E\left(\frac{\partial f(v_t;\theta)}{\partial \theta'}\right) = p$. The GMM estimator of $\theta$ is defined as

$$\hat{\theta}_T := \underset{\theta \in \Theta}{argmin} \frac{1}{T} g_T(\theta)' W_T g_T(\theta)$$

where $g_t(\theta) = \frac{1}{T}\sum_{j=1}^{t} f(v_j;\theta)$, and $W_T$ is a $m \times m$ weighting matrix. The matrix $W_T$ has a subscript $T$ to emphasize that this weighting matrix can be dependent on the sample. Define

$$\begin{aligned} G_t(\theta) :=& \frac{\partial g_t(\theta)}{\partial \theta'} \\ =& \frac{1}{T}\sum_{j=1}^{t} \frac{\partial f(v_j;\theta)}{\partial \theta} \\ G_o(\theta) :=& E\left[\frac{\partial f(v_j;\theta)}{\partial \theta}\right]. \end{aligned}$$

We consider the following set of assumptions which are common under the fixed-$b$ asymptotic framework [24, 25, 27, 54]. Let $\xrightarrow{p}$ and $\xrightarrow{d}$ indicate convergence in probability and convergence in distribution, respectively.

**Assumption 1:** $\hat{\theta}_T \xrightarrow{p} \theta_0$ *and* $\theta_0$ *is an interior point of* $\Theta$.

**Assumption 2:** $G_{\lfloor rT \rfloor}(\tilde{\theta}_T) \xrightarrow{p} rG_0$ *uniformly in* $r \in [0,1]$ *for any* $\tilde{\theta}_T$ *whose elements are between the corresponding elements of* $\hat{\theta}_T$ *and* $\theta_0$.

**Assumption 3:** $W_T$ *is positive semi-definite and* $W_T \xrightarrow{p} W_\infty$ *where* $W_\infty$ *is a matrix of constants and* $G_0' W_\infty G_0$ *is positive definite.*

**Assumption 4:** $T^{-1/2}\sum_{t=1}^{\lfloor rT \rfloor} u_t \xrightarrow{d} \Lambda B_d(r)$ *where*

$$\begin{aligned} \Omega =& \sum_{j=-\infty}^{\infty} E\left[f(v_t;\theta_0) f(v_{t-j},\theta_0)'\right] \qquad (2.1) \\ =& \sum_{j=-\infty}^{\infty} E[u_t, u_{t-j}'] \\ =& \sum_{j=-\infty}^{\infty} \Gamma(j) \\ =& \Lambda\Lambda' \end{aligned}$$

*where $B_d(r)$ is a standard d-dimensional Brownian motion, and $\Gamma(j)$ is the autocovariance matrix at lag $j$.*

**Assumption 5:** *(i) The sequence $u_t$ is a stationary Gaussian process. (ii) For any $c \in \mathbb{R}^d$, the spectral density of $c'u_t$ is bounded above and away from zero in a neighborhood around the origin. (iii) For some $r \in [0, 2+\eta]$ and $\eta > 0$,*

$$h_r := \sum_{s=-\infty}^{\infty} |s|^r |\Gamma(s)| < \infty. \tag{2.2}$$

Assumptions 1-3 are straight forward and standard in the literature. Assumption 4 is required for the functional central limit theorem (FCLT). This assumption is less strict than in other settings. For example, Andrews [1] requires that $f(v_t; \theta_0)$ is a zero mean fourth-order $\alpha$-mixing process which implies assumption 4. See works by [45, 54, 58, 64] for a larger discussion on sufficient conditions under various settings. Assumption 5 provides conditions for deriving the bias of LRV estimators and obtaining expressions for the distribution of the test statistics under fixed-$b$ assumptions [27, 54].

**Lemma 1:** *Under assumptions 1 - 4,*

$$\begin{aligned} T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} &(G_0 W_\infty G_0)^{-1} G_0 W_\infty \Lambda B_p(1) + o_p(1) \\ \sim &N(0, V) \end{aligned}$$

*where $V = (G_0' W_\infty G_0)^{-1}(G_0' W_\infty \Lambda\Lambda' W_\infty G_0')(G_0' W_\infty G_0)^{-1}$.*

*Begin proof.*

Following the work of [25, 54], under Assumption 2 we have the following Taylor Series approximation

$$g_T(\hat{\theta}_T) \approx g_T(\theta_0) + G_T(\theta_0)(\hat{\theta}_T - \theta_0) + o_p(1).$$

Multiply both sides of the equation by $G_T(\hat{\theta}_T)'W_T$,

$$G_T(\hat{\theta}_T)'W_T g_T(\hat{\theta}_T) \approx G_T(\hat{\theta}_T)'W_T g_T(\theta) + G_T(\hat{\theta}_T)'W_T G_T(\theta_0)(\hat{\theta}_T - \theta_0) + o_p(1).$$

Observe that we pick $\hat{\theta}_T$ such that it minimizes $g_T(\hat{\theta}_T)W_T g_T(\hat{\theta}_T)$. Alternatively, we can think of $\hat{\theta}_T$ as the estimator such that $G_T'(\hat{\theta}_T)W_T g_T(\hat{\theta}_T) = 0$ by using the chain rule,

$$0 \approx G_T(\hat{\theta}_T)'W_T g_T(\theta) + G_T(\hat{\theta}_T)'W_T G_T(\theta_0)(\hat{\theta}_T - \theta_0) + o_p(1).$$

Solve for $(\hat{\theta}_T - \theta_0)$ and multiply both sides by $T^{-1/2}$,

$$T^{1/2}(\hat{\theta}_T - \theta_0) \approx \left[G_T(\hat{\theta}_T)'W_T G_T(\theta_0)\right]^{-1} G_T(\hat{\theta}_T)'W_T T^{-1/2} g_T(\theta_0).$$

Under Assumption 3 and Assumption 4 we have the following result,

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} (G_0 W_\infty G_0)^{-1} G_0 W_\infty \Lambda B_p(1) + o_p(1).$$

■

Let $r()$ be a $d \times 1$ vector of twice differentiable functions, and $R(\theta) = \frac{\partial r(\theta)}{\partial \theta'}$. Consider the following set of hypotheses,

$$\begin{aligned} H_0 : &\quad r(\theta_0) = 0 \\ H_A : &\quad r(\theta_0) \neq 0. \end{aligned} \tag{2.3}$$

Using the $\delta$-method and Lemma 1 we know that

$$T^{1/2} r(\hat{\theta}_T) \xrightarrow{d} N\left(0, R(\theta_0)'V R(\theta_0)\right).$$

The typical Wald test statistics in this setting would be

$$\begin{aligned} F_T =& (T) r(\hat{\theta}_T) \left[ R(\hat{\theta}_T)' \hat{V} R(\hat{\theta}_T) \right]^{-1} r(\hat{\theta}_T)/d \\ t_T =& T^{-1/2} r(\hat{\theta}_T) \left[ R(\hat{\theta}_T)' \hat{V} R(\hat{\theta}_T) \right]^{-1} \end{aligned} \tag{2.4}$$

where

$$\hat{V} = \left[ G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T) \right]^{-1} G_T(\hat{\theta}_T)' W_T \hat{\Sigma}_T W_T G_T(\hat{\theta}_T) \left[ G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T) \right]^{-1}$$

and $(\hat{\theta}_T, \hat{\Sigma}_T)$ are consistent estimates of $(\theta, \Sigma)$.

## 2.2 Long Run Variance Estimators

Many estimators for the LRV have been proposed, and SV estimators are perhaps the most common. SV estimators are seen in many applications and fields, and go by different names. In econometrics they are called heteroskedastic and autocorrelation consistent/robust (HAC/HAR) estimators [41, 65], and in time series they may be called kernel estimators [1, 11]. These estimators are also used in the frequency domain [49], and in MCMC applications [62]. SV estimators consist of two tuning parameters, a kernel function and a bandwidth parameter. We can construct the SV estimator in the following ways,

$$\begin{aligned} \hat{\Omega}_T =& \sum_{h=-(T-1)}^{T-1} \kappa^* \left( \frac{h}{bT} \right) \hat{\Gamma}(h) \\ =& \sum_{h=-(T-1)}^{T-1} \kappa^* \left( \frac{h}{bT} \right) \frac{1}{T} \sum_{t=1}^{T-h} \hat{u}_t \hat{u}_{t+h}' \\ =& \frac{1}{T} \sum_{t=1}^{T} \sum_{h=1}^{T} \kappa^* \left( \frac{t-h}{bT} \right) \hat{u}_t \hat{u}_h' \end{aligned} \tag{2.5}$$

where $\hat{\Gamma}(h)$ is an estimator for the autocovariance function $\hat{\Gamma}(h)$, and $\hat{u}_t$ is a plug in estimator for $u_t$,

$$\hat{u}_t = -R(\hat{\theta}_T)\left(G_T'(\hat{\theta}_T)W_T G_T(\hat{\theta}_T)\right)^{-1} G_T'(\hat{\theta}_T)W_T f(v_t, \hat{\theta}_T).$$

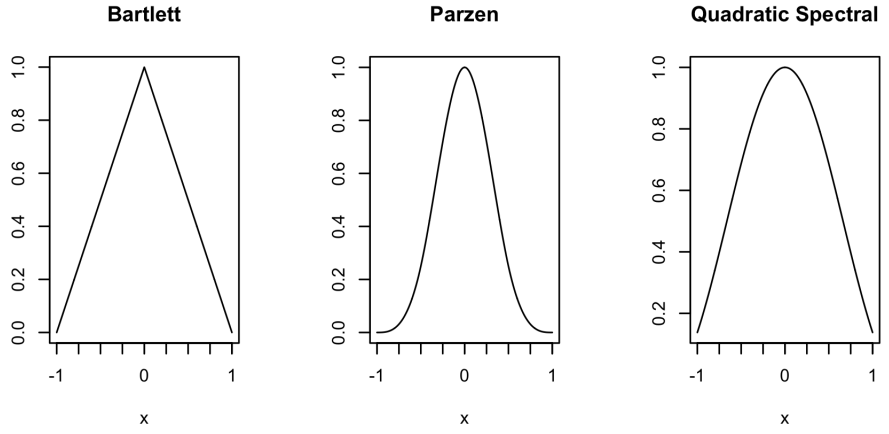


Figure 2.1: Mother Kernels, class $\mathscr{K}_1$

We typically choose the kernel for an SV from the following class of functions:

$$\mathscr{K}_1 = \left\{\kappa^* : \mathbb{R} \to [-1,1] \middle| \kappa^*(x) = \kappa^*(-x), \kappa^*(0) = 1, c_2 < \infty, \int_0^\infty \kappa^*(x)x dx < \infty, \right.$$
$$\left. \kappa^* \text{ is piece wise smooth}, K^*(\omega) \geq 0, \forall \omega > 0 \right\},$$

where $K^*(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \kappa^*(u)\exp(-iu\omega)du$ is the Fourier transformation of $\kappa^*$ and $c_i = \int_{-\infty}^{\infty} (\kappa^*(u))^i \, du$. The most common functions in this class only produce non-negative weights, with a notable exception being the quadratic spectral kernel, defined below. Requiring that kernel functions be piece wise smooth is convenient both mathematically and in practice. Symmetry ensures that are estimator will not have imaginary numbers [49].

Restricting the squared integral and first moment of the kernel to be finite is necessary to derive fixed-$b$ distributional expressions, and requiring $K^*(u)$ to be non-negative ensures that the SV estimator is positive semi-definite. The following kernel functions are all common, belong to class $\mathscr{K}_1$, and are illustrated in Figure 2.1:

$$\text{Bartlett} \qquad \kappa(x) = \begin{cases} 1 - |x| & \text{if } |x| \leq 1 \\ 0 & \text{if otherwise} \end{cases};$$

$$\text{Parzen} \qquad \kappa(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{if } 0 \leq |x| \leq \frac{1}{2} \\ 2(1 - |x|)^3 & \text{if } \frac{1}{2} \leq |x| \leq 1 \\ 0 & \text{if otherwise} \end{cases};$$

$$\text{Quadratic Spectral} \qquad \kappa(x) = \frac{25}{12\pi^2 x^2} \left( \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right).$$

Alternatively, we can construct the SV estimator in the frequency domain by evaluating the spectral density function at frequency zero,

$$\begin{aligned} \hat{h}(\omega) &= \int_{-\pi}^{\pi} I_T(\theta) W_T(\omega - \theta) d\theta \\ &= \frac{1}{2\pi} \sum_{h=-(T-1)}^{T-1} \kappa^* \left( \frac{h}{bT} \right) \hat{\Gamma}(h) \exp(-ih\omega) \end{aligned}$$

where

$$I_T(\omega) = \frac{1}{2\pi} \sum_{s=-(T-1)}^{T-1} \hat{\Gamma}(s) \exp(-is\omega)$$

$$W_T(\omega) = \frac{1}{2\pi} \sum_{s=-(T-1)}^{T-1} \kappa^* \left(\frac{s}{bT}\right) \exp(-is\omega)$$

$$= bT \sum_{j=-\infty}^{\infty} K^* \left(bT(\omega + 2\pi j)\right).$$

The function $I_T(\omega)$ is the estimated periodogram and $W_T(\omega)$ is the weight assigned to each frequency. The weight function is directly related to the Fourier transformation of the kernel, $W_T(\omega) = \lfloor bT \rfloor \sum_{j=-\infty}^{\infty} K^* \left(\lfloor bT \rfloor (\omega + 2\pi j)\right)$ [49, page 447].

It is well known that using standard small-$b$ limiting theory results in an inflated Type 1 error rate [1, 25]. The inflated error rate is largely attributed to the bias and the variability of the SV estimator. For illustrative purposes we outline sources of bias and variability below for a positively correlated second order stationary process.

$$
\begin{aligned}
\Omega &= \sum_{h=-\infty}^{\infty} \Gamma(h) \\
&\geq \sum_{h=-(T-1)}^{T-1} \Gamma(h) && \text{Finite Sampling} \quad (2.6) \\
&\approx \sum_{h=-(T-1)}^{T-1} \sum_{t=1}^{T-h} \frac{1}{T-h} E[\hat{u}_t, \hat{u}'_{t+|h|}] && \text{Estimation} \quad (2.7) \\
&\geq \sum_{h=-(T-1)}^{T-1} \sum_{t=1}^{T-|h|} \frac{1}{T} E[\hat{u}_t \hat{u}'_{t+|h|}] && \text{Autocov. Estimator} \quad (2.8) \\
&= \sum_{h=-(T-1)}^{T-1} \hat{\Gamma}(h) \\
&= \sum_{h=-\lfloor bT \rfloor}^{\lfloor bT \rfloor} E[\hat{\Gamma}(h)] + \sum_{h=-(T-1)}^{-(\lfloor bT \rfloor+1)} E[\hat{\Gamma}(h)] + \sum_{h=(\lfloor bT \rfloor+1)}^{T-1} E[\hat{\Gamma}(h)] \\
&\geq \sum_{h=-\lfloor bT \rfloor}^{\lfloor bT \rfloor} E[\hat{\Gamma}(h)] + 0 + 0 && \text{Bandwidth} \quad (2.9) \\
&\geq \sum_{h=-\lfloor bT \rfloor}^{\lfloor bT \rfloor} \kappa^* \left( \frac{h}{bT} \right) E[\hat{\Gamma}(h)] && \text{Kernel Function} \quad (2.10)
\end{aligned}
$$

The first source of bias (2.6) is unavoidable in the estimation process. Recall the LRV is comprised of an infinite sum of the autocovariance matrices, but due to finite sampling it is only feasible to estimate the first $T-1$ autocovariance matrices. However, it is usually not concerning to ignore the autocovariance matrices beyond lag $T-1$ because they typically tend towards zero as the lag increases.

The next component (2.7) introduces variability due to the estimation process of $\Omega$ in general. This variability is also influenced by estimating parameters to obtain an estimate for $u_t$. We typically use plug-in estimated parameters instead of the hypothesized parameters in the estimation process because the former has higher power [10, 55]. Observe

that the test statistic $F_T$ has both a random numerator and denominator. By Slutsky's theorem the test statistic converges to a $\chi^2$ distribution as the sample goes to infinity, but this limiting distribution does not capture the variability in the denominator. For finite samples, the variability inherent to estimating $\Omega$ has a cascading affect on the variability of test statistic. The tails of the true distribution of the test statistic using finite samples are heavier than the $\chi^2$ limiting distribution because of the two sources of variability. A analogous effect is observed in the classical univariate iid setting where we use the $t$-distribution instead of the standard normal distribution to capture the extra variability induced by estimating the variance. When the sample size exceeds 30 the distributions are the same up to several decimal places. In contrast, for highly correlated data the sample size must be exceptionally large to have a similar effect.

Next we see that the lag-$h$ autocovariance estimator is biased by having $T$ in the denominator instead of $T - |h|$. This biased autocovariance estimator is common, and is used in most computer programs because it has less variability. This bias is less pronounced when $h$ is small relative to $T$, and when $h$ is closer to $T$ the decrease in variability by using the biased estimator is typically a worthy trade [49, Section 5.3.3]. It is also needed to ensure the estimate is positive semi-definite.

The bandwidth parameter (2.9) dictates the proportion of non-zero autocovariance matrices we considered for the estimator, further limiting the amount of autocovariance matrices incorporated in the estimation process. It also proportionally effects the weights of the kernel function. We will refer to the expected value of the sum of the autocovariance matrices beyond lag $\lfloor bT \rfloor$, including those beyond lag $T - 1$, as the *finite sampling bias*.

In general, the bias and variability of the estimator of $\Omega$ have an inverse relationship with $b$. The variability of the estimator decreases as the bandwidth parameter decreases, and in contrast, the bias increases as the bandwidth parameter decreases. This relationship will be further illustrated in Theorem 1.

We also have bias attributed to the kernel function (2.10), which is typically one of the largest sources of bias and directly affects variability [5]. The kernel function down weights the remaining non-zero autocovariance matrices creating a negative bias; however, this down weight also scales the variability attributed to each autocovariance matrix to be smaller. We refer to the *kernel bias* as the expected value of the sum of the difference between the down weighted autocovariances and their respective true values.

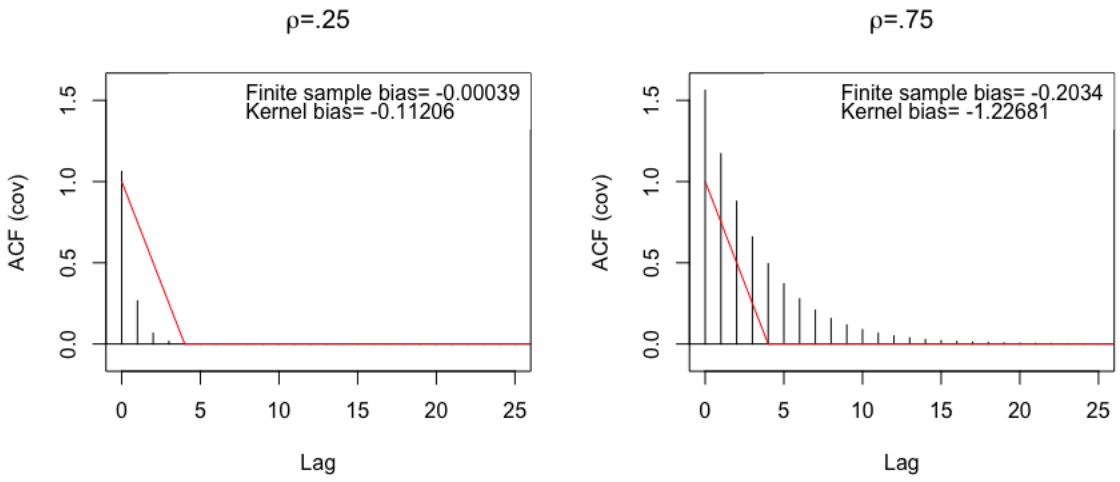


Figure 2.2: Autocovariance functions for an AR(1) model correlation coefficient $\rho = .25$ and $\rho = .75$ using a bandwidth of $b = 0.022$.

Not explicitly represented here is the dimensionality bias, the biases listed above become more impactful when the number of restrictions under null hypotheses is larger,

which is related to the number of parameters in the model because $d \geq p$. There is also the strength of the correlation to consider. Consider estimating the LRV from a standard AR(1) model with Bartlett kernel $\kappa^*$, sample size $T = 200$, and $b = 0.022$. In Figure 2.2 we plot the theoretical autocovariance function with autocorrelation coefficient $\rho = 0.25$ and $\rho = 0.75$, and overlaid the kernel function with the given bandwidth. The estimated LRV is the sum of the product of the estimated autocovariance and the weight generated by the kernel function. The kernel function and the finite sampling bias are listed in the top corner of each figure. Observe that correlation strength effects both the kernel and finite sampling bias, both of which are more pronounced in the presence of higher correlation.

Individually the components of the SV estimation process listed above are not typically of concern and several are unavoidable. However, together these components can have a cascading effect on the test statistic and consequently the testing procedures. Most of these aspects cause the test statistic to be larger than what we would otherwise expect, which is why we can observe an inflated Type 1 error rate for positively correlated second order stationary processes.

### 2.2.1 Lugsail Estimators

Several works have focused on optimizing SV estimators while restricting the kernels to class $\mathscr{K}_1$ [1, 27, 49, 57], the constant challenge being balancing bias and variability. Recently a new estimation procedure has been proposed by Vats and Flegal [61] to offset the negative bias inherent to the kernels in $\mathscr{K}_1$. This estimator has two characterizations.

It can be thought of as a linear combination of SV estimators with different bandwidths where $\kappa^* \in \mathscr{K}_1$,

$$\hat{\Omega}_T^{(L)} = \frac{1}{1-c}\hat{\Omega}_{T,b} - \frac{c}{1-c}\hat{\Omega}_{T,b/r}, \tag{2.11}$$

or as a SV estimator with a transformed kernel function that belongs in a broader class of kernels, $\mathscr{K}_2$. We reference the first interpretation as the lugsail estimator. For the second interpretation we refer to the transformed kernel as the *lugsail kernel*, and the original kernel function as the *mother kernel*, which is part of class $\mathscr{K}_1$. This terminology is similar to that used for the steep origin kernels [46]. Let $\mathscr{K}_2$ be the set of lugsail kernels,

$$\mathscr{K}_2 = \left\{ \kappa(\cdot) : \mathbb{R} \to (-\infty, \infty) \middle| \kappa^* \in \mathscr{K}_1, c \in [0,1), r \geq 1, \kappa(x) = \frac{1}{1-c}\kappa^*(x) - \frac{c}{1-c}\kappa^*(xr) \right\}.$$

Class $\mathscr{K}_2$ expands the range of kernel functions we can consider for SV estimators. When $r = 1$ and $c = 0$ then the lugsail kernel is equal to it's mother kernel, i.e. $\mathscr{K}_1 \subset \mathscr{K}_2$. Lugsail kernels do not necessarily share the same properties as the mother kernels. For example, for $\kappa \in \mathscr{K}_2 - \mathscr{K}_1$ the weights may exceed 1, and $K(\theta)$ is no longer guaranteed to be non-negative. For simplicity assume $\kappa \in \mathscr{K}_2$ and $\kappa^* \in \mathscr{K}_1$ unless otherwise stated.

For all kernels $\kappa \in \mathscr{K}_2$ we define their corresponding characteristic component as defined by Parzen [42],

$$q = max \left\{ q : q \in \mathbb{N}^+, g_q = \lim_{x \to 0} \frac{1 - \kappa(x)}{|x|^q} < \infty \right\} \tag{2.12}$$

where $g_q$ is the generalized derivative of the kernel. The statistics $q$ and $g_q$ are useful for describing specific kernel characteristics. For the Bartlett, Parzen, and quadratic spectral kernels $q$ is equal to 1, 2, and 2, respectively. A necessary condition for positive semi-definiteness is that $q \leq 2$ [49].

This lugsail estimator has four tuning parameters $b$, $c$, $r$, and the mother kernel. The aim of the new tuning parameters $(r, c)$ is to inflate the weight function to offset the negative kernel and finite sampling bias inherent to the mother kernel. There are no firm or optimal guidelines for picking $(r, c)$, but a rule-of-thumb has been recommended [61]. If any arbitrary sequence of random variables that has an underlying process similar to that of 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. The recommended settings for these three situations are in Table 2.1. The recommendations are in response to the strength of the correlation because the biases of the SV estimator are more pronounced for data sets with higher correlation [61]. Observe that the adaptive setting relies on an initial estimate for the parameter $b$. We recommend using the optimal zero lugsail bandwidth in Chapter 5 as the initial estimate, or another comparable testing optimal rule.

| Correlation | Lugsail Transformation | $r$ | $c_r$ |
|---|---|---|---|
| | Mother | 1 | 0 |
| Moderate | Zero | 2 | $r^{-q}$ |
| Moderate-High | Adapt | 2 | $\frac{\log(T)-\log(\lfloor bT \rfloor)+1}{r^q(\log(T)-\log(\lfloor bT \rfloor))+1}$ |
| High-Extreme | Over | 3 | $\frac{2}{(1+r^q)}$ |

Table 2.1: Lugsail Parameter Recommendations.

The zero lugsail kernel is designed to have an asymptotic kernel bias of zero, which will be apparent in Theorem 1. It is similar to that of a flat-top estimator in that it 'flattens' the kernel function around the origin [48]. In contrast, the adaptive and over lugsail kernels generate weights that exceed one, which over correct for the kernel bias to offset the finite

sampling bias. Figure 2.3 contains an illustration of the lugsail kernels that correspond to the settings listed in Table 2.1.

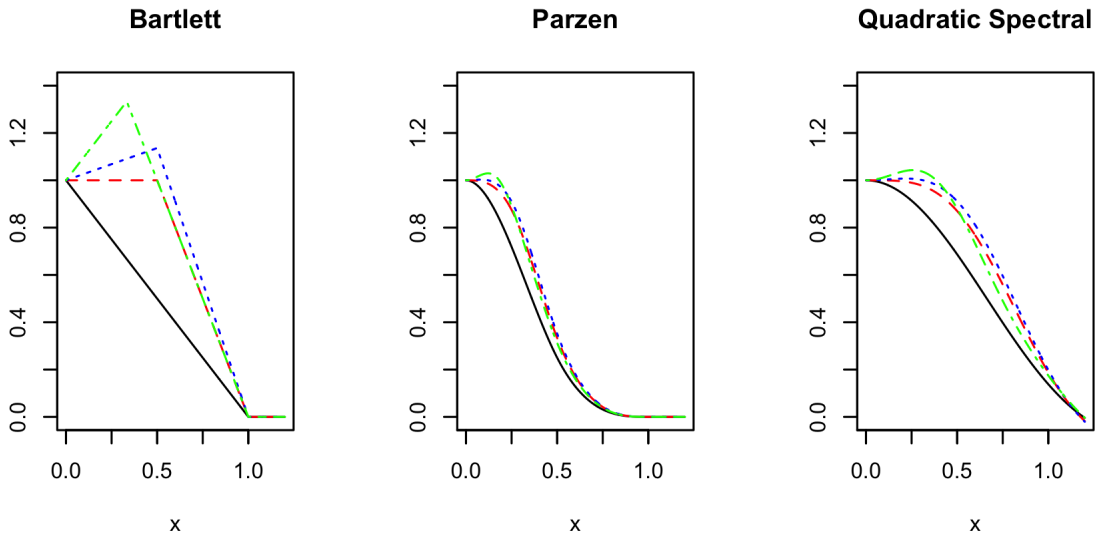


Figure 2.3: Kernel functions in $\mathscr{K}_2$ with mother (solid black), zero (dashed red), adaptive (dotted blue), and over (dot-dashed green) lugsail settings. Adaptive lugsail kernel was generated with the initial bandwidth $b = 0.022$.

**Theorem 1:** *Let $\kappa \in \mathscr{K}_2$, $q_0 \leq q$, and assume that $bT \to \infty$ as $T \to \infty$, but at a slower rate. Under Assumption 5iii*

$$\begin{aligned} Bias(\hat{\Omega}_T) &= -(bT)^{-q_0} g_{q_0} h_{q_0} + o\left((bT)^{-q_0}\right) \\ &= \left(\frac{1 - cr^{q_0}}{1 - c}\right)(bT)^{-q_0} g^*_{q_0} h_{q_0} + o\left((bT)^{-q_0}\right) \\ Var(vec(\hat{\Omega}_T)) &= (c_2 b)\left(I_{d^2} + \mathbb{K}_{dd}\right)\Omega \otimes \Omega + o(b) \end{aligned}$$

*where $\otimes$ is the Kronecker product, and $\mathbb{K}_{nm}$ is a $n \times m$ commutation matrix.*

*Begin proof.*

We follow the work of Priestly [49, pg 459] and Parzen [42]. Begin by plugging in the estimator for $\hat{\Omega}_T$, accounting for the bias of autocovariance function $\hat{\Gamma}(s)$, and rearranging terms,

$$\begin{aligned}
Bias(\hat{\Omega}_T) =& E\left(\hat{\Omega}_T\right) - \sum_{s=-\infty}^{\infty} \Gamma(s) \\
=& E\left(\sum_{s=-(T-1)}^{T-1} \kappa^*\left(\frac{s}{bT}\right)\hat{\Gamma}(s)\right) - \sum_{s=-\infty}^{\infty} \Gamma(s) \\
=& \sum_{s=-(T-1)}^{T-1} \kappa^*\left(\frac{s}{bT}\right) E\left(\hat{\Gamma}(s)\right) - \sum_{s=-\infty}^{\infty} \Gamma(s) \\
=& \sum_{s=-(T-1)}^{T-1} \kappa^*\left(\frac{s}{bT}\right)\left(1-\frac{|s|}{T}\right)\Gamma(s) - \sum_{s=-\infty}^{\infty} \Gamma(s) \\
=& \sum_{s=-(T-1)}^{T-1} \left(\kappa^*\left(\frac{s}{bT}\right)\left(1-\frac{|s|}{T}\right) - 1\right)\Gamma(s) \\
& - \sum_{|s|\geq T}^{\infty} \Gamma(s) \\
=& \sum_{s=-(T-1)}^{T-1} \left(\kappa^*\left(\frac{s}{bT}\right) - 1\right)\Gamma(s) - \frac{1}{T}\sum_{s=-(T-1)}^{T-1} |s|\Gamma(s) \\
& - \sum_{|s|\geq T}^{\infty} \Gamma(s). \qquad (2.13)
\end{aligned}$$

The bias expression in (2.13) has three components: the first term is caused by the kernel function alone, the second term is the 'interaction' between the kernel function and the autocovariance estimator, and the last term comes from finite sampling. Our given assumption implies that $((bT)^{q_0}/T) \to 0$ as $T \to \infty$. This lets us replace the last two terms with $o((bT)^{-q_0})$, and we are left with

$$Bias(\hat{\Omega}_T) = \sum_{s=-(T-1)}^{T-1} \left(\kappa^*\left(\frac{s}{bT}\right) - 1\right)\Gamma(s) + o\left((bT)^{-q_0}\right).$$

To assess the asymptotic bias we take the limit of the first term as $T$ goes to infinity. Under the given assumptions,

$\lim_{T\to\infty} \sum_{s=-(T-1)}^{T-1} \left(\kappa^*\left(\frac{s}{bT}\right) - 1\right)\Gamma(s)$

$$\begin{aligned}
&= -\lim_{T\to\infty} \sum_{s=-(T-1)}^{T-1} \left(1 - \kappa^*\left(\frac{s}{bT}\right)\right)\Gamma(s) \\
&= -\lim_{bT\to\infty} \sum_{s=-(T-1)}^{T-1} \left(1 - \kappa^*\left(\frac{s}{bT}\right)\right)\Gamma(s) \\
&= -\sum_{s=-(T-1)}^{T-1} \lim_{bT\to\infty} \left(1 - \kappa^*\left(\frac{s}{bT}\right)\right)\Gamma(s) \\
&= -\sum_{s=-(T-1)}^{T-1} \lim_{1/bT\to 0} \left(1 - \kappa^*\left(\frac{s}{bT}\right)\right)\Gamma(s) \\
&= -\sum_{s=-(T-1)}^{T-1} \lim_{s/bT\to 0} \left(1 - \kappa^*\left(\frac{s}{bT}\right)\right)\Gamma(s) \\
&= -\sum_{s=-(T-1)}^{T-1} \lim_{s/bT\to 0} \frac{\left(1 - \kappa^*\left(\frac{s}{bT}\right)\right)}{(s/bT)^{q_0}}\Gamma(s) bT^{-q_0}|s|^{q_0} \\
&= -\sum_{s=-(T-1)}^{T-1} g_q \Gamma(s) bT^{-q_0}|s|^{-q_0} \\
&= -(bT)^{-q_0} g_{q_0} h_{q_0}.
\end{aligned}$$

By linearity of expectations we can find the bias expression for $\kappa \in \mathscr{K}_2$. The proof in Theorem 1b can be found in [49, 61] among others.

■

| | | Bartlett | Parzen | QS |
|---|---|---|---|---|
| | $q$ | 1 | 2 | 2 |
| *Mother* | $c_1^*$ | 1.00 | 0.75 | 1.25 |
| | $c_2^*$ | 0.67 | 0.54 | 1.00 |
| | $g_q^*$ | 1.00 | 6.00 | 1.42 |
| *Zero* | $c_1$ | 1.5 | 0.80 | 1.52 |
| | $c_2$ | 1.33 | 0.60 | 1.29 |
| | $g_q$ | 0 | 0 | 0 |
| *Adaptive* | $c_1$ | 1.64 | 0.81 | 1.57 |
| | $c_2$ | 1.57 | 0.62 | 1.36 |
| | $g_q$ | -0.273 | -1.35 | -0.32 |
| *Over* | $c_1$ | 2 | 0.81 | 1.51 |
| | $c_2$ | 2.33 | 0.57 | 1.31 |
| | $g_q$ | -1.00 | -6.00 | -1.42 |

Table 2.2: Kernel Summary Statistics.

Through Theorem 1 we observe that the asymptotic kernel bias and asymptotic variability of the SV estimator is reflected through $g_q$ and $c_2$, respectively. In addition, the characteristic component $q$ plays a large role in the estimation process. We also note that the bias expression captures the bias from the kernel but nothing more. Table 2.2 contains a summary of useful kernel statistics. Notice how the generalized derivative changes under the different settings. The mother kernel's generalized derivative is $g_q^*$, and the corresponding lugsail generalized derivative is $g_q = \left(\frac{1-cr^q}{1-c}\right) g_q^*$. This indicates that the lugsail generalized derivative is typically less than the mother kernel's generalized derivative. We can also see the effect of over correction for the adaptive and over lugsail kernel because the sign for the expression of bias will change, which is by design. We observe from Table 2.2 and Theorem 1 that the more inflated the kernel function is, the more $c_2$ increases, and hence the variability of $\hat{\Omega}_T$.

With the exception of the zero lugsail kernel, lugsail kernels will have the same characteristic component ($q$) as the mother kernel. Due to its 'flatness' the zero-lugsail kernel has an infinite characteristic component, just like the flat top kernels and the truncated kernel ($\kappa(x) = 1$ for $x < b$, 0 elsewhere). Therefore $q_0$ in Theorem 1 is limited by the underlying model instead of the kernel function. For the standard class of autoregressive processes with moving average (ARMA) residuals we can think of the spectral density as possessing an essentially infinite number of derivatives [48]. This is a major advantage for the zero lugsail kernel function because the overall bias of the SV estimator is dominated by $(bT)^{-q}$.

### 2.2.2 Centered Kernels

Another modification we implement is centering our error terms. Let $\kappa_{bT}(t,s) = \kappa\left(\frac{t-s}{bT}\right)$, and observe $\kappa_b(\cdot,\cdot) : [0,1] \times [0,1] \to \mathbb{R}$. We continue to use the SV estimator with lugsail kernels but instead of using the raw estimated errors we use the de-meaned estimated error,

$$\hat{\Omega}_T = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \kappa_{bT}\,(t,s)\,(\hat{u}_t - \hat{\overline{u}})(\hat{u}_s - \hat{\overline{u}})'. \tag{2.14}$$

We can think of this procedure as another transformation of the kernel function, which we can see by factoring and rearranging terms, i.e.

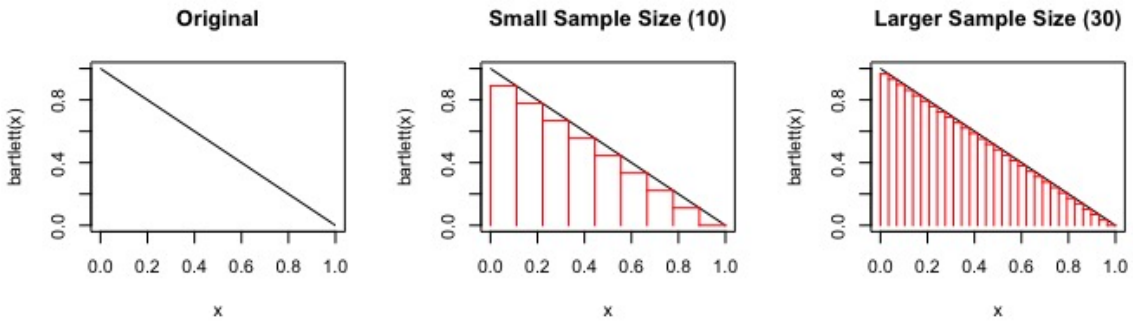


Figure 2.4: Riemann-Sum Approximation of Bartlett Kernel

$$\begin{aligned}\hat{\Omega}_T =& \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_{bT}(t,s)\left[\hat{u}_t\hat{u}_s' - \hat{u}_t\hat{\bar{u}}' - \hat{u}_s'\hat{\bar{u}} + \hat{\bar{u}}\hat{\bar{u}}'\right] \\ =& \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_{bT}(t,s)\hat{u}_t\hat{u}_s' - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_{bT}(t,s)\hat{u}_t\hat{\bar{u}}' \\ & - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_{bT}(t,s)\hat{\bar{u}}\hat{u}_s' + \hat{\bar{u}}\hat{\bar{u}}'\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_{bT}(t,s).\end{aligned}$$

Notice that $\sum_{s=1}^{T}\kappa_{bT}(t,s)\frac{1}{T} = \int_0^1 \kappa_b\left(\frac{t}{T},r\right)dr$ as $T\to\infty$ for any $t$, a Riemann integral. An illustration of this is in Figure (2.4). This lets us further simplify,

$$\begin{aligned}\hat{\Omega}_T =& \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_{bT}(t,s)\hat{u}_t\hat{u}_s' - \sum_{t=1}^{T}\left[\int_0^1 \kappa_b\left(\frac{t}{T},r\right)dr\right]\hat{u}_t\hat{\bar{u}}' \\ & - \sum_{s=1}^{T}\left[\int_0^1 \kappa_b\left(\tau,\frac{s}{T}\right)d\tau\right]\hat{\bar{u}}\hat{u}_s' + T\hat{\bar{u}}\hat{\bar{u}}'\int_0^1\int_0^1 \kappa_b(\tau,r)d\tau dr.\end{aligned}$$

We further substitute $\hat{\bar{u}}$ and rearrange terms,

$$\begin{aligned}
\hat{\Omega}_T =& \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_{bT}(t,s)\hat{u}_t\hat{u}_s' - \sum_{t=1}^{T}\left[\int_0^1 \kappa_b\left(\frac{t}{T},r\right)dr\right]\hat{u}_t\left\{\frac{1}{T}\sum_{s=1}^{T}\hat{u}_s\right\}' \\
&- \sum_{s=1}^{T}\left[\int_0^1 \kappa_b\left(\tau,\frac{s}{T}\right)d\tau\right]\left\{\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t\right\}\hat{u}_s' \\
&+ T\left\{\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t\right\}\left\{\frac{1}{T}\sum_{s=1}^{T}\hat{u}_s\right\}'\int_0^1\int_0^1 \kappa_b(\tau,r)d\tau dr \\
=& \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_{bT}(t,s)\hat{u}_t\hat{u}_s' - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\int_0^1 \kappa_b\left(\frac{t}{T},r\right)dr\right]\hat{u}_t\{\hat{u}_s\}' \\
&- \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\int_0^1 \kappa_b\left(\tau,\frac{s}{T}\right)d\tau\right]\{\hat{u}_t\}\hat{u}_s' \\
&+ \frac{1}{T}\sum_{t=1}^{T}\sum_{t=1}^{T}\{\hat{u}_t\}\{\hat{u}_s\}'\int_0^1\int_0^1 \kappa_b(\tau,r)d\tau dr \\
=& \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\kappa_{bT}(t,s) - \int_0^1 \kappa_b\left(\frac{t}{T},r\right)dr - \int_0^1 \kappa_b\left(\tau,\frac{s}{T}\right)d\tau\right. \\
&\left. + \int_0^1\int_0^1 \kappa_b(\tau,r)d\tau dr\right]\hat{u}_t\hat{u}_s'.
\end{aligned}$$

Therefore, the estimator in equation (2.14) is equal to

$$\hat{\Omega}_T = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{\kappa}\left(\frac{t-s}{bT}\right)\tilde{u}_t\tilde{u}_s', \tag{2.15}$$

where

$$\tilde{\kappa}_{bT}(t,s) := \kappa_{bT}(t,s) - \int_0^1 \kappa_b\left(\frac{t}{T},r\right)dr - \int_0^1 \kappa_b\left(\tau,\frac{s}{T}\right)d\tau + \int_0^1\int_0^1 \kappa_b(r,\tau)drd\tau \tag{2.16}$$

and $\tilde{u}_t = \hat{u}_t - \hat{\bar{u}}$. The kernel function $\tilde{\kappa}$ is known as the 'centered' kernel. We define a new class of centered lugsail kernels $\mathscr{K}_3$,

$$\mathscr{K}_3 = \left\{\tilde{\kappa} : \mathbb{R} \to [-1,1]\middle| \text{ where } \tilde{\kappa} \text{ is defined in (2.16) with } \kappa \in \mathscr{K}_2\right\}.$$

The centered kernel function was proposed by Hall [10] and used by others [30, 54, 55]. In GMM estimation procedures it is assumed that the model is correctly specified

and that the moment conditions in (2.1) are valid. However, when tests are over specified or mis-specified the power of the test suffers. A more powerful test is obtained when used the centered kernel regardless if the moment conditions in (2.1) are correct. The centered kernels inherit the properties described in Table 2.2 of the kernel they are constructed with, although they induce an additional bias term due to centering the errors [49, 63], which we refer to as the *de-meaning bias*.

**Corollary 1:** *Let $\tilde{\kappa} \in \mathscr{K}_3$, $q_0 \leq q$, and assume that $bT \to \infty$ as $T \to \infty$, but at a slower rate. Under Assumption 4 and 5iii*

$$Bias(\hat{\Omega}_T) = \left(\frac{1 - cr^{q_0}}{1 - c}\right)(bT)^{-q_0} g^*_{q_0} h_{q_0} - \Omega c_1 b + o\left((bT)^{-q_0}\right).$$

*Begin proof.*

In Theorem 1a we used errors terms that are not de-meaned, which have the expected value: $E(\hat{\Gamma}(s)) = \left(1 - \frac{|s|}{T}\right)\Gamma(s)$, due to the bias of the autocovariance estimator. Suppose that $E(\hat{u}_t) = \mu$, then this standard estimator for $\Gamma(s)$ can be rewritten as, $\hat{\Gamma}(s) = \frac{1}{T}\sum_{t=1}^{T-s}(\hat{u}_t - \mu)(\hat{u}_{t+s} - \mu)'$. In most frameworks under the null hypothesis we have $\mu = 0$. However, we are instead relying on the de-meaned errors and want to find the expected value for $\tilde{\Gamma}(s) = \frac{1}{T}\sum_{t=1}^{T-|s|}(\hat{u}_t - \hat{\overline{u}})(\hat{u}_{t+|s|} - \hat{\overline{u}})'$. We can do this by breaking down the components of $\hat{\Gamma}(s)$,

$$
\begin{aligned}
\hat{\Gamma}(s) =& \frac{1}{T}\sum_{t=1}^{T-|s|}(\hat{u}_t-\mu)(\hat{u}_{t+|s|}-\mu)' \\
=& \frac{1}{T}\sum_{t=1}^{T-|s|}\left[(\hat{u}_t-\hat{\overline{u}})+(\hat{\overline{u}}-\mu)\right]\left[(\hat{u}_{t+|s|}-\hat{\overline{u}})+(\hat{\overline{u}}-\mu)\right]' \\
=& \frac{1}{T}\sum_{t=1}^{T-|s|}\left[(\hat{u}_t-\hat{\overline{u}})(\hat{u}_{t+s}-\hat{\overline{u}})'+(\hat{u}_t-\hat{\overline{u}})(\hat{\overline{u}}-\mu)'\right. \\
& \left.+(\hat{\overline{u}}-\mu)(\hat{u}_{t+|s|}-\hat{\overline{u}})'+(\hat{\overline{u}}-\mu)(\hat{\overline{u}}-\mu)'\right].
\end{aligned}
$$

We approximate $\frac{1}{T}\sum_{t=1}^{T}(u_t-\mu)\approx\frac{1}{T}\sum_{t=1}^{T-|s|}(u_t-\mu)\approx 0$ to further simplify,

$$
\hat{\Gamma}(s)\approx\frac{1}{T}\sum_{t=1}^{T-|s|}\left[(\hat{u}_t-\hat{\overline{u}})(\hat{u}_{t+|s|}-\hat{\overline{u}})'\right]+\frac{T-|s|}{T}(\mu-\hat{\overline{u}})(\hat{\overline{u}}-\mu)'.
$$

By rearranging terms, using Assumption 4, and taking the expected value we obtain,

$$
\begin{aligned}
\hat{\Gamma}(s)\approx&\tilde{\Gamma}(s)+\frac{T-|s|}{T}Var\left(\hat{\overline{u}}\right) \\
\tilde{\Gamma}(s)\approx&\hat{\Gamma}(s)-\frac{T-|s|}{T}Var\left(\hat{\overline{u}}\right) \\
E\left(\tilde{\Gamma}(s)\right)=&\left(1-\frac{|s|}{T}\right)\Gamma(s)-\frac{T-|s|}{T^2}\Omega.
\end{aligned}
$$

Therefore, the expected value of $\hat{\Omega}_T$ with kernel function $\tilde{\kappa}$ is equal to equation (2.13) from Theorem 1a plus the additional term,

$$
\begin{aligned}
-\sum_{s=-(T-1)}^{T-1}\kappa^*\left(\frac{s}{bT}\right)\frac{(T-|s|)}{T^2}\Omega &= -\Omega\sum_{s=-(T-1)}^{T-1}\left(1-\frac{|s|}{T}\right)\kappa^*\left(\frac{s}{bT}\right)\frac{1}{T}\\
&= -\Omega\sum_{s=-(T-1)}^{T-1}\left(1-\frac{|s|}{T}\right)\kappa^*\left(\frac{|s|}{bT}\right)\frac{1}{T}\\
&= -\Omega\int_{-(T-1)}^{T-1}\left(1-\frac{|s|}{T}\right)\kappa^*\left(\frac{|s|}{bT}\right)ds\\
&= -\Omega\int_{-(bT)}^{bT}\left(1-\frac{|s|}{T}\right)\kappa^*\left(\frac{|s|}{bT}\right)ds.
\end{aligned}
$$

In the last step we recognize Riemann-Stieltjes integral. To simplify we use a u-substitution with $w=\frac{s}{b}$, and recall properties of mother kernels,

$$
\begin{aligned}
-\sum_{s=-(T-1)}^{T-1}\kappa^*\left(\frac{s}{bT}\right)\frac{(T-|s|)}{T^2}\Omega &= -\Omega\int_{-T}^{T}b\kappa^*\left(\frac{w}{T}\right)dw+2\Omega\int_{0}^{T}wb^2\kappa^*\left(\frac{w}{T}\right)dw\\
&= -\Omega c_1 b+o\left(bT^{-q_0}\right).
\end{aligned}
$$

We have shown the result for a centered mother kernel. For a centered lugsail kernel we can use linearity of expectations.

■

The additional bias term in Corollary 1 is due to de-meaning the errors. It follows the results of Priestley [49, Section 5.3], and also corresponds the Edgeworth expansion result by Velasco & Robinson [63, Lemma 2 & Lemma 6].

We must issue a note of caution. We have used the terms finite sampling bias, bandwidth bias, kernel bias, de-meaning bias, and dimensionality bias; however, these as-

pects of the estimation process are also intimately related to the variability of $\hat{\Omega}_T$. For convenience we follow convention and continue referring to these components as biases.

# Chapter 3

# Asymptotic Behavior of the Test Statistics

## 3.1 Test Statistic Corrections

Most applications utilize large sample tests which rely on familiar limiting distributions for the test statistic by assuming the sample size is infinite. Large sample tests are typically valid under a wide variety of settings which results in robust procedures and streamline processes. They are particularly useful because exact distributions are not always tractable, difficult to compute, and not as versatile. However, large sample tests may not be 'good enough' to accurately capture the behavior of the test statistic in practice. For example, a standard one sample mean test with iid normal data with unknown variance can rely on the standard normal distribution or a $t$-distribution, the former being the large sample test and the latter being the exact test. The distributions are the same up to a

certain margin of error when the sample size is over 30, so for many applications there is no practical discrepancy between the two procedures. In other scenarios the sample size needs to be impractically large to observe a similar behavior, causing the observed error rate to be off by an egregious amount from the prescribed rate when using large sample distributions, and resulting in a statistically invalid test.

Rubin [51] describes two different criteria for statistical validity in the frequentest (non-Bayesian) context, *randomized validity* and *confidence validity*. Randomized validity is when the observed Type I error rate is equal to $\alpha$, and confidence validity is when the observed Type I error rate is less than or equal to $\alpha$. In either case, the rejection rate of the procedure should be no larger than $\alpha$, which is a standard requirement for Neyman-Pearson based testing procedures [50]. In cases where our test statistic is not well represented by the large sample limiting distribution there are generally two techniques used to correct the discrepancy [3]: an alternative limiting distribution that better captures the properties of finite sampling, or a new test statistic that is better approximated by the large sample limiting distribution.

A conservative version of a combination of techniques for dependent stationary univariate data has been purposed by Tukey [60] for modeling the distribution of $\hat{\Omega}_T$,

$$\nu \frac{\hat{h}(\omega)}{h(\omega)} \sim \chi^2_\nu$$

where $\nu = (bc_2)^{-1}$ is known as the equivalent degrees of freedom. The thought process, as described by Priestly [49, page 466], is if a linear process $\{Y_t\}$ has normally distributed error terms then each $I_T(\omega)$ is an independent $\chi^2$ random variable and $\hat{h}(\omega)$ is a weighted linear combination. However, these weights are typically unequal and hence Tukey's adjustment

is used to scale the test statistic and select a $\chi^2$ distribution such that the finite sample mean and variance of the test statistic match the limiting distribution.

For our purposes we focus on the distribution of the test statistic $F_T$, which has a $\chi^2$ large sample limiting distribution [1, 49]. This limiting distribution relies on the common assumption that $T \to \infty$, and an additional assumption specific to this setting that $Tb \to \infty$ at a slower rate than $T$. Because of this additional assumption the limiting distribution and related findings are referred to as small-$b$ asymptotic theory. As already discussed in Chapters 1 and 2, this limiting distribution typically does not represent the test statistic well because it does not capture the variability due to estimating the LRV and the bias the of the estimator.

Like the Tukey procedure above, we propose a combination of techniques to better capture the finite sample properties of the test statistic. The test procedure relies on an alternative nonstandard limiting distribution under the assumption that $T \to \infty$ but holds $b$ as a fixed value. This nonstandard limiting distribution is not tractable, so we further approximate the probabilities using various procedures. The resulting liming distribution and related findings are referred to as fixed-$b$ asymptotic theory [24, 25]. We denote the fixed-$b$ random variable as $F_\infty(d,b) \sim \mathscr{F}_\infty(d,b)$ and observe that

$$\begin{aligned} F_T \xrightarrow{d} & \mathscr{F}_\infty(d,b) \\ =& B_d(1)' Q_d(b)^{-1} B_d(1)/d \\ =& B_d(1)' \left[ \int_0^1 \int_0^1 \kappa\left(\frac{r-s}{bT}\right) d\tilde{B}_d(r) d\tilde{B}_d(s) \right]^{-1} B_d(1)/d \end{aligned} \tag{3.1}$$

where $Q_d(b)$ is a random variable independent of the standard $d$-dimensional Brownian process $B_d(1)$, and $\tilde{B}_d(r)$ is standard $d$-dimensional Brownian bridge process [25, 29, 54].

The fixed-$b$ distribution $\mathscr{F}_\infty(d,b)$ is dependent on the kernel function, bandwidth, and the number of restrictions under the null. This is in contrast to the standard small-$b$ limiting distribution which only takes into account the number of restrictions under the null. We will denote $c_d^\alpha(b)$ as the $\alpha$ fixed-$b$ critical value, i.e. $P\left(F_\infty \geq c_d^\alpha(b)\right) = \alpha$.

The other adjustments we make are to the test statistic through the estimation process of $\hat{\Omega}_T$. We utilize the kernels in $\mathscr{K}_3$, i.e. centered lugsail kernels. The lugsail kernel addresses the Type 1 error rate distortion due to the kernel bias, and the centering of the kernel results in higher power when the model is mis-specified. With the alternative fixed-$b$ limiting theory and kernel adjustments we have a more powerful testing procedure that achieves the prescribed Type 1 error rate at a faster rate.

## 3.2 Fixed-$b$ Limiting Distribution

This section derives some of the more technical components of the fixed-$b$ distribution under robust general assumptions. Proposition 1 derives different representations of the fixed-$b$ distribution. Lemma 2 obtains general statistics of interest for kernels in class $\mathscr{K}_3$ that do not rely on model assumptions for the data. Lemma 3 derives moments and related statistics for a random variable presented in Proposition 1b. Lastly, in Theorem 2 we present an expression for the fixed-$b$ distribution under general settings. The following results largely follow Sun [54] but in the context of lugsail estimators.

**Proposition 1:** *Let $\kappa \in \mathscr{K}_2$ and it's corresponding center kernel be $\tilde{\kappa} \in \mathscr{K}_3$.*

*(a) The following are all equal,*

$$\int_0^1 \int_0^1 \kappa\left(\frac{t-s}{b}\right) d\tilde{B}_d(t) d\tilde{B}_d(s) = \int_0^1 \int_0^1 \tilde{\kappa}\left(\frac{t-s}{b}\right) d\tilde{B}_d(t) d\tilde{B}_d(s)$$
$$= \int_0^1 \int_0^1 \tilde{\kappa}\left(\frac{t-s}{b}\right) dB_d(t) dB_d(s). \tag{3.2}$$

*(b) For fixed-b random variable $F_\infty(d,b)$ with kernel $\tilde{\kappa}$,*

$$dF_\infty(d,b) \stackrel{d}{=} \frac{||\eta||^2}{v_{11\cdot 2}} \tag{3.3}$$

*where $\eta \sim N(0, \mathbb{I}_d)$, and $v_{11\cdot 2} = v_{11} - v_{12}v_{22}^{-1}v_{21}$ is a random quantity such that*

$$\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

*is the sum of independent but not necessarily identically distributed Whisart random variables where $v_{11} \in \mathbb{R}$, $v_{21} \in \mathbb{R}^{(d-1)\times 1}$ , $v_{12} \in \mathbb{R}^{1\times(d-1)}$, and $v_{22} \in \mathbb{R}^{(d-1)\times(d-1)}$.*

*Begin proof.*

The result from Proposition 1a can be shown using properties of the centered kernels [54, page 662]. For Proposition 1b we begin by recalling some properties of Itô integrals which will be useful for proving part (a) and (b) [26]. For some function $k(t,s) : [0,1]\times[0,1] \to \mathbb{R}$ the Itô integral is,

$$\int_a^b \int_a^b k(t,s) dB_d(t) dB_d(s) = \lim_{n\to\infty} \sum_{i=1}^n \sum_{j=1}^n k(t_{i-1}, s_{j-1})$$
$$\times \left[B_d(t_i) - B_d(t_{i-1})\right] \left[B_d(s_j) - B_d(s_{j-1})\right]'. \tag{3.4}$$

If we let $t$ and $s$ have the same increments we can rewrite this again as

$$\int_a^b \int_a^b k(t,s) dB_d(t) dB_d(s) = \lim_{n\to\infty} \sum_{i=1}^n \sum_{j=1}^n k(t_{i-1}, t_{j-1})$$
$$\times \left[B_d(t_i) - B_d(t_{i-1})\right] \left[B_d(t_j) - B_d(t_{t-1})\right]'. \tag{3.5}$$

We further note a property about the expectation of the product of two Brownian motions,

$$E\left\{\left[B_d(t_i) - B_d(t_{i-1})\right]\left[B_d(t_j) - B_d(t_{j-1}))\right]\right\} = \begin{cases} 0 & \text{if } t_i \neq t_j \\ t_i - t_{i-1} & \text{if } t_i = t_j. \end{cases} \tag{3.6}$$

When $(t_i \neq t_j)$ the expect value of the product is 0 due to independent increments. When $(t_i = t_j)$ the expected value is $(t_i - t_{i-1})$ which is equal to the variance for the increment of the Brownian motion.

We now begin with Proposition 1b. Consider a series of constants $\{\lambda_n\}_{n=1}^{\infty}$ and orthonormal functions $\{\phi_n(t)\}_{n=1}^{\infty}$, i.e. $\int \phi_n(s)\phi_m(s)ds = 0$ if $m \neq n$ and 1 if otherwise, such that

$$\tilde{\kappa}_b(t,s) = \sum_{n=1}^{\infty} \lambda_n \phi_n(t)\phi_n(s). \tag{3.7}$$

We can choose any sequence of othornomal functions and corresponding constants so (3.7) holds. For example, we can consider the discrete Fourier transform for the even function $\tilde{\kappa}$ with $\phi_k(t) = \sqrt{2}\cos(2\pi kr)$ when $k = (0, 1, 2, \dots)$ [16, 52]. From the expansion in (3.7) we can approximate the equation (3.2),

$$\int_0^1 \int_0^1 \tilde{\kappa}\left(\frac{t}{b}, \frac{s}{b}\right) dB_d(t)dB_d'(s) = \int_0^1 \int_0^1 \sum_{n=1}^{\infty} \lambda_n f_n(t) f_n(s) dB_d(t) dB_d'(s) \tag{3.8}$$

$$= \sum_{n=1}^{\infty} \lambda_n \left[\int_0^1 f_n(t) dB_d(t)\right]\left[\int_0^1 f_n(s) dB_d'(s)\right] \tag{3.9}$$

$$= \sum_{n=1}^{\infty} \lambda_n \zeta_n \zeta_n' \tag{3.10}$$

where $\zeta_n \overset{iid}{\sim} N(0, I_p)$ because we used orthonormal functions, and $\zeta_n\zeta_n' \overset{iid}{\sim} Wishart_d(\mathbb{I}_d, 1)$. Using (3.10) we can now use an alternative version of (3.1) which is easier to work with,

$$dF_{\infty}(d,b) \overset{d}{=} \eta' \left[\sum_{n=1}^{\infty} \lambda_n \zeta_n \zeta_n'\right]^{-1} \eta \tag{3.11}$$

where $\eta \sim N(0, I_d)$ by properties of Brownian motions. We define a new orthornormal matrix

$$H = \begin{bmatrix} \underset{(1\times d)}{\eta'/||\eta||} \\ \underset{((d-1)\times d)}{\prod'} \end{bmatrix}$$

where $||\cdot||$ is the Euclidean norm. With this new orthornomal matrix we have the following property

$$\underset{(d\times d)}{H} \underset{(d\times 1)}{\eta} = ||\eta|| \underset{(d\times 1)}{e_1}$$

where $e_1 = (1, 0, 0, \ldots, 0)'$. Using this matrix we can form a new expression,

$$\begin{aligned}
dF_\infty(d,b) \overset{d}{=} & \eta' \left[\sum_{n=1}^{\infty} \lambda_n \zeta_n \zeta_n'\right]^{-1} \eta \\
&= \eta' \left[\sum_{n=1}^{\infty} \lambda_n (H'H)\zeta_n \zeta_n' (H'H)\right]^{-1} \eta \\
&= \eta' \left[H' \sum_{n=1}^{\infty} \lambda_n (H\zeta_n)(H\zeta_n)' H\right]^{-1} \eta \\
&= \eta'(H)^{-1} \left[\sum_{n=1}^{\infty} \lambda_n (H\zeta_n)(H\zeta_n)'\right]^{-1} (H)\eta \\
&= (H\eta)' \left[\sum_{n=1}^{\infty} \lambda_n (H\zeta_n)(H\zeta_n)'\right]^{-1} (H\eta) \\
&= ||\eta|| e_1' \left[\sum_{n=1}^{\infty} \lambda_n (H\zeta_n)(H\zeta_n)'\right]^{-1} ||\eta|| e_1 \\
&= ||\eta||^2 e_1' \left[\sum_{n=1}^{\infty} \lambda_n (H\zeta_n)(H\zeta_n)'\right]^{-1} e_1.
\end{aligned}$$

Observe that $H\zeta_n | H$ has the same distribution as $\zeta_n$ [54], therefore,

$$dF_\infty(d,b) \overset{d}{=} ||\eta||^2 e_1' \left[\sum_{n=1}^{\infty} \lambda_n \zeta_n \zeta_n'\right]^{-1} e_1.$$

To find an expression for the inverse of the random matrix we use a block matrix method [20]. Let $M$ be some $m \times m$ matrix with the following partition,

$$M = \begin{bmatrix} \underset{(1\times 1)}{c} & \underset{((m-1)\times 1)}{b} \\ \underset{(1\times(m-1))}{b'} & \underset{((m-1)\times(m-1))}{A} \end{bmatrix}.$$

Then the inverse of the matrix is

$$M^{-1} = \begin{bmatrix} \frac{1}{k} & -\frac{1}{k}b'A^{-1} \\ -\frac{1}{k}b'A^{-1} & A^{-1} + \frac{1}{k}A^{-1}bb'A^{-1} \end{bmatrix}$$

where $k = c - b'A^{-1}b$. Using this property,

$$\begin{aligned} e_1' \left( \sum_{n=1}^{\infty} \lambda_n \zeta_n \zeta_n' \right)^{-1} e_1 =& e_1' \left( \begin{bmatrix} \underset{(1\times 1)}{v_{11}} & \underset{((d-1)\times 1)}{v_{12}} \\ \underset{(1\times(d-1))}{v_{21}} & \underset{((d-1)\times(d-1))}{v_{22}} \end{bmatrix} \right)^{-1} e_1 \\ =& (v_{11} - v_{12}v_{22}^{-1}v_{21})^{-1} \\ =& v_{11.2}^{-1}. \end{aligned}$$

Thus, $dF_{\infty}(d,b) \overset{d}{=} \frac{||\eta||^2}{v_{11\cdot 2}}$ as desired.

■

In Proposition 1a we observe that when relying on $\tilde{\kappa} \in \mathscr{K}_3$ we can rewrite the fixed-$b$ distribution expression in equation (3.1) using standard Brownian motions instead of a Brownian bridge processes, which can make calculations easier. We also rely on a series representation of a kernel function, a common tool used for WOS and series estimators [27, 39, 44, 53, 54]. An important distinction between this result and Sun's [54] is that we are not restricting ourselves to an eigen value series expansion in equation (3.10). The

only kernels to guarantee equality in equation (3.10) using the eigen value version of the expansion are the positive semi-definite kernels [18] of $\mathscr{K}_1$. We instead utilize a Fourier series expansion that is applicable to class $\mathscr{K}_2$. However, the eigen value version of the expansion for a $\mathscr{K}_1$ kernel could be used to construct a basis for some $\mathscr{K}_2$ kernels but it would not be guaranteed to be orthonormal.

We note that sometimes the WOS and series estimators rely on orthonormal basis functions and sometimes replace the weights $\{\lambda_n\}$ with a constant value. For example, they may use only the first $K$ elements in the series and replace the weights with $\frac{1}{K}$. This creates an equally weighted sum of orthonormal functions, and when integrating over Brownian motions we obtain an equally weighted linear combination of independent standard Wishart random variables.

Recall the Wishart distribution is a multivariate generalization of the Gamma distribution and is commonly used to model the covariance matrix for random vectors [21]. As discussed, for univariate data sets $\hat{\Omega}_T$ can be thought of as a unequally weighted sum of independent $\chi^2$ random variables. For multivariate processes with *equal* weights, $\hat{\Omega}_T$ is the equal in distribution to a scaled finite sum of independent Wishart random variables. This makes our test statistic $F_T$ resemble a Hotelling's $T^2$ distribution, which is directly related to an standard $F$ distribution. Hotelling's $T^2$ distribution is a multivariate generalization of the $t$-distribution. Observe, if $v \sim N_d(0, \mathbb{I}_d)$ and $M \sim \text{Wishart}(\mathbb{I}_d, K)$, then

$$\begin{aligned} X :=& K v' M^{-1} v \\ \sim& T^2(d, K) \\ \to& \left(\frac{K-d+1}{dK}\right) X \sim F_{d,K-d+1}. \end{aligned}$$

In these situations one may forgo the typical $F_T$ statistic and instead capitalize on the relationship between the Wishart distribution and the $F$ distribution and use the following scaled test statistic

$$F_T^* = \left( \frac{K - d + 1}{K} \right) F_T$$

for $d > 1$. This scaled version of the test statistic is common [27, 29, 53] and similar to that of a Bonferroni-type correction [3]. However, our estimation procedure is not equally weighted thus we continue using the standard $F_T$ test statistic.

Since $\hat{\Omega}_T$ does not have a Wishart distribution we cannot take advantage of the Hotelling $T^2$ and $F$ distribution connection. Instead we utilize alternative methods. Proposition 1b obtains yet another expression for the fixed-$b$ random variable. The pivotal component in this expression is the random variable $v_{11 \cdot 2}$ which captures the variability of the estimation process for $\hat{\Omega}_T$.

Another useful observation is that the random variables $\zeta_n$ and $\eta$ are independent for all $n$ [54]. We can observe this using Itô's integral (3.4), expectation properties for Brownian motions as described in equation (3.6), and Riemann-Stieltjes integration,

$$
\begin{aligned}
Cov(\zeta_n, \eta) =& E\left[\zeta_n \eta\right] \\
=& E\left[\int_0^1 \phi_n(t) dB_d(t) \int_0^1 dB_d(s)\right] \\
=& \lim_{n\to\infty} E\left[\left(\sum_{i=1}^{n} \phi_n(t_{i-1})[B_d(t_i) - B_d(t_{i-1})]\right)\left(\sum_{j=1}^{n}[B_d(t_j) - B_d(t_{j-1})]\right)\right] \\
=& \lim_{n\to\infty}\left(\sum_{i=1}^{n} \phi_n(t_{i-1})[t_i - t_{i-1}]\right) \\
=& \int_0^1 \phi_n(t) dt \\
=& 0.
\end{aligned}
$$

This mirrors the representation in equation (3.1) derived in the original fixed-$b$ works by Kiefer and Vogelsang [24, 25] in that the random variables in the numerator and the denominator of the expressions are independent.

We next derive general center kernel properties that will be useful for modeling the behavior of $v_{11\cdot 2}$.

**Lemma 2:** *Let $\tilde{\kappa} \in \mathscr{K}_3$. As $b \to 0$ and $T \to \infty$, we have*

*(a)* $\mu_1 := \int_0^1 \tilde{\kappa}_b(t,t)dt \approx 1 - bc_1 + O(b^2)$

*(b)* $\mu_2 := \int_0^1 \int_0^1 \left[\tilde{\kappa}_b(t,s)\right]^2 dtds = \sum_{n=1}^{\infty}(\lambda_m)^2 \approx bc_2 + O(b^2)$.

*Begin proof of Lemma 2a.*

Define

$$\begin{aligned}
\mu_1 :=& \int_0^1 \tilde{\kappa}_b(t,t)dt \\
=& 1 - \int_0^1 \int_0^1 \kappa_b(t,s)dtds \\
\mu_2 :=& \int_0^1 \int_0^1 [\tilde{\kappa}_b(t,s)]^2 dtds \\
=& \left( \int_0^1 \int_0^1 \kappa_b(t-s)dtds \right)^2 + \int_0^1 \int_0^1 \kappa_b^2(t-s)dtds \\
& - 2\int_0^1 \int_0^1 \int_0^1 \kappa_b(t-p)\kappa_b(t-q)dtdqdp.
\end{aligned}$$

Recall the following identities and representations [59],

$$\sin^2(x) = \frac{1-\cos(2x)}{2} \tag{3.12}$$

$$\cos^2(x) + \sin^2(x) = 1 \tag{3.13}$$

$$\int_{-\infty}^{\infty} \left( \frac{\sin(x/c)}{x} \right)^2 dx = \frac{\pi}{|c|}, \forall c \in \mathbb{R} \tag{3.14}$$

$$\int_0^1 \cos\left(\lambda(t-s)\right) ds = \frac{1}{\lambda}\sin\left(\lambda t\right) - \frac{1}{\lambda}\sin\left(\lambda(t-1)\right) \tag{3.15}$$

$$\exp(it) = \cos(x) + i\sin(x). \tag{3.16}$$

Define the following Fourier transformations and the corresponding inverse transformations,

$$\begin{aligned}
K_1(\lambda) =& \frac{1}{2\pi}\int_{-\infty}^{\infty} \kappa\left(\frac{x}{b}\right) \exp\left(\frac{-i\lambda x}{b}\right) dx \qquad \kappa\left(\frac{x}{b}\right) = \int_{-\infty}^{\infty} K_1(\lambda) \exp\left(\frac{i\lambda x}{b}\right) d\lambda \\
K_2(\lambda) =& \frac{1}{2\pi}\int_{-\infty}^{\infty} \kappa^2\left(\frac{x}{b}\right) \exp\left(\frac{-i\lambda x}{b}\right) dx \qquad \kappa^2\left(\frac{x}{b}\right) = \int_{-\infty}^{\infty} K_2(\lambda) \left(\frac{i\lambda x}{b}\right) d\lambda.
\end{aligned}$$

Using the inverse Fourier transformations we derive an expression for $\int_0^1 \int_0^1 \kappa_b(t-s)dtds$,

$$\begin{aligned}
\int_0^1 \int_0^1 \kappa_b(t-s)dtds &= \int_0^1 \int_0^1 \int_{-\infty}^{\infty} K_1(\lambda) \exp\left(-i\lambda\left(\frac{t}{b}-\frac{s}{b}\right)\right) dtds \\
&= \int_0^1 \int_0^1 \int_{-\infty}^{\infty} K_1(\lambda) \exp\left(-i\lambda\frac{t}{b}\right) \exp\left(i\lambda\frac{s}{b}\right) dtdsd\lambda \\
&= \int_{-\infty}^{\infty} K_1(\lambda) \left[\int_0^1 \exp\left(-i\lambda\frac{t}{b}\right) dt\right] \left[\int_0^1 \exp\left(i\lambda\frac{s}{b}\right) ds\right] d\lambda.
\end{aligned}$$

We integrate the inside terms, use Euler's formula in equation (3.16), and properties of imaginary numbers,

$$\begin{aligned}
\int_0^1 \int_0^1 \kappa_b(t-s)dtds &= \int_{-\infty}^{\infty} K_1(\lambda) \left[\frac{-i\left(\exp(i\lambda/b)-1\right)}{(\lambda/b)}\right] \left[\frac{\sin(\lambda/b)+i\left(\cos(\lambda/b)-1\right)}{(\lambda/b)}\right] d\lambda \\
&= \int_{-\infty}^{\infty} K_1(\lambda) \left(\frac{b}{\lambda}\right)^2 \left[-i\left(\cos(\lambda/b)+i\sin(\lambda/b)-1\right)\right] \\
&\quad \times \left[\sin(\lambda/b)+i\left(\cos(\lambda/b)-1\right)\right] d\lambda \\
&= \int_{-\infty}^{\infty} K_1(\lambda) \left(\frac{b}{\lambda}\right)^2 \left[-i\left(\cos(\lambda/b)-1\right)+\sin(\lambda/b)\right] \\
&\quad \times \left[\sin(\lambda/b)+i\left(\cos(\lambda/b)-1\right)\right] d\lambda \\
&= \int_{-\infty}^{\infty} K_1(\lambda) \left(\frac{b}{\lambda}\right)^2 \left[\left\{i\left(\cos(\lambda/b)-1\right)\right\}^2+\sin^2(\lambda/b)\right] d\lambda \\
&= \int_{-\infty}^{\infty} K_1(\lambda) \left(\frac{b}{\lambda}\right)^2 \left[\left(1-\cos(\lambda/b)\right)^2+\sin^2(\lambda/b)\right] d\lambda \\
&= \int_{-\infty}^{\infty} K_1(\lambda) \left(\frac{b}{\lambda}\right)^2 \left[\left(\cos^2(\lambda/b)-2\cos(\lambda/b)+1\right)+\sin^2(\lambda/b)\right] d\lambda.
\end{aligned}$$

We further use the identities in equations (3.13), (3.12), and (3.14) to continue to simplify,

$$\begin{aligned}
\int_0^1 \int_0^1 \kappa_b(t-s)dtds &= \int_{-\infty}^{\infty} K_1(\lambda) \left(\frac{b}{\lambda}\right)^2 [2 - 2\cos(\lambda/b)]\, d\lambda \\
&= \int_{-\infty}^{\infty} 2K_1(\lambda) \left(\frac{b}{\lambda}\right)^2 [1 - 1\cos(\lambda/b)]\, d\lambda \\
&= \int_{-\infty}^{\infty} 4K_1(\lambda) \left(\frac{b}{\lambda}\right)^2 \sin^2\left(\frac{\lambda}{2b}\right) d\lambda \\
&= \int_{-\infty}^{\infty} 4b^2 K_1(\lambda) \left(\frac{\sin\left(\frac{\lambda}{2b}\right)}{\lambda}\right)^2 d\lambda \\
&= \int_{-\infty}^{\infty} [K_1(\lambda) - K_1(0) + K_1(0)]\, 4b^2 \left(\frac{\sin\left(\frac{\lambda}{2b}\right)}{\lambda}\right)^2 d\lambda \\
&= \int_{-\infty}^{\infty} K_1(0) 4b^2 \left(\frac{\sin\left(\frac{\lambda}{2b}\right)}{\lambda}\right)^2 d\lambda \\
&\quad + \int_{-\infty}^{\infty} [K_1(\lambda) - K_1(0)]\, 4b^2 \left(\frac{\sin\left(\frac{\lambda}{2b}\right)}{\lambda}\right)^2 d\lambda \\
&= K_1(0) 4b^2 \frac{\pi}{2b} + \int_{-\infty}^{\infty} [K_1(\lambda) - K_1(0)]\, 4b^2 \left(\frac{\sin\left(\frac{\lambda}{2b}\right)}{\lambda}\right)^2 d\lambda \\
&= K_1(0) 2b\pi + \int_{-\infty}^{\infty} [K_1(\lambda) - K_1(0)]\, 4b^2 \left(\frac{\sin\left(\frac{\lambda}{2b}\right)}{\lambda}\right)^2 d\lambda. \qquad (3.17)
\end{aligned}$$

Now we concentrate on the the last component of (3.17) and drop the constants. Recall the identity in equation (3.12) and rearrange terms,

$$\begin{aligned}
&\textstyle\int_{-\infty}^{\infty} \left[\frac{K_1(\lambda) - K_1(0)}{\lambda^2}\right] \left(\sin^2\left(\frac{\lambda}{2b}\right)\right) d\lambda \\
&= \int_{-\infty}^{\infty} \left[\frac{K_1(\lambda) - K_1(0)}{\lambda^2}\right] \left(\sin^2\left(\frac{\lambda}{2b}\right) - \frac{1}{2} + \frac{1}{2}\right) d\lambda \\
&= \int_{-\infty}^{\infty} \left[\frac{K_1(\lambda) - K_1(0)}{\lambda^2}\right] \left(\sin^2\left(\frac{\lambda}{2b}\right) - \frac{1}{2}\right) d\lambda + \int_{-\infty}^{\infty} \left[\frac{K_1(\lambda) - K_1(0)}{\lambda^2}\right] \left(\frac{1}{2}\right) d\lambda \\
&= \int_{-\infty}^{\infty} \left[\frac{K_1(\lambda) - K_1(0)}{\lambda^2}\right] \left(\cos\left(\frac{\lambda}{b}\right)\right) d\lambda + \left(\frac{1}{2}\right) \int_{-\infty}^{\infty} \left[\frac{K_1(\lambda) - K_1(0)}{\lambda^2}\right] d\lambda. \qquad (3.18)
\end{aligned}$$

Recall the Riemann-Lebesgue Lemma [8], if the Lebesgue integral of $|f|$ is finite, then the Fourier transform of $f$ satisfies

$$\int f(x)exp(-izx)dx \to 0$$

as $|z| \to \infty$. Observe $K_1(0) = \frac{c_1}{2\pi}$ is a constant, $K_1(\lambda)$ is finite and even because the Fourier transform of an even function is even, and thus consequentially $\frac{K_1(\lambda)-K_1(0)}{\lambda^2}$ is a finite even function. We further observe that

$$\int_{-\infty}^{\infty} \left[\frac{K_1(\lambda) - K_1(0)}{\lambda^2}\right] \left(\cos\left(\frac{\lambda}{b}\right)\right) d\lambda$$

is the cosine representation of Fourier transformations for even functions. Thus, we expect it to tend towards 0 by Riemann-Lebesgue Lemma as $\frac{1}{b} \to \infty$ (or $b \to 0$). Therefore we have the following,

$$\int_0^1 \int_0^1 \kappa_b(t-s)dtds = K_1(0)2b\pi + \left(\frac{4b^2}{2}\right) \int_{-\infty}^{\infty} \left[\frac{K_1(\lambda) - K_1(0)}{\lambda^2}\right] d\lambda + o(b^2). \tag{3.19}$$

Again by recognizing properties of Fourier transformations of even functions we can substitute: $K_1(\lambda) = (2\pi)^{-1} \int_{-\infty}^{\infty} \kappa\left(\frac{x}{b}\right) \cos\left(\lambda\frac{x}{b}\right) dx$,

$$\int_0^1 \int_0^1 \kappa_b(t-s)dtds = K_1(0)2b\pi + \left(\frac{2b^2}{\pi}\right) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \kappa\left(\frac{x}{b}\right) \left(\frac{cos\left(\frac{\lambda x}{b}\right) - c_1}{\lambda^2}\right) dxd\lambda + o(b^2). \tag{3.20}$$

We now consider three cases for $c_1$: $c_1 = 1$, $c_1 < 1$, and $c_1 > 1$. We first consider the case when $c_1 = 1$. Using these identities in equations (3.12) and (3.14), and properties of mother kernels we obtain

$$\begin{aligned}
\int_0^1 \int_0^1 \kappa_b(t-s)dtds =& K_1(0)2b\pi + \left(\frac{2b^2}{\pi}\right) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \kappa\left(\frac{x}{b}\right) \frac{sin^2\left(\frac{\lambda}{2}\frac{x}{b}\right)}{\lambda^2} dxd\lambda + o(b^2) \\
=& K_1(0)2b\pi + \left(\frac{2b^2}{\pi}\right) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \kappa(y) \frac{sin^2\left(\frac{\lambda}{2}y\right)}{\lambda^2} dyd\lambda + o(b^2) \\
=& K_1(0)2b\pi + \left(\frac{2b^2}{\pi}\right) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \kappa(y) \frac{sin^2(\lambda y/2)}{\lambda^2} dyd\lambda + o(b^2) \\
=& K_1(0)2b\pi + b^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \kappa(y)|y| dyd\lambda + o(b^2) \\
\approx& K_1(0)2b\pi + O(b^2) + o(b^2) \\
\approx& bc_1 + O(b^2).
\end{aligned}$$

Next we consider the case when $c_1 < 1$. Observe that $\forall y$ and $\forall \lambda$

$$\begin{aligned}
-\kappa(y)\left(\frac{cos(\lambda y) - 1}{\lambda^2}\right) < \kappa(y)\left(\frac{cos(\lambda y) - c_1}{\lambda^2}\right) < \kappa(y)\left(\frac{cos(\lambda y) - 1}{\lambda^2}\right) \\
-\kappa(y)\left(\frac{sin^2(\lambda y/2)}{\lambda^2}\right) < \kappa(y)\left(\frac{cos(\lambda y) - c_1}{\lambda^2}\right) < \kappa(y)\left(\frac{sin^2(\lambda y/2)}{\lambda^2}\right) \\
-\int_{-\infty}^{\infty} k(y)|y|dy < \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \kappa(y)\left(\frac{cos(\lambda y) - c_1}{\lambda^2}\right) dyd\lambda < \int_{-\infty}^{\infty} k(y)|y|dy.
\end{aligned}$$

Define $s := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \kappa(y)\left(\frac{cos(\lambda y) - c_1}{\lambda^2}\right) dyd\lambda < \infty$ and hence

$$\begin{aligned}
\int_0^1 \int_0^1 \kappa(t-s)dtds \approx& bc_1 - b^2 s + o(b^2) \\
\approx& bc_1 + O(b^2).
\end{aligned}$$

Lastly for the case when $c_1 > 1$. We can rewrite $c_1$ into components $c_1 = c_{1.1} + c_{1.r}$, where $c_{1.1} = 1$ and $c_{1.r} = c_1 - 1$. Then,

$\left(\frac{2b^2}{\pi}\right)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\kappa(y)\left(\frac{cos(\lambda y)-c_1}{\lambda^2}\right)dyd\lambda + o(b^2)$

$$
\begin{aligned}
&= \left(\frac{2b^2}{\pi}\right)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\kappa(y)\left(\frac{cos(\lambda y)-c_{1.1}-c_{1.r}}{\lambda^2}\right)dyd\lambda + o(b^2)\\
&= \left(\frac{2b^2}{\pi}\right)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\kappa(y)\left(\frac{cos(\lambda y)-c_{1.1}}{\lambda^2}\right)dyd\lambda\\
&\quad - \left(\frac{2b^2}{\pi}\right)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\kappa(y)\left(\frac{c_{1.r}}{\lambda^2}\right)dyd\lambda + o(b^2)\\
&\approx O(b^2).
\end{aligned}
$$

Thus $\mu_1 = \int_0^1 \tilde{\kappa}_b(t,t)dt \approx 1 - bc_1 + O(b^2)$ as desired.

■

*Begin proof of Lemma 2b.*

First we focus on the term $\int_0^1\int_0^1 \kappa_b^2(t-s)dtds$. Using the same arguments as Lemma 2a, we can replace all instances of $K_1$ with $K_2$. Recalling the mother kernel property: $\int_{-\infty}^{\infty}\kappa^2(x)x^2dx < \infty$, we get the following,

$$\int_0^1\int_0^1 \kappa_b(t-s)dtds \approx bc_2 + O(b^2). \tag{3.21}$$

We next consider: $\int_0^1\int_0^1\int_0^1 \kappa_b(t-p)\kappa_b(t-q)dtdqdp$. We start with an alternative representation of the familiar integral using the Fourier transforms and inverses from Lemma 2a with kernel function $\kappa$, and use properties of Fourier transforms and even functions

$$
\begin{aligned}
\int_0^1 \kappa_b(t-s)ds &= \int_0^1\left(\int_{-\infty}^{\infty} K_1(\lambda)\exp\left(\frac{-i\lambda(t-s)}{b}\right)d\lambda\right)ds\\
&= \int_0^1\left(\int_{-\infty}^{\infty} K_1(\lambda)\cos\left(\frac{\lambda(t-s)}{b}\right)d\lambda\right)ds\\
&= \int_{-\infty}^{\infty}\int_0^1 K_1(\lambda)cos\left(\left(\frac{\lambda(t-s)}{b}\right)dx\right)d\lambda.
\end{aligned}
$$

We use the identity in equation (3.15) and rewrite the integral using a u-substitution with $x = \lambda/b$,

$$\begin{aligned}\int_0^1 \kappa_b(t-s)ds &= -\int_{-\infty}^{\infty} K_1(\lambda)\left(\frac{b}{\lambda}\right)\left[\sin\left(\frac{\lambda(t-1)}{b}\right)-\sin\left(\frac{\lambda t}{b}\right)\right]d\lambda \\ &= -b^2\int_{-\infty}^{\infty} K_1(xb)\left(\frac{1}{x}\right)\left[\sin\left(x(t-1)\right)\sin\left(xt\right)\right]dx. \qquad (3.22)\end{aligned}$$

We rewrite the triple integral with the expression in equation (3.22),

$\int_0^1\int_0^1\int_0^1 \kappa_b(t-p)\kappa_b(t-q)dtdqdp$

$$\begin{aligned}&= \int_0^1\left[\int_0^1 \kappa_b(t-p)dp\right]\left[\int_0^1 \kappa_b(t-q)dq\right]dt \\ &= \int_0^1\left[\int_0^1 \kappa_b(t-p)dp\right]^2 dt \\ &= b^2\int_0^1\left[\int_{-\infty}^{\infty} K_1(xb)\left(\frac{1}{x}\right)\left[\sin\left(x(t-1)\right)-\sin\left(xt\right)\right]dx\right]^2 dt \\ &= b^2\int_0^1\left[\int_{-\infty}^{\infty} K_1(xb)\left(\frac{1}{x}\right)\sin\left(x(t-1)\right)dx - \int_{-\infty}^{\infty} K_1(xb)\left(\frac{1}{x}\right)\sin\left(xt\right)dx\right]^2 dt.\end{aligned}$$

For each $x$, $K_1(xb) \approx K_1(0) + o(1) = K_1(0)(1+o(1))$ as $b \to 0$,

$\int_0^1\int_0^1\int_0^1 \kappa_b(t-p)\kappa_b(t-q)dtdqdp$

$$\approx b^2K_1^2(0)\int_0^1\left[\int_{-\infty}^{\infty}\frac{1}{x}\sin\left(x(t-1)\right)dx - \int_{-\infty}^{\infty}\frac{1}{x}\sin\left(xt\right)dx\right]^2 dt(1+o(1)).$$

We further use a u-substitution, the property that $K_1(0) = \frac{c_1}{2\pi}$, and the identity (3.14) to simplify terms,

$\int_0^1 \int_0^1 \int_0^1 \kappa_b(t-p)\kappa_b(t-q)dtdqdp$

$$\begin{aligned}
&\approx b^2 K_1^2(0) \int_0^1 \left[ 2 \int_{-\infty}^{\infty} \frac{1}{y} sin\,(y)\, dy \right]^2 dt(1+o(1)) \\
&= b^2 c_1^2 (2\pi)^{-2} \int_0^1 \left[ 2 \int_{-\infty}^{\infty} \frac{1}{y} sin\,(y)\, dy \right]^2 dt(1+o(1)) \\
&= b^2 c_1^2 (2\pi)^{-2} \int_0^1 2^2 \pi^2 dt(1+o(1)) \\
&= b^2 c_1^2 + o(b^2).
\end{aligned}$$

Therefore, we have the following

$$\begin{aligned}
\mu_2 \approx & [bc_1 + O(b^2)]^2 + [bc_2 + O(b^2)] - 2[b^2 c_1^2 + O(b^2)] \\
= & [bc_1]^2 + bc_2 - 2[bc_1]^2 + O(b^2) \\
= & bc_2 + O(b^2).
\end{aligned}$$

■

In Lemma 2 no model or distributional properties of the data are used except for the implicit assumption that $T \to \infty$. The terms $\mu_1$ and $\mu_2$ can be thought of as adjacent summary statistics to those presented in Table 2.2 for $\tilde{\kappa} \in \mathscr{K}_3$ when accounting for the bandwidth parameter. We will see in Lemma 3 that these summary statistics for $\mathscr{K}_3$ are related to the first two moments of the random variable $v_{11\cdot 2}$ in a similar way as the summary statistics in Table 2.2 for the kernels in class $\mathscr{K}_2$ are related to the bias and variance of $\hat{\Omega}_T$ as presented in Theorem 1.

When using $\tilde{\kappa} \in \mathscr{K}_3$ constructed with $\kappa^* \in \mathscr{K}_1$, i.e. a centered mother kernel, we have the additional observation that $\mu_1$ is equal to the sum of the eigen values because of positive semi-definite properties. In contrast, $\tilde{\kappa} \in \mathscr{K}_3$ constructed with $\kappa \in \mathscr{K}_2$, i.e. all

centered lugsail kernels, $\mu_2$ is equal to the sum of squared eigen values despite the lack of guaranteed positive semi-definiteness.

With the information in Lemma 2 we can derive the first two moments of $v_{11\cdot 2}$.

**Lemma 3:** *Let $\tilde{\kappa}_b \in \mathscr{K}_3$. As $b \to 0$ and $T \to \infty$, we have*

*(a)* $E(v_{11} - v_{12}v_{22}^{-1}v_{21}) = 1 - bc_1 - bc_2(d-1) + o(b)$

*(b)* $E\left[(v_{11} - v_{12}v_{22}^{-1}v_{21})^2\right] = 1 - 2b(c_1 - c_2) - 2(d-1)bc_2 + o(b)$

*(c)* $E\left[(v_{11} - v_{12}v_{22}^{-1}v_{21}) - 1\right]^2 = 2bc_2 + o(b)$

*Begin proof of Lemma 3a.*

We start with expectation of $v_{11}$, and utilize Itô's integral (3.4). Recall that $(t, s)$ have the same support $[0, 1]$ and we can choose the same partition for both dimensions,

$$\begin{aligned}
E\{v_{11}\} =& E\left\{\int_0^1 \int_0^1 \tilde{\kappa}_b(t,s) dB_1(t) dB_1'(s)\right\} \\
=& \lim_{n\to\infty} E\left\{\sum_{i=1}^n \sum_{j=1}^n \tilde{\kappa}_b(t_{i-1}, s_{j-1}) \left[B_1(t_i) - B_1(t_{i-1})\right] \left[B_1(s_j) - B_1(s_{j-1}))\right]\right\} \\
=& \lim_{n\to\infty} E\left\{\sum_{i=1}^n \sum_{j=1}^n \tilde{\kappa}_b(t_{i-1}, t_{j-1}) \left[B_1(t_i) - B_1(t_{i-1})\right] \left[B_1(t_j) - B_1(t_{j-1}))\right]\right\} \\
=& \lim_{n\to\infty} \sum_{i=1}^n \sum_{j=1}^n \tilde{\kappa}_b(t_{i-1}, t_{j-1}) E\left\{\left[B_1(t_i) - B_1(t_{i-1})\right] \left[B_1(t_j) - B_1(t_{j-1}))\right]\right\}.
\end{aligned}$$

We further recall that that non-overlapping increments are independent as illustrated in equation (3.6). We use this property along with Riemann-Stieltjes integration, and Lemma 2 to obtain

$$\begin{aligned} E\left\{v_{11}\right\} =& \sum_{i=1}^{n} \tilde{\kappa}_b(t_{i-1}, t_{i-1}) E\left\{\left[B_1(t_i) - B_1(t_{i-1})\right]^2\right\} \\ =& \sum_{i=1}^{n} \tilde{\kappa}_b(t_{i-1}, t_{i-1}) \left[t_i - t_{i-1}\right] \\ =& \int_0^1 \tilde{\kappa}_b(t,t) dt \\ =& \mu_1 \\ =& 1 - bc_1 + O(b^2). \end{aligned}$$

Similarly for $v_{22}$,

$$E\left\{v_{22}\right\} = (1 - bc_1)\mathbf{1}_{(d-1)} + O(b^2). \tag{3.23}$$

Using the representation in equation (3.7) we observe the following properties.

$$\begin{aligned} \int_0^1 \tilde{\kappa}_b(t,s)\phi_n(t)dt =& \lambda_n \phi_n(s) \\ \int_0^1 \phi_n(t)dt =& 0. \end{aligned}$$

Let $\zeta_n := \int_0^1 \phi_n(t) dB_d(t) \in \mathbb{R}^d$, then

$$\begin{aligned} \int_0^1 \int_0^1 \tilde{\kappa}_b(t,s) dB_p(t) dB_p'(s) =& \int_0^1 \int_0^1 \sum_{n=1}^{\infty} \lambda_n \phi_n(t)\phi_n(s) dB_d(t) dB_d(s) \\ =& \sum_{n=1}^{\infty} \lambda_n \left(\int_0^1 \phi_n(t) dB_d(t)\right)\left(\int_0^1 \phi_n(s) dB_d(s)\right) \\ =& \sum_{n=1}^{\infty} \lambda_n \zeta_n \zeta_n' \end{aligned} \tag{3.24}$$

where $\zeta_n \overset{iid}{\sim} N(0,1)$ and $\zeta_n\zeta_n' \overset{iid}{\sim} Wishart(\mathbb{I}_d, 1)$. Observe that

$$\sum_{n=1}^{\infty} \lambda_n \zeta_n \zeta_n' \xrightarrow{p} \sum_{n=1}^{\infty} \lambda_n \mathbf{1}_d \tag{3.25}$$

where $\sum_{n=1}^{\infty} \lambda_n \in \mathbb{R}^1$. We further observe that $\xi_n := \int_0^1 \phi_n(t) dB_{d-1}(t) \in \mathbb{R}^{d-1}$, which inherits the same properties. Next we consider $v_{12} v_{22}^{-1} v_{21}$. Observe that $v_{12} \perp v_{22} \perp v_{21}$. We utilize the property in equation (3.24) for $\xi_n$, the results in Lemma 2, and convergence properties of random variables and continuous functions [50, pg 290] to obtain

$$
\begin{aligned}
E\left\{v_{12}v_{22}^{-1}v_{21}\right\} =& E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_1(t)dB'_{d-1}(s)\right)\right.\\
&\times\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_{d-1}(t)dB'_{d-1}(s)\right)^{-1}\\
&\left.\times\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB'_1(t)dB_{d-1}(s)\right)\right\}\\
=& E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_1(t)dB'_{d-1}(s)\right)\right\}\\
&\times E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_{d-1}(t)dB'_{d-1}(s)\right)^{-1}\right\}\\
&\times E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB'_1(t)dB_{d-1}(s)\right)\right\}\\
=& E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_1(t)dB'_{d-1}(s)\right)\right\}\\
&\times E\left\{\left(\sum_{n=1}^{\infty}\lambda_n\xi_n\xi'_n\right)^{-1}\right\}\\
&\times E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB'_1(t)dB_{d-1}(s)\right)\right\}\\
=& E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_1(t)dB'_{d-1}(s)\right)\right\}\\
&\times \mu_1^{-1}\mathbf{1}_{d-1}\\
&\times E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB'_1(t)dB_{d-1}(s)\right)\right\}\\
=& \mu_1^{-1}E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_1(t)dB'_{d-1}(s)\right)\right\}\\
&\times E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB'_1(t)dB_{d-1}(s)\right)\right\}\\
=& \mu_1^{-1}E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_1(t)dB'_{d-1}(s)\right)\right.\\
&\left.\times\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB'_1(t)dB_{d-1}(s)\right)\right\}.
\end{aligned}
$$

Replace $s$ with $\tau_1$ and $\tau_2$, and use the definition of Itô's integral in equation (3.4) for the inner integrals of the last two terms and choose the same partition. We further use the property that non-overlapping increments are independent as illustrated in equation (3.6), and Riemann-Stieltjes integration, to simplify

$$\begin{aligned}
E\left\{v_{12}v_{22}^{-1}v_{21}\right\} =& \mu_1^{-1}E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,\tau_1)dB_1(t)dB'_{d-1}(\tau_1)\right)\right. \\
&\left.\times\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,\tau_2)dB'_1(t)dB_{d-1}(\tau_2)\right)\right\} \\
=& \mu_1^{-1}\lim_{n\to\infty}E\left\{\left(\int_0^1\sum_{i=1}^n \tilde{\kappa}_b(t,\tau_1)\left[B_1(t_i)-B_1(t_{i-1})\right]dB'_{d-1}(\tau_1)\right)\right. \\
&\left.\times\left(\int_0^1\sum_{i=1}^n \tilde{\kappa}_b(t,\tau_2)\left[B_1(t_i)-B_1(t_{i-1})\right]dB_{d-1}(\tau_2)\right)\right\} \\
=& \mu_1^{-1}\lim_{n\to\infty}E\Big\{\int_0^1\int_0^1\sum_{i=1}^n\sum_{j=1}^n \tilde{\kappa}_b(t_{i-1},\tau_1)\tilde{\kappa}_b(t_{j-1},\tau_2) \\
&\times\left[B_1(t_i)-B_1(t_{i-1})\right]\left[B_1(t_j)-B_1(t_{j-1})\right]dB'_{d-1}(\tau_1)dB_{d-1}(\tau_2)\Big\} \\
=& \mu_1^{-1}\lim_{n\to\infty}E\Big\{\int_0^1\int_0^1\sum_{i=1}^n \tilde{\kappa}_b(t_{i-1},\tau_1)\tilde{\kappa}_b(t_{i-1},\tau_2) \\
&\times\left[t_i-t_{i-1}\right]dB'_{d-1}(\tau_1)dB_{d-1}(\tau_2)\Big\} \\
=& \mu_1^{-1}E\left\{\int_0^1\int_0^1\left(\int_0^1 \tilde{\kappa}_b(t,\tau_1)\tilde{\kappa}_b(t,\tau_2)dt\right)dB'_{d-1}(\tau_1)dB_{d-1}(\tau_2)\right\}
\end{aligned}$$

Observe the non-random middle component above. We use substitution with the representation in equation (3.7) to obtain

$$\begin{aligned}\int_0^1 \tilde{\kappa}_b(t,\tau_1)\tilde{\kappa}_b(t,\tau_2)dt &= \int_0^1 \left(\sum_{m=1}^{\infty} \lambda_m \phi_m(t)\phi_m(\tau_1)\right)\left(\sum_{n=1}^{\infty} \lambda_n \phi_n(t)\phi_n(\tau_2)\right) dt \\ &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \lambda_m\lambda_n \left(\int_0^1 \phi_m(t)\phi_n(t)dt\right)\phi_m(\tau_1)\phi_n(\tau_2) \\ &= \sum_{n=1}^{\infty} \lambda_n^2 \phi_n(\tau_1)\phi_n(\tau_2).\end{aligned}$$

We use the property in equation (3.24) for $\xi_n$ to further simplify,

$$\begin{aligned}&\left(\int_0^1\int_0^1 \left[\int_0^1 \tilde{\kappa}_b(t,\tau_1)\tilde{\kappa}_b(t,\tau_2)dt\right] dB'_{d-1}(\tau_1)dB_{d-1}(\tau_2)\right) \\ &\quad= \left(\int_0^1\int_0^1 \left[\sum_{n=1}^{\infty}(\lambda_n)^2\phi_n(\tau_1)\phi_n(\tau_2)\right] dB'_{d-1}(\tau_1)dB_{d-1}(\tau_2)\right) \\ &\quad= \sum_{n=1}^{\infty}\lambda_n^2\xi'_n\xi_n.\end{aligned}$$

We recognize that $\xi'_i\xi_i \overset{iid}{\sim} \chi^2_{(d-1)}$, which means that $\sum_{n=1}^{\infty}\lambda_n^2\xi'_n\xi_n \xrightarrow{p} \sum_{n=1}^{\infty}\lambda_n^2(d-1)$ where $\sum_{n=1}^{\infty}\lambda_n^2 = \mu_2 \in \mathbb{R}^1$. Therefore,

$$\begin{aligned}E\left\{v_{12}v_{22}^{-1}v_{21}\right\} =& \mu_1^{-1}E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,\tau_1)dB_1(t)dB'_{d-1}(\tau_1)\right)\right. \\ &\left.\times\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t,\tau_2)dB'_1(t)dB_{d-1}(\tau_2)\right)\right\} \\ =& \mu_1^{-1}E\left\{\sum_{n=1}^{\infty}\lambda_n^2\xi'_n\xi_n\right\} \\ =& \frac{\mu_2}{\mu_1}(d-1)(1+o(1)) \\ =& \frac{bc_2+O(b^2)}{1-bc_1+O(b^2)}(d-1)(1+o(1)) \\ =& bc_2(d-1)+o(b).\end{aligned}$$

■

*Begin proof of Lemma 3b.*

Next we consider $v_{11}^2$. We use Itô's integral in equation (3.4),

$$\begin{aligned}
E[v_{11}^2] =& E\left\{\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t_1,s_1)dB_1(t_1)dB_1'(s_1)\right)\right. \\
&\left.\left(\int_0^1\int_0^1 \tilde{\kappa}_b(t_2,s_2)dB_1(t_2)dB_1'(s_2)\right)\right\} \qquad (3.26)\\
=& \lim_{n\to\infty} E\left(\sum_{i=1}^n\sum_{j=1}^n \tilde{\kappa}_b(t_{i-1},s_{j-1})\left[B_1(t_i)-B_1(t_{i-1})\right]\left[B_1(s_j)-B_1(s_{j-1})\right]\right) \\
&\times\left(\sum_{l=1}^n\sum_{h=1}^n \tilde{\kappa}_b(t_{l-1},s_{h-1})\left[B_1(t_l)-B_1(t_{l-1})\right]\left[B_1(s_h)-B_1(s_{h-1})\right]\right) \qquad (3.27)\\
=& \lim_{n\to\infty}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n\sum_{h=1}^n \tilde{\kappa}_b(t_{i-1},s_{j-1})\tilde{\kappa}_b(t_{l-1},s_{h-1})E\left\{\left[B_1(t_i)-B_1(t_{i-1})\right]\right. \\
&\left.\times\left[B_1(s_j)-B_1(s_{j-1})\right]\left[B_1(t_l)-B_1(t_{l-1})\right]\left[B_1(s_h)-B_1(s_{h-1})\right]\right\}. \qquad (3.28)
\end{aligned}$$

There are three cases to consider: $(i=j,\, l=h)$, $(i=l,\, j=h)$, $(i=h,\, j=l)$. We review the first case using the format in equation (3.27). Using expecation properties for Brownian motions illustrated in equation (3.6) coupled with Riemann-Stieltjes integration we obtain

$$\begin{aligned}
&E\left(\textstyle\sum_{i=1}^n \tilde{\kappa}_b(t_{i-1},t_{i-1})\left[B_1(t_i)-B_1(t_{i-1})\right]^2\right)\left(\textstyle\sum_{l=1}^n \tilde{\kappa}_b(t_{l-1},t_{l-1})\left[B_1(t_l)-B_1(t_{l-1})\right]^2\right) \\
&\qquad= \left(\sum_{i=1}^n \tilde{\kappa}_b(t_{i-1},t_{i-1})\left[t_i-t_{i-1}\right]\right)\left(\sum_{l=1}^n \tilde{\kappa}_b(t_{l-1},t_{l-1})\left[t_l-t_{l-1}\right]\right) \\
&\qquad= \left(\int_0^1 \tilde{\kappa}_b(t,t)dt\right)^2.
\end{aligned}$$

We next review the second case, i.e. $(i=l,\, j=h)$, using the format of equation (3.28). Again by using expectation properties for Brownian motions illustrated in equation (3.6) coupled with Riemann-Stieltjes integration we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{\kappa}_b(t_{i-1},s_{j-1})\tilde{\kappa}_b(t_{i-1},s_{j-1})E\left\{\left[B_1(t_i)-B_1(t_{i-1})\right]^2\left[B_1(s_j)-B_1(s_{j-1})\right]^2\right\}$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{\kappa}_b(t_{i-1},s_{j-1})\tilde{\kappa}_b(t_{i-1},s_{j-1})\left[t_i-t_{i-1}\right]\left[s_j-s_{j-1}\right]$$

$$=\int_0^1\int_0^1\tilde{\kappa}_b(t,s)^2dtds.$$

By symmetry, the second and third cases return the same result. Therefore, by using Lemma 1 we can rewrite the expected value of equation $v_{11}^2$ as

$$\begin{aligned}E(v_{11}^2)=&E\left\{\left(\int_0^1\int_0^1\tilde{\kappa}_b(t_1,s_1)dB_1(t_1)dB_1'(s_1)\right)\left(\int_0^1\int_0^1\tilde{\kappa}_b(t_2,s_2)dB_1(t_2)dB_1'(s_2)\right)\right\}\\ =&\left(\int_0^1\tilde{\kappa}_b(t,r)dt\right)^2+2\int_0^1\int_0^1\tilde{\kappa}_b(t,s)^2dtds\\ =&\mu_1^2+2\mu_2\\ =&(1-bc_1+o(b))^2+2bc_2+o(b)\\ =&1-2b(c_1-c_2)+o(b).\end{aligned}$$

Next we consider $v_{11}v_{12}v_{22}^{-1}v_{21}$. Recall $v_{11}\perp v_{12}\perp v_{22}\perp v_{21}$, which implies that $v_{11}\perp v_{12}v_{22}^{-1}v_{21}$. Using this independence relationship we obtain

$$
\begin{aligned}
E[v_{11}v_{12}v_{22}^{-1}v_{21}] =& E\left\{\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_1(t)dB_1'(s)\right. \\
& \int_0^1\int_0^1 \tilde{\kappa}_b(t,\tau_1)dB_1(t)dB_{d-1}'(\tau_1) \\
& \left[\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_1(t)dB_{d-1}'(\tau_1)\right]^{-1} \\
& \left.\int_0^1\int_0^1 \tilde{\kappa}_b(t,\tau_2)dB_1(t)dB_{d-1}'(s)\right\} \\
=& E\{v_{11}\}E\{v_{12}v_{22}^{-1}v_{21}\} \\
\approx& \{(1-bc_1+o(b)\}\{bc_2(p-1)+o(b)\} \\
=& bc_2(p-1)+o(b)-b^2c_1c_2(p-1)+o(b)bc_1 \\
=& bc_2(p-1)+o(b)+o(b)+o(b) \\
=& bc_2(p-1)+o(b).
\end{aligned}
$$

Next we consider $v_{12}v_{22}^{-2}v_{21}v_{12}v_{22}^{-1}v_{21}$. Let $V = \left[\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_{d-1}(t)dB_{d-1}'(s)\right]^{-1}$ for ease of notation. Use Itô's integral from equation (3.4), and let $u_1,\ldots,u_n$ be the partition for $t$ and $s$,

$E[v_{12}v_{22}^{-2}v_{21}v_{12}v_{22}^{-1}v_{21}]$

$$
\begin{aligned}
=&E\left\{\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_1(t)dB'_{d-1}(s)\right.\\
&\left[\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_1(t)dB'_{d-1}(s)\right]^{-1}\\
&\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB'_1(t)dB_{d-1}(s)\\
&\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_1(t)dB'_{d-1}(s)\\
&\left[\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB_1(t)dB'_{d-1}(s)\right]^{-1}\\
&\left.\int_0^1\int_0^1 \tilde{\kappa}_b(t,s)dB'_1(t)dB_{d-1}(s)\right\}\\
=&\lim_{n\to\infty} E\left\{\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n\sum_{h=1}^n\int_0^1 \tilde{\kappa}_b(u_{i-1},s^{(1)})[B_1(u_i)-B_1(u_{i-1})]dB'_{d-1}(s^{(1)})V\right.\\
&\times\int_0^1 \tilde{\kappa}_b(t^{(2)},u_{j-1})[B_1(u_j)-B_1(u_{j-1})]dB_{d-1}(t^{(2)})\\
&\times\int_0^1 \tilde{\kappa}_b(u_{l-1},s^{(3)})[B_1(u_l)-B_1(u_{l-1})]dB'_{d-1}(s^{(3)})V\\
&\left.\times\int_0^1 \tilde{\kappa}_b(t^{(4)},u_{h-1})[B_1(u_h)-B_1(u_{h-1})]dB_{d-1}(t^{(4)})\right\}. \qquad (3.29)
\end{aligned}
$$

We only need to consider three cases: $(i=j,l=h)$, $(i=l,j=h)$, $(i=h,j=l)$. Under the first case equation (3.29) is equal to the following,

$$
\begin{aligned}
\lim_{n\to\infty} E&\left\{\sum_{i=1}^n\sum_{j=1}^n\int_0^1 \tilde{\kappa}_b(u_{i-1},s^{(1)})[u_i-u_{i-1}]dB'_{d-1}(s^{(1)})V\int_0^1 \tilde{\kappa}_b(t^{(2)},u_{i-1})dB_{d-1}(t^{(2)})\right.\\
&\left.\int_0^1 \tilde{\kappa}_b(u_{j-1},s^{(3)})[u_j-u_{j-1}]dB'_{d-1}(s^{(3)})V\int_0^1 \tilde{\kappa}_b(t^{(4)},u_{j-1})dB_{d-1}(t^{(4)})\right\}
\end{aligned}
$$

$$= \lim_{n\to\infty} E\left\{ \sum_{i=1}^{n} \int_0^1 \int_0^1 \tilde{\kappa}_b(u_{i-1}, s^{(1)}) \tilde{\kappa}_b(t^{(2)}, u_{i-1})[u_i - u_{i-1}] dB'_{d-1}(s^{(1)}) V dB_{d-1}(t^{(2)}) \right.$$
$$\left. \times \sum_{j=1}^{n} \int_0^1 \int_0^1 \tilde{\kappa}_b(u_{j-1}, s^{(3)}) \tilde{\kappa}_b(\tau^{(4)}, u_{j-1})[u_j - u_{j-1}] dB'_{d-1}(s^{(3)}) V dB_{d-1}(t^{(4)}) \right\}.$$

We utilize Riemann-Stieltjes integration and the property in equation (3.24) to further simplify $E(v_{11}v_{12}v_{22}^{-1}v_{21})$ under the first case,

$$\begin{aligned}
& E\left\{ \int_0^1 \int_0^1 \left[ \int_0^1 \tilde{\kappa}_b(\tau_1, s^{(1)}) \tilde{\kappa}_b(t^{(2)}, \tau_1) d\tau_1 \right] dB'_{d-1}(s^{(1)}) V dB_{d-1}(t^{(2)}) \right. \\
& \left. \times \int_0^1 \int_0^1 \left[ \int_0^1 \tilde{\kappa}_b(\tau_2, s^{(3)}) \tilde{\kappa}_b(t^{(4)}\tau_2) d\tau_2 \right] dB'_{d-1}(s^{(3)}) V dB_{d-1}(t^{(4)}) \right\} \\
=& E\left\{ \int_0^1 \int_0^1 \left[ \sum_{m=1}^{\infty} \lambda_m^2 \phi_m(s^{(1)}) \phi_m(t^{(2)}) \right] dB'_{d-1}(s^{(1)}) V dB_{d-1}(t^{(2)}) \right. \\
& \left. \times \int_0^1 \int_0^1 \left[ \sum_{m=1}^{\infty} \lambda_m^2 \phi_m(s^{(1)}) \phi_m(t^{(2)}) \right] dB'_{d-1}(s^{(3)}) V dB_{d-1}(t^{(4)}) \right\} \\
=& E\left\{ trace\left( \int_0^1 \int_0^1 \left[ \sum_{m=1}^{\infty} \lambda_m^2 \phi_m(s^{(1)}) \phi_m(t^{(2)}) \right] dB'_{d-1}(s^{(1)}) V dB_{d-1}(t^{(2)}) \right) \right. \\
& \left. \times trace\left( \int_0^1 \int_0^1 \left[ \sum_{m=1}^{\infty} \lambda_m^2 \phi_m(s^{(1)}) \phi_m(t^{(2)}) \right] dB'_{d-1}(s^{(3)}) V dB_{d-1}(t^{(4)}) \right) \right\}.
\end{aligned}$$

The last step was obtained by recognizing that for an arbitrary for a $d \times 1$ vector $\mathbf{a}$ and $d \times d$ matrix $V$, $\mathbf{a}'V\mathbf{a} = trace(\mathbf{a}'V\mathbf{a})$. We further simplify $E(v_{11}v_{12}v_{22}^{-1}v_{21})$ under the first

case by using the cyclical property of trace operator, the property in equation (3.25), and properties of convergence in probability and continuous functions [50, pg 290],

$$
\begin{aligned}
& E\left\{ trace\left( V\int_0^1\int_0^1\left[\sum_{m=1}^{\infty}\lambda_m^2\phi_m(s^{(1)})\phi_m(t^{(2)})\right]dB_{d-1}(t^{(2)})dB'_{d-1}(s^{(1)})\right)\right. \\
& \left.\times\, trace\left( V\int_0^1\int_0^1\left[\sum_{m=1}^{\infty}\lambda_m^2\phi_m(s^{(3)})\phi_m(t^{(4)})\right]dB_{d-1}(t^{(4)})dB'_{d-1}(s^{(3)})\right)\right\} \\
=& E\left\{ trace\left( V\sum_{m=1}^{\infty}\lambda_m^2\xi_m\xi'_m)\right) trace\left( V\sum_{m=1}^{\infty}\lambda_m^2\xi_m\xi'_m\right)\right\} \\
=& E\left\{ trace\left( V\sum_{m=1}^{\infty}\lambda_m^2\xi_m\xi'_m)\right)^2\right\} \\
=& E\left\{ trace\left(\left[\sum_{n=1}^{\infty}\lambda_n\xi_n\xi'_n\right]\sum_{m=1}^{\infty}\lambda_m^2\xi_m\xi'_m)\right)^2\right\} \\
\approx& \frac{\mu_2^2}{\mu_1^2}(d-1)^2(1+o(1)).
\end{aligned}
$$

We next review the second case, $(i = l, j = h)$. From equation (3.29),

$$
\begin{aligned}
\lim_{n\to\infty} E\Bigg\{ & \sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^1 \tilde{\kappa}_b(u_{i-1}, s^{(1)})[u_i - u_{i-1}]dB'_{d-1}(s^{(1)})V \\
& \times\int_0^1 \tilde{\kappa}_b(t^{(2)}, u_{j-1})[u_j - u_{j-1}]dB_{d-1}(t^{(2)}) \\
& \times\int_0^1 \tilde{\kappa}_b(u_{i-1}, s^{(3)})dB'_{d-1}(s^{(3)})V\int_0^1 \tilde{\kappa}_b(t^{(4)}, u_{j-1})dB_{d-1}(t^{(4)})\Bigg\}.
\end{aligned}
$$

We further simplify $E(v_{11}v_{12}v_{22}^{-1}v_{21})$ under case 2 by using Riemann-Stieltjes integration and symmetric properties,

$$\begin{aligned}
&\lim_{n\to\infty} E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^1\int_0^1 \tilde{\kappa}_b(u_{i-1},s^{(1)})\tilde{\kappa}_b(u_{i-1},s^{(3)})[u_i-u_{i-1}]dB'_{d-1}(s^{(1)})V\right.\\
&\left.\times\int_0^1\int_0^1 \tilde{\kappa}_b(t^{(2)},u_{j-1})\tilde{\kappa}_b(t^{(4)},u_{j-1})[u_j-u_{j-1}]dB_{d-1}(t^{(2)})dB'_{d-1}(s^{(3)})VdB_{d-1}(t^{(4)})\right\}\\
=&E\left\{\int_0^1\int_0^1\left[\int_0^1 \tilde{\kappa}_b(\tau_3,s^{(1)})\tilde{\kappa}_b(\tau_3,s^{(3)})d\tau_3\right]dB'_{d-1}(s^{(1)})V\right.\\
&\left.\times\int_0^1\int_0^1\left[\int_0^1 \tilde{\kappa}_b(t^{(2)},\tau_4)\tilde{\kappa}_b(t^{(4)},\tau_4)d\tau_4\right]dB_{d-1}(t^{(2)})dB'_{d-1}(s^{(3)})VdB_{d-1}(t^{(4)})\right\}\\
=&E\left\{\int_0^1\int_0^1\int_0^1\int_0^1\left[\int_0^1 \tilde{\kappa}_b(\tau_3,s^{(1)})\tilde{\kappa}_b(\tau_3,s^{(3)})d\tau_3\right]\left[\int_0^1 \tilde{\kappa}_b(t^{(2)},\tau_4)\tilde{\kappa}_b(t^{(4)},\tau_4)d\tau_4\right]\right.\\
&\left.\times dB'_{d-1}(s^{(1)})VdB_{d-1}(t^{(2)})dB'_{d-1}(s^{(3)})VdB_{d-1}(t^{(4)})\right\}\\
=&E\left\{\int_0^1\int_0^1\int_0^1\int_0^1\left[\int_0^1 \tilde{\kappa}_b(t^{(2)},\tau_3)\tilde{\kappa}_b(\tau_3,s^{(3)})d\tau_3\right]\left[\int_0^1 \tilde{\kappa}_b(\tau_4,s^{(1)})\tilde{\kappa}_b(t^{(4)},\tau_4)d\tau_4\right]\right.\\
&\left.\times dB'_{d-1}(s^{(1)})VdB_{d-1}(t^{(2)})dB'_{d-1}(s^{(3)})VdB_{d-1}(t^{(4)})\right\}.
\end{aligned}$$

We continue to simplify under case 2 in a similar fashion as in case 1. Namely, we use a relationship between vectors and matrices ($\mathbf{a}'V\mathbf{a} = trace(\mathbf{a}'V\mathbf{a})$), the cyclical property of the trace operator, the property equation (3.24), and properties of convergence in probability and continuous functions [50, pg 290],

$$
\begin{aligned}
&Etrace\left\{\int_0^1\int_0^1\int_0^1\int_0^1\left[\int_0^1\tilde{\kappa}_b(t^{(2)},\tau_3)\tilde{\kappa}_b(\tau_3,s^{(3)})d\tau_3\right]\right.\\
&\left.\times\left[\int_0^1\tilde{\kappa}_b(\tau_4,s^{(1)})\tilde{\kappa}_b(t^{(4)},\tau_4)d\tau_4\right]dB'_{d-1}(s^{(1)})VdB_{d-1}(t^{(2)})dB'_{d-1}(s^{(3)})VdB_{d-1}(t^{(4)})\right\}\\
=&Etrace\left\{V\int_0^1\int_0^1\left[\int_0^1\tilde{\kappa}_b(t^{(2)},\tau_3)\tilde{\kappa}_b(\tau_3,s^{(3)})d\tau_3\right]dB_{d-1}(t^{(2)})dB'_{d-1}(s^{(3)})\right.\\
&\left.\times V\int_0^1\int_0^1\left[\int_0^1\tilde{\kappa}_b(\tau_4,s^{(1)})\tilde{\kappa}_b(t^{(4)},\tau_4)d\tau_4\right]dB_{d-1}(t^{(4)})dB'_{d-1}(s^{(1)})\right\}\\
=&Etrace\left\{V\sum_{n=1}^{\infty}\lambda_n^2\xi_n\xi_n'V\sum_{m=1}^{\infty}\lambda_m^2\xi_m\xi_m'\right\}\\
=&trace\left\{\mu_1^{-1}\mathbf{1}_{(d-1)}\mu_2\mathbf{1}_{(d-1)}\mu_1^{-1}\mathbf{1}_{(d-1)}\mu_2\mathbf{1}_{(d-1)}\right\}\\
\approx&(d-1)\frac{\mu_2^2}{\mu_1^2}(1+o(1)).
\end{aligned}
$$

By symmetry the third and second case yield the same result. Therefore,

$$
\begin{aligned}
E[v_{12}v_{22}^{-2}v_{21}v_{12}v_{22}^{-1}v_{21}]\approx&\left(\frac{\mu_2}{\mu_1}\right)^2\left((d-1)^2+2(d-1)\right)(1+o(1))\\
=&\left(\frac{\mu_2}{\mu_1}\right)^2(d-1)(d+1)(1+o(1))\\
\approx&o(b)
\end{aligned}
$$

$$
\begin{aligned}
E\left[(v_{11}-v_{12}v_{22}^{-1}v_{21})^2\right]=&E\left[v_{11}^2-2v_{11}v_{12}v_{22}^{-1}v_{21}+v_{12}v_{22}^{-1}v_{21}v_{12}v_{22}^{-1}v_{21}\right]\\
\approx&1-2b(c_1-c_2)-2bc_2(d-1)+o(b).
\end{aligned}
$$

■

*Begin proof of Lemma 3c.*

The expected value of $\left[(v_{11}-v_{12}v_{22}^{-1}v_{21})-1\right]^2$ is obtained directly from the previous results,

$$E\left[(v_{11} - v_{12}v_{22}^{-1}v_{21}) - 1\right]^2$$

$$\begin{aligned}
&= E\left[(v_{11} - v_{12}v_{22}^{-1}v_{21})^2\right] - 2E(v_{11} - v_{12}v_{22}^{-1}v_{21}) + 1 \\
&= 1 - 2b(c_1 - c_2) - 2bc_2(d-1) - 2\left[1 - bc_1 - bc_2(d-1)\right] + 1 + o(b) \\
&= 2bc_2 + o(b).
\end{aligned}$$

■

Lemma 3 is concerned with calculating the first two moments of the random variable $v_{11\cdot 2}$. We see that this function is concentrated around 1, but is influenced by $b, c_1$ and $c_2$. This is expected given the relationship with the $\chi^2$ distribution in the univariate case and the standard Wishart distribution in the multivariate case. For both Lemma 2 and Lemma 3 we rely on the assumptions that $b \to 0$ and $T \to \infty$ but there is no restrictions on the relationship between these quantities.

**Theorem 2:** *As $b \to 0$ and $T \to \infty$, we have*

$$P(dF_\infty(d,b) \leq z) = G_d(z) - G'_d(z)z\left(c_1 + c_2(d-1)\right)b + G''_d(z)z^2c_2b + o(b)$$

*where $G_d(z)$ is the cumulative distribution function function of a $\chi^2$ random variable with degrees of freedom $d$ at $z$, and $G'_d(z)$ is its first derivative.*

*Begin proof.*

We start by applying Proposition 1, and then take a Taylor series expansion around $z$. Let $\tilde{z}$ be some real value between $z$ and $zv_{11\cdot 2}$,

$$\begin{aligned}
P(dF_\infty(d,b) \leq z) =& E\left[G_d(z(v_{11\cdot 2}))\right] \\
\approx& G_d(z) + G_d'(z)E\left[(zv_{11\cdot 2} - z)\right] + \frac{1}{2}G_d''(\tilde{z})E\left[(zv_{11\cdot 2} - z)^2\right] \\
=& G_d(z) + G_d'(z)E\left[(zv_{11\cdot 2} - z)\right] \\
&+ \frac{1}{2}\left(G_d''(\tilde{z}) + G_d''(z) - G_d''(z)\right)E\left[(zv_{11\cdot 2} - z)^2\right] \\
=& G_d(z) + G_d'(z)zE\left[(v_{11\cdot 2} - 1)\right] + \frac{1}{2}G_d''(z)z^2E\left[(v_{11\cdot 2} - 1)^2\right] \\
&+ \frac{1}{2}\left(G_d''(\tilde{z}) - G_d''(z)\right)z^2E\left[(v_{11\cdot 2} - 1)^2\right].
\end{aligned}$$

We further approximate by using the results in Lemma 3,

$$\begin{aligned}
P(dF_\infty(d,b) \leq z) \approx& G_d(z) + G_d'(z)z\left(-bc_1 - bc_2(d-1)\right) + \frac{1}{2}G_d''(z)z^2\left(2bc_2\right) + o(b) \\
=& G_d(z) - G_d'(z)z\left(c_1 + c_2(d-1)\right)b + G_d''(z)z^2\left(c_2\right)b + o(b).
\end{aligned}$$

■

The first component of Theorem 2 is the standard small-$b$ $\chi^2$ distributional approximation. As our sample size increases and $b$ decreases this term alone is acceptable as there will be no practical contribution from the remaining terms and we can use the large sample inference procedures. However, as already discussed, this is typically not sufficient for finite samples.

The random variable $v_{11\cdot 2}$ captures many of the bias and variance properties of $\hat{\Omega}_T$. In all components beyond the small-$b$ distributional approximation we see a dependence on the kernel function and the bandwidth as expected. The value $c_2(d-1)$ in the second term is

dependent on the number of hypotheses restrictions, disappearing in the 1-dimensional case and increasing as more restrictions are added. The value $bc_1$ in the second term arises from calculating the first moment of $v_{11\cdot 2}$ and is generally reflecting the uncertainty in parameter estimation through use of the centered kernel [54]. The value $c_2 b$ in the third term can be viewed as the approximate asymptotic variability of $v_{11\cdot 2}$, which is observed via Lemma 3 and recognizing $Var(v_{11\cdot 2}) \approx 2bc_2 + o(b^2)$. The random variable $v_{11\cdot 2}$ reflects variability of parameter estimation through the use of the centered kernel, is dependent on a bandwidth parameter and kernel function, and suffers from dimensionality bias.

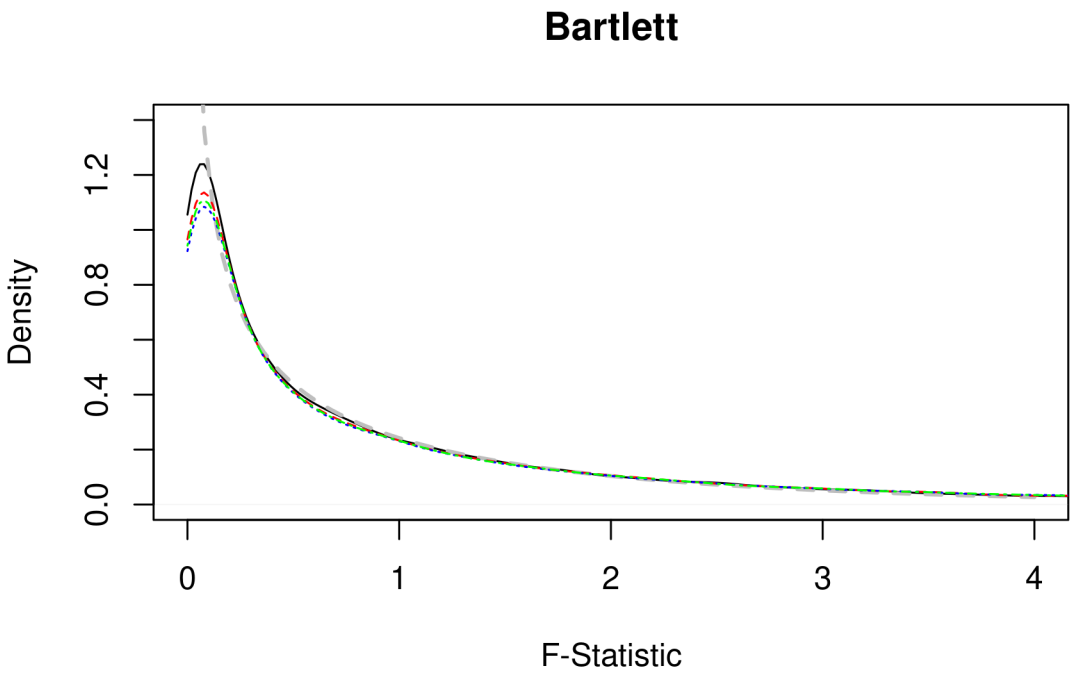


Figure 3.1: Fixed-$b$ density of $F_T$ using the Bartlett kernel at $b = 0.07$ with mother (solid black), zero (dashed red), adaptive (dotted blue), and over (dot-dashed green) lugsail settings. Small-$b$ density (dashed grey) provided as a reference.

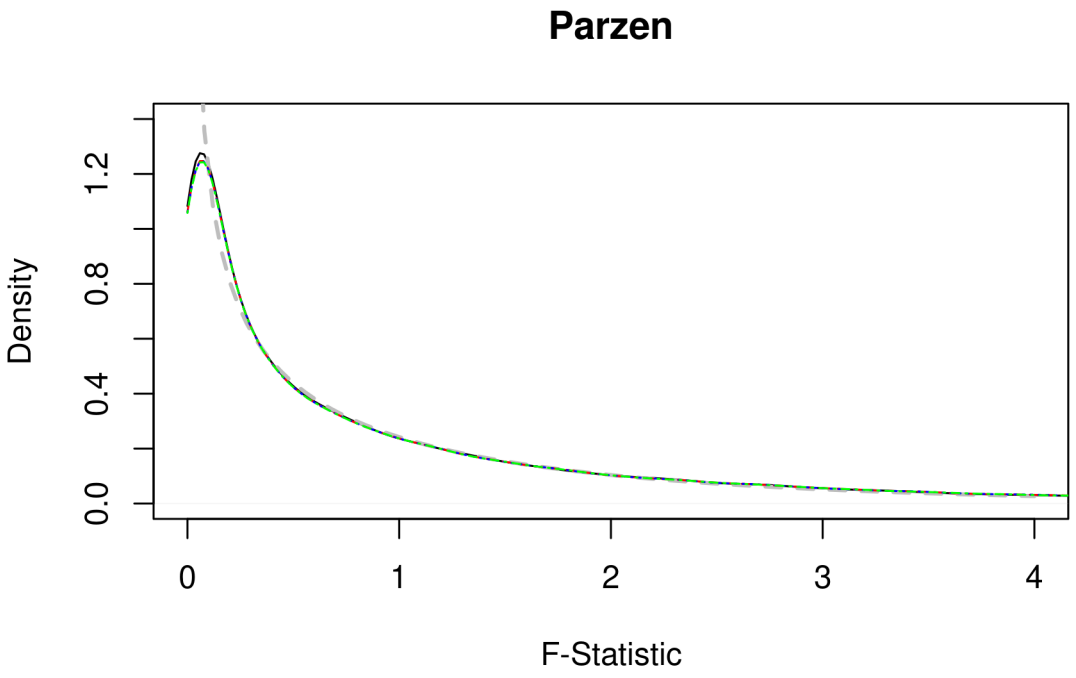


Figure 3.2: Fixed-$b$ density of $F_T$ using the Parzen kernel at $b = 0.07$ with mother (solid black), zero (dashed red), adaptive (dotted blue), and over (dot-dashed green) lugsail settings. Small-$b$ density (dashed grey) provided as a reference.

However, as evident in Theorem 1, the asymptotic bias attributed to the kernel function is reflected through $(bT)^{-q} g_q h_q$, which is not present in the expression for the fixed-$b$ distribution in Theorem 2. In general, the fixed-$b$ random variables capture the variability of $\hat{\Omega}_T$, but not the bias. This is reflected in Figures 3.1-3.3 as we observe the longer tails for the more extreme lugsail settings. In these figures we have the density functions for univariate fixed-$b$ random variables $F_\infty$ estimated using 10,000 simulations with a sample size of $T = 1000$ and $b = 0.07$. In each figure there are five density functions: the standard $\chi^2_1$ (dashed grey), mother kernel (solid black), zero lugsail kernel (dashed red), adaptive lugsail kernel (dotted blue), and over lugsail (dot-dashed green). Each fixed-$b$ density relies

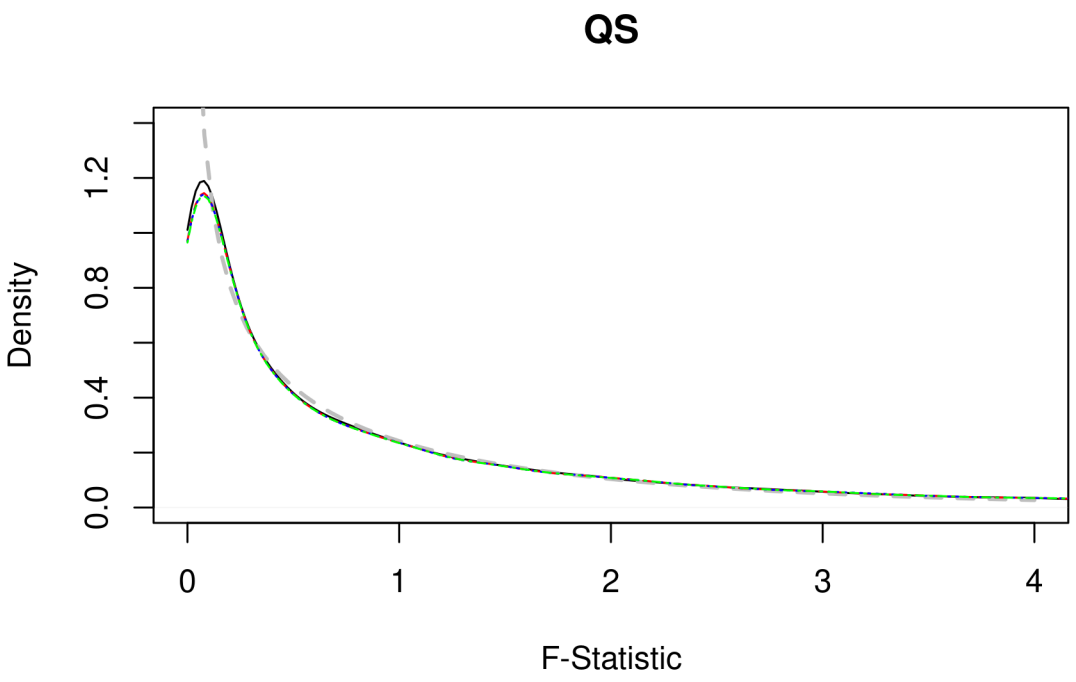


Figure 3.3: Fixed-$b$ density of $F_T$ using the quadratic spectral kernel at $b = 0.07$ with mother (solid black), zero (dashed red), adaptive (dotted blue), and over (dot-dashed green) lugsail settings. Small-$b$ density (dashed grey) provided as a reference.

on the centered version of the kernel, using the Bartlett, Parzen, and quadratic spectral mother kernels in Figures 3.1, 3.2, and 3.3, respectively.

# Chapter 4

# Finite Sample Behavior of the Lugsail Estimator

The previous chapter presented distributional representations and expressions for probabilities for the fixed-$b$ random variable $F_\infty$. The results use a fixed bandwidth and rely on Brownian motions, and hence model the behavior of the test statistic $F_T$ using an infinite sample [1, 40]. The fixed-$b$ assumption was imposed for the distributional representations in Proposition 1, but relaxed when approximating associated probabilities. This limiting framework is applicable to a wide variety of settings; however, it is well known that their may be large discrepancies between the exact distribution of a test statistic with a finite sample and a large sample limiting distribution. In this chapter we produce a probability expression for $F_T$ generated with a finite sample under a classic but common and versatile setting. We revisit the fixed-$b$ distribution again in Chapter 5.

We also introduce a new assumption that to the best of our knowledge is not used elsewhere in this space. This assumption expands the expression for the bias of $\hat{\Omega}_T$ introduced in Theorem 1 and Corollary 1 in order to incorporate the finite sample bias.

**Assumption 6:** *Let $\tilde{\kappa} \in \mathscr{K}_3$, and assume that $bT \to \infty$ as $T \to \infty$, but at a slower rate. Then there exists some value $c_b$ as a function of $b$ such that,*

$$\begin{aligned} Bias(\hat{\Omega}) =& E[\hat{\Omega}] - \Omega \\ =& -(bT)^{-q} g_q h_q - \Omega c_1 b - \Omega c_b - o\left((bT)^q\right). \end{aligned} \tag{4.1}$$

The first term comes from Theorem 1, it is an asymptotic expression for the kernel bias utilizing $\kappa \in \mathscr{K}_2 \cup \mathscr{K}_3$. The second component comes from Corollary 1 and arises from the de-meaning of the error terms. The term $\Omega c_b$ is novel and represents the finite sampling bias, which we assume is proportional to the true LRV as a function of $b$. Provided the data is finite, for data sets generated by even the simplest of models with serial correlation we expect $c_b$ to be nonzero.

Assumption 6 is natural and critical for understanding the bias of lugsail estimators. For $\kappa^* \in \mathscr{K}_1$ the dominating source of bias is typically attributed to the kernel, which we assume has weights that decrease at a faster rate then the autocovariance function. Typically other components are comparatively negligible and difficult to represent robustly, hence they are usually ignored. Since zero lugsail estimators have $g_\infty = 0$, the other sources of bias become more crucial. We can estimate $c_b$ with moderate model structures which we explore in the subsequent chapter.

For now we focus on deriving general probability expressions with a finite sample using the Gaussian location model,

$$y_t = \theta_0 + w_t, \tag{4.2}$$

where $y_t$, $\theta_0$, and $w_t$ are all $p \times 1$ vectors, $t = (1, \ldots, T)$, and $w_t$ is a Gaussian random process with zero mean and variance-covariance matrix $\Sigma_w$. We consider the following set of hypotheses,

$$\begin{aligned} H_0 : &\qquad R\theta = R\theta_0 \\ H_A : &\quad R\theta = R\theta_0 + R\Lambda T^{-1/2}\delta, \end{aligned} \tag{4.3}$$

where $R$ is a $d \times p$ real valued matrix, $\Omega = \Lambda\Lambda'$, and $\delta$ is uniformly distributed on the real $d$ dimensional sphere centered at the origin and with radius $\delta$ [27, 54]. The Guassian location model at (4.2) with moment conditions $f(y_t, \theta) = (\overline{y} - \theta)$ is a special version of a GMM model [11, 54]. With these exactly identified moment conditions we can select any arbitrary weighting matrix because we have the same number of moment conditions as unknown parameters ($m = p$). With $G_T = -\mathbb{I}_p$ and $W_T = \mathbb{I}_p$, under Lemma 1 we have

$$\sqrt{T}R\left[\hat{\theta}_T - \theta_0\right] \xrightarrow{d} N(0, \Omega).$$

The typical $F$ and $t$ test statistics are given by

$$\begin{aligned} F_T =& \left[\sqrt{T}R\left(\hat{\theta}_T - \theta_0\right)\right]' \hat{\Omega}_T^{-1} \left[\sqrt{T}R\left(\hat{\theta}_T - \theta_0\right)\right] / d \\ t_T =& \sqrt{T}R\hat{\theta}_T\hat{\Omega}_T^{-1/2} \end{aligned} \tag{4.4}$$

where $\hat{\Omega}_T$ is a SV estimator defined in (2.5) with $\tilde{\kappa} \in \mathscr{K}_3$, and $u_t = R(y_t - \hat{\theta}_T)$. Using the small-$b$ limiting distribution under the null hypothesis the test statistic $dF_T$ converges to a $\chi^2$ random variable with $d$ degrees of freedom.

Up until this point we have discussed the estimator for $\hat{\Omega}_T$ but not $\hat{\theta}_T$. We will primarily focus on the ordinary least squares (OLS) estimator $\theta$, which for the model in equation (4.2) is $\hat{\theta}_{OLS} = \frac{1}{T}\sum_{t=1}^{T} y_t$. However, there are some properties we will utilize from generalized least squares (GLS) estimators, $\hat{\theta}_{GLS} = \left[(l_T \otimes \mathbb{I}_p)'\Omega_v^{-1}(l_T \otimes \mathbb{I}_p)\right]^{-1}(l_T \otimes \mathbb{I}_p)'\Omega_v^{-1}y$, where $\underset{(T\times 1)}{l_T}$ is a vector of ones, $\underset{(pT\times 1)}{v} = [v_1', \ldots, v_T']'$ is the observed data, and $Var\,(vec(v)) = \Omega_v$. For the Gaussian location model the observed data is $v_t = y_t$. Observe that the $F_T$ test statistic in equation (2.4) relies on the function $r(\cdot)$ and the estimator $\hat{\Omega}_T$, both of which depend on $\hat{\theta}_T$. The estimate for $\hat{\theta}_T$ does not need to be the same for both components for Lemma 1 to hold [11, 54]. In fact, if both components rely on $\hat{\theta}_{OLS}$, then we can not say that $r(\hat{\theta}_{OLS})$ and $\hat{\Omega}_T(\hat{\theta}_{OLS})$ are generally independent. Instead we can rely on $r(\hat{\theta}_{GLS})$ and $\hat{\Omega}_T(\hat{\theta}_{OLS})$, which results in a better test statistic to work with analytically and practically [11]. From this point on assume $\hat{\Omega}_T$ relies on the OLS estimator $\hat{\theta}_{OLS}$ unless otherwise stated, let $\hat{\theta}_T$ represent any consistent estimator of $\theta$, and let $\hat{\theta}_{OLS}$ and $\hat{\theta}_{GLS}$ be the OLS and GLS estimators of $\theta$. Furthermore,

$$F_{OLS} = \left[\sqrt{T}R\left(\hat{\theta}_{OLS} - \theta_0\right)\right]' \hat{\Omega}_T^{-1}\left[\sqrt{T}R\left(\hat{\theta}_{OLS} - \theta_0\right)\right]/d$$
$$F_{GLS} = \left[\sqrt{T}R\left(\hat{\theta}_{GLS} - \theta_0\right)\right]' \hat{\Omega}_T^{-1}\left[\sqrt{T}R\left(\hat{\theta}_{GLS} - \theta_0\right)\right]/d.$$

In the following Proposition we confirm the independence of $(\hat{\theta}_{GLS} - \theta)$ and $u_t$, which lets us establish independence between the numerator and denominator of $F_{GLS}$. This again mimics the independence seen in the fixed-$b$ distribution representation in equation (3.1) and Proposition 1, and will be useful for obtaining probability expressions. Later in Lemma 4 we establish a connection between $F_{GLS}$ and $F_{OLS}$.

**Proposition 2:** *Under Assumption 5i and 5ii,*

$$\hat{\theta}_{GLS} - \theta \perp \hat{u}_t, \tag{4.5}$$

*for* $t = (1, \ldots, T)$.

*Begin proof.*

We use similar steps as Sun [54]. Let $\underset{(dT\times 1)}{\hat{u}} = [\hat{u}_1', \ldots, \hat{u}_T]'$. Observe, that

$$\begin{aligned}
\hat{u}_t &= \underset{(d\times p)}{R}\,(y_t - \hat{\theta}_{OLS}) \\
&= R(y_t - \overline{y}) \\
&= R(v_t - \overline{v})
\end{aligned}$$

and $(l_T l_T'/T \otimes \mathbb{I}_p)v$ is equal to the $Tp \times 1$ vector $[\overline{v}, \ldots \overline{v}]'$. Thus $[(\mathbb{I}_T - l_T l_T'/T) \otimes \mathbb{I}_p]\, v = [v_1 - \overline{v}, \ldots, v_T - \overline{v}]'$ and $\hat{u} = (R \otimes \mathbb{I}_T)\,[(\mathbb{I}_T - l_T l_T'/T) \otimes \mathbb{I}_p]\, v$. If $\mathrm{Cov}(\hat{\theta}_{GLS} - \theta, \hat{u}_t) = \underset{(p\times d)}{0}$

for all $t$, then equivalently, $\text{Cov}\left(\left[(l_T \otimes \mathbb{I}_p)'\Omega_v^{-1}(l_T \otimes \mathbb{I}_p)\right]^{-1}(l_T \otimes \mathbb{I}_p)\Omega_v^{-1}v, \hat{u}\right) = \underset{(p\times Td)}{0}$. We check this relation using algebraic manipulation,

$$
\begin{aligned}
\text{Cov}(\hat{\theta}_{GLS} - \theta, \hat{u}) =& \text{Cov}\Big(\left[(l_T \otimes \mathbb{I}_p)'\Omega_v^{-1}(l_T \otimes \mathbb{I}_p)\right]^{-1}(l_T \otimes \mathbb{I}_p)\Omega_v^{-1}v, \\
& \qquad (R \otimes \mathbb{I}_T)\left[(\mathbb{I}_T - l_T l_T'/T) \otimes \mathbb{I}_p\right]v\Big) \\
=& E\left\{\left[(l_T \otimes \mathbb{I}_p)'\Omega_v^{-1}(l_T \otimes \mathbb{I}_p)\right]^{-1}(l_T \otimes \mathbb{I}_p)\Omega_v^{-1}vv'\left[(\mathbb{I}_T - l_T l_T'/T) \otimes \mathbb{I}_p\right]\right\} \\
& \times (R \otimes \mathbb{I}_T)' + 0 \\
=& E\left\{\left[(l_T \otimes \mathbb{I}_p)'\Omega_v^{-1}(l_T \otimes \mathbb{I}_p)\right]^{-1}(l_T \otimes \mathbb{I}_p)\Omega_v^{-1}\Omega_v\left[(\mathbb{I}_T - l_T l_T'/T) \otimes \mathbb{I}_p\right]\right\} \\
& \times (R \otimes \mathbb{I}_T)' \\
=& E\left\{\left[(l_T \otimes \mathbb{I}_p)'\Omega_v^{-1}(l_T \otimes \mathbb{I}_p)\right]^{-1}(l_T \otimes \mathbb{I}_p)'\left[(\mathbb{I}_T - l_T l_T'/T) \otimes \mathbb{I}_p\right]\right\} \\
& \times (R \otimes \mathbb{I}_T)' \\
=& E\left\{\left[(l_T \otimes \mathbb{I}_p)'\Omega_v^{-1}(l_T \otimes \mathbb{I}_p)\right]^{-1}\right\} \\
& \times (l_T \otimes \mathbb{I}_p)'\left[(\mathbb{I}_T - l_T l_T'/T) \otimes \mathbb{I}_p\right](R \otimes \mathbb{I}_T)' \\
=& E\left\{\left[(l_T \otimes \mathbb{I}_p)'\Omega_v^{-1}(l_T \otimes \mathbb{I}_p)\right]^{-1}\right\} \\
& \times (l_T' \otimes \mathbb{I}_p)\left[(\mathbb{I}_T - l_T l_T'/T) \otimes \mathbb{I}_p\right](R \otimes \mathbb{I}_T)' \\
=& E\left\{\left[(l_T \otimes \mathbb{I}_p)'\Omega_v^{-1}(l_T \otimes \mathbb{I}_p)\right]^{-1}\right\} \\
& \times \left(l_T'(\mathbb{I}_T - l_T l_T'/T)\right) \otimes [\mathbb{I}_p]\ (R \otimes \mathbb{I}_T)'.
\end{aligned}
$$

Observe that $l_T'(\mathbb{I}_T - l_T l_T'/T)$ produces a $1 \times T$ vector with elements $\left[(1 - \frac{1}{T}) - \sum_{t=1}^{T-1}\frac{1}{T}\right] = 0$. Therefore, $(l_T'(\mathbb{I}_T - l_T l_T'/T)) \otimes [\mathbb{I}_p] = \underset{(p\times pT)}{0}$ and $\text{Cov}(\hat{\theta}_{GLS} - \theta, \hat{u}) = \underset{(p\times dT)}{0}$, as desired.

■

**Lemma 4:** *Under Assumption 4 and 5,*

*(a)* $P(dF_{GLS} \leq z) = E\left[G_p(z\Xi_T^{-1})\right] + O(T^{-1})$

*(b)* $P(dF_{OLS} \leq z) = P(pF_{GLS} \leq z) + O(T^{-1})$

*where*

$$\underset{(1\times 1)}{\Xi_T}\left(\hat{\Omega}_T\right) = e_T'[\Omega^{1/2}\hat{\Omega}_T^{-1}\Omega^{1/2}]e_T$$

$$\underset{(d\times 1)}{e_T} = \frac{\Omega_{GLS}^{-1/2}\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)}{||\Omega_{GLS}^{-1/2}\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)||}$$

*and* $\Omega_{GLS}$ *is the variance of* $\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)$.

*Begin proof of Lemma 4a.*

Let $\Xi_T(\hat{\Omega}_T) = \Xi_T$ for ease of notation. Define $\Upsilon_T := ||\Omega_{GLS}^{-1/2}\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)||^2$ and $\Xi_{T,GLS} := e_T'\Omega_{GLS}^{1/2}\hat{\Omega}_T^{-1}\Omega_{GLS}^{1/2}e_T$. We can rewrite the definition of $F_{GLS}$ in equation (4.4) using these values,

$$\begin{aligned}
dF_{GLS} &= \left[\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)\right]'\hat{\Omega}_T^{-1}\left[\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)\right] \\
&= \left\{\left[\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)\right]'\Omega_{GLS}^{-1/2}\right\}\Omega_{GLS}^{1/2}\hat{\Omega}_T^{-1}\Omega_{GLS}^{1/2}\left\{\Omega_{GLS}^{-1/2}\left[\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)\right]\right\} \\
&= ||\Omega_{GLS}^{-1/2}\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)||e_T'\Omega_{GLS}^{1/2}\hat{\Omega}_T^{-1}\Omega_{GLS}^{1/2}e_T||\Omega_{GLS}^{-1/2}\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)|| \\
&= ||\Omega_{GLS}^{-1/2}\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)||^2e_T'\Omega_{GLS}^{1/2}\hat{\Omega}_T^{-1}\Omega_{GLS}^{1/2}e_T \\
&= \Upsilon_T\Xi_{T,GLS}.
\end{aligned}$$

By Proposition 2, $\Upsilon_T$ and $\Xi_{T,GLS}$ are independent. Also observe that $e_T'\Omega_{GLS}^{1/2}\hat{\Omega}_T^{-1}\Omega_{GLS}^{1/2}e_T = ||e_T'\Omega_{GLS}^{1/2}\hat{\Omega}_T^{-1/2}||^2 > 0$, thus we can take the inverse of $\Xi_{T,GLS}$. We obtain an expression for $P(dF_{GLS} \leq z)$ using these facts and the tower rule of expectation,

$$\begin{aligned}
P(dF_{GLS} \leq z) =& P(\Upsilon_T \Xi_{T,GLS} \leq z) \\
=& P(\Upsilon_T \leq z\left(\Xi_{T,GLS}\right)^{-1}) \\
=& E\left(G_p(z\left(\Xi_{T,GLS}\right)^{-1})\right) \\
=& E\left(G_p\left(z\left(e_T'\Omega_{GLS}^{1/2}\hat{\Omega}_T^{-1}\Omega_{GLS}^{1/2}e_T\right)^{-1}\right)\right).
\end{aligned}$$

By the mean value theorem, for $\epsilon, M \in \mathbb{R}^+$ and $\epsilon' \in (0, \epsilon)$

$$\begin{aligned}
\frac{G_d(x+\epsilon T) - G_d(x)}{\epsilon T} =& G_d'(x+\epsilon' T) \\
G_d(x+\epsilon T) - G_d(x) =& (\epsilon T)^{-1} G_d'(x+\epsilon' T),
\end{aligned}$$

and because $|G_d'(x)| < M < \infty$ (bounded function), we further observe,

$$\begin{aligned}
|G_d(x+\epsilon T) - G_d(x)| <& (\epsilon T)^{-1} M = O(T^{-1}) \\
\Rightarrow G_d(x+\epsilon T) - G_d(x) =& O(T^{-1}) \\
G_d(x+\epsilon T) =& G_d(x) + O(T^{-1}).
\end{aligned}$$

Using the mean value theorem result and recognizing $\Omega_{GLS} = \Omega(1 + O(T^{-1}))$, we obtain the final result

$$\begin{aligned}
P(dF_{GLS} \leq z) =& E\left(G_p\left(z\left(e_T'\left\{\Omega(1+O(T^{-1}))\right\}^{1/2}\hat{\Omega}_T^{-1}\left\{\Omega(1+O(T^{-1}))\right\}^{1/2}e_T\right)^{-1}\right)\right) \\
=& E\left[G_p\left(z\Xi_T^{-1}\right)\right] + O(T^{-1}).
\end{aligned}$$

■

*Proof of Lemma 4b.*

Define the following:

- $\zeta_{1T} = 2(\sqrt{T}R\Delta)'\hat{\Omega}^{-1}\Omega_{T,GLS}^{1/2}e_T$
- $\zeta_{2T} = (\sqrt{T}R\Delta)'\hat{\Omega}_T^{-1}(\sqrt{T}R\Delta)$
- $\zeta_T = \zeta_{2T} + \zeta_{1T}\sqrt{\Upsilon_T}$
- $\Delta = \hat{\theta}_{GLS} - \theta_0 - (\hat{\theta}_{OLS} - \theta_0) \Rightarrow (\hat{\theta}_{OLS} - \theta_0) = \Delta + (\hat{\theta}_{GLS} - \theta_0)$.

We begin by deriving an alternative expression for $dF_{OLS}$ in terms of the $dF_{GLS}$,

$$\begin{aligned}
dF_{OLS} =& \left[\sqrt{T}R(\hat{\theta}_{OLS} - \theta_0)\right]' \hat{\Omega}_T^{-1} \left[\sqrt{T}R(\hat{\theta}_{OLS} - \theta_0)\right] \\
=& \left[\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0) + \sqrt{T}R\Delta\right]' \hat{\Omega}_T^{-1} \left[\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0) + \sqrt{T}R\Delta\right] \\
=& \left[\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)\right]' \hat{\Omega}_T^{-1} \left[\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)\right] + \left[\sqrt{T}R\Delta\right]' \hat{\Omega}_T^{-1} \left[\sqrt{T}R\Delta\right] \\
&+ 2\left[\sqrt{T}R\Delta\right]' \hat{\Omega}_T^{-1} \left[\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)\right] \\
=& dF_{GLS} + \zeta_{2T} + 2(\sqrt{T}R\Delta)'\hat{\Omega}_T^{-1}\Omega_{T,GLS}^{1/2} \frac{\Omega_{T,GLS}^{-1/2}\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)}{||\Omega_{T,GLS}^{-1/2}\sqrt{T}R(\hat{\theta}_{GLS} - \theta_0)||}\sqrt{\Upsilon_T} \\
=& dF_{GLS} + \zeta_{2T} + 2(\sqrt{T}R\Delta)'\hat{\Omega}_T^{-1}\Omega_{T,GLS}^{1/2} e_T\sqrt{\Upsilon_T} \\
=& dF_{GLS} + \zeta_{2T} + \zeta_{1T}\sqrt{\Upsilon_T}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
P\left(dF_{OLS} \leq z\right) =& P(dF_{GLS} + \zeta_{2T} + \zeta_{1T}\sqrt{\Upsilon_T} \leq z) \\
=& P\left(\Upsilon_T\Xi_{T,GLS} + \zeta_{2T} + \zeta_{1T}\sqrt{\Upsilon_T} \leq z\right).
\end{aligned}$$

Define $F(a,b,c) = P\left(\Upsilon_T c + b + a\sqrt{\Upsilon_T} \leq z\right)$. Let

$$F_a = \tfrac{\partial}{\partial a}F \qquad F_{aa} = \left(\tfrac{\partial}{\partial a}\right)^2 F$$
$$F_b = \tfrac{\partial}{\partial b}F \qquad F_{bb} = \left(\tfrac{\partial}{\partial b}\right)^2 F$$
$$F_{ab} = \tfrac{\partial}{\partial a}\tfrac{\partial}{\partial b}F.$$

Take a multivariate Taylor Series approximation for variables $a$ and $b$, both around 0,

$$\begin{aligned} F(\zeta_{1T},\zeta_{2T},\Xi_{T,GLS}) \approx & F(0,0,\Xi_{T,GLS}) \\ & + F_a(0,0,\Xi_{T,GLS})(\zeta_{1T}) + \frac{1}{2}F_{aa}(0,0,\Xi_{T,GLS})(\zeta_{1T})^2 \\ & + F_b(0,0,\Xi_{T,GLS})\zeta_{2T} + \frac{1}{2}F_{bb}(0,0,\Xi_{T,GLS})(\zeta_{2T})^2 \\ & + F_{ab}(0,0,\Xi_{T,GLS})(\zeta_{1T}\zeta_{2T}). \end{aligned} \tag{4.6}$$

From Proposition 2 we observe that $\Upsilon_T$ is independent of $(\zeta_{1T},\zeta_{2T},\Xi_{T,GLS})$. Therefore,

$$\begin{aligned} E\left[F(\zeta_{1T},\zeta_{2T},\Xi_{T,GLS})\right] \approx & E\left[F(0,0,\Xi_{T,GLS})\right] \\ & + E[F_a(0,0,\Xi_{T,GLS})(\zeta_{1T})] + \frac{1}{2}E[F_{aa}(0,0,\Xi_{T,GLS})(\zeta_{1T})^2] \\ & + E[F_b(0,0,\Xi_{T,GLS})\zeta_{2T}] + \frac{1}{2}E[F_{bb}(0,0,\Xi_{T,GLS})(\zeta_{2T})^2] \\ & + E[F_{ab}(0,0,\Xi_{T,GLS})(\zeta_{1T}\zeta_{2T})] \\ = & E\left[F(0,0,\Xi_{T,GLS})\right] \\ & + E[F_a(0,0,\Xi_{T,GLS})]E[\zeta_{1T}] + \frac{1}{2}E[F_{aa}(0,0,\Xi_{T,GLS})]E[\zeta_{1T}^2] \\ & + E[F_b(0,0,\Xi_{T,GLS})]E[\zeta_{2T}] + \frac{1}{2}E[F_{bb}(0,0,\Xi_{T,GLS})]E[\zeta_{2T}^2] \\ & + E[F_{ab}(0,0,\Xi_{T,GLS})]E[\zeta_{1T}\zeta_{2T}]. \end{aligned}$$

Observe that $E[\zeta_{2T}] = O(T^{-1})$ and $E[\zeta_{1T}^2] = O(T^{-1})$, hence

$$\begin{aligned} E\left[F(\zeta_{1T}, \zeta_{2T}, \Xi_{T,GLS})\right] \approx & E\left[F(0,0,\Xi_{T,GLS})\right] + E[F_a(0,0,\Xi_{T,GLS})]E[\zeta_{1T}] \\ & + O\left(E[\zeta_{1T}^2]\right) + O\left(E[\zeta_{2T}]\right) + O\left(E[\zeta_{2T}^2]\right) + O\left(E[\zeta_{1T}\zeta_{2T}]\right) \\ = & E\left[F(0,0,\Xi_{T,GLS})\right] + E[F_a(0,0,\Xi_{T,GLS})]E[\zeta_{1T}] + O(T^{-1}). \end{aligned}$$

Let $f_e(x)$ be the probability density function (pdf) of $e_T$, then

$$\begin{aligned} E[F_a(0,0,\Xi_{T,GLS})\zeta_{1T}] &= \int E[F_a(0,0,\Xi_{T,GLS})\zeta_{1T}|e_T = x]f_e(x)dx \\ &= \int E[F_a(0,0,x'\Omega_{GLS}^{1/2}\hat{\Omega}_T^{-1}\Omega_{GLS}^{1/2}x)]2(\sqrt{T}R\Delta)'\hat{\Omega}^{-1}\Omega_{GLS}^{1/2}xf_e(x)dx. \end{aligned}$$

We note that $e_T$ is a standard normal vector, $\hat{\Omega}_T$ is an even function of the estimated errors, and $\Delta$ is an odd function of the estimated errors. The product of an odd function and even functions results in an odd function, and the integral of an odd function on a symmetric interval about zero is zero. Thus, $E[F_a(0,0,\Xi_{T,GLS})\zeta_{1T}] = 0$, and

$$\begin{aligned} P\left(dF_{OLS} \leq z\right) \approx & E\left[F(0,0,\Xi_{T,GLS})\right] + O(T^{-1}) \\ = & P(\Upsilon_T \Xi_{T,GLS} < z) + O(T^{-1}) \\ = & P(dF_{GLS} < z) + O(T^{-1}). \end{aligned}$$

■

The test statistic $F_{GLS}$ was used to obtain a probability expression in terms of a familiar distribution by using the tower property. In Lemma 4b we further establish an expression for the probability of $F_{OLS}$, and see the two statistics are asymptotically equivalent. A similar property holds for $\hat{\theta}_{OLS}$ and $\hat{\theta}_{GLS}$ which are also asymptotically equivalent [9].

The random variable $\Xi_T^{-1}$ serves a similar role as $v_{11\cdot 2}$ in that it is harnessing the variability and the general behavior of $\hat{\Omega}_T$. In the next Lemma we establish an expansion for $\Xi_T^{-1}$ in a similar spirit as Lemma 3 for $v_{11\cdot 2}$.

**Lemma 5:** *Let $\kappa \in \mathscr{K}_2 \cup \mathscr{K}_3$ and assume that $bT \to \infty$ as $T \to \infty$, but at a slower rate. Under Assumption 5 and 6,*

*(a) The Taylor series expansion of $\Xi_T^{-1}$ around $\Xi_T(\Omega)$ is*

$$\Xi_T^{-1} \approx 1 + L + Q + remainder$$

*(b)* $E(L) = -(bT)^{-q} g_q w_q - c_1 b - c_b + o((bt)^{-q})$

*(c)* $E(Q) = -bc_2(d-1) + o(b)$

*(d)* $E(L^2) = 2c_2 b + o(b)$

*where*

- $\underset{(1\times 1)}{L} = Dvec(\hat{\Omega}_T - \Omega)$
- $\underset{(1\times 1)}{Q} = \frac{1}{2} vec(\hat{\Omega}_T - \Omega)'(J_1 + J_2) vec(\hat{\Omega}_T - \Omega)$
- $\underset{(1\times d^2)}{D} = \left(\left[e_T'\Omega^{-1/2}\right] \otimes \left[e_T'\Omega^{-1/2}\right]\right)$
- $\underset{(d^2\times d^2)}{J_1} = \left[2\Omega^{-1/2}(e_T e_T')\Omega^{-1/2}\right] \otimes \left[\Omega^{-1/2}(e_T e_T')\Omega^{-1/2}\right]$
- $\underset{(d^2\times d^2)}{J_2} = -\left[\Omega^{-1/2} e_T e_T' \Omega^{-1/2} \otimes \Omega^{-1}\right] \mathbb{K}_{dd} \left(\mathbb{I}_{d^2} + \mathbb{K}_{dd}\right)$
- $w_q = d^{-1} trace\left(\sum_{j=-\infty}^{\infty} |j|^q \Gamma(j)\Omega^{-1}\right) = d^{-1} trace\left(h_q \Omega^{-1}\right)$, *the trace of the scaled version of the Parzen generalized $q^{th}$ derivative at the origin.*

*Begin proof of Lemma 5a.*

First we observe that

$$\begin{aligned}\Xi_T(\Omega) =& e_T'[\Omega^{1/2}\Omega_T^{-1}\Omega^{1/2}]e_T \\ =& e_T'e_T \\ =& 1.\end{aligned}$$

Therefore, $\Xi_T^{-1}(\Omega) = 1$. We next compute the first derivative of $\Xi_T^{-1}$ with respect to $\hat{\Omega}_T$ using the chain rule, and adopt the notation $\frac{\partial}{\partial vec(\hat{\Omega})} = d$,

$$\begin{aligned}\frac{\partial}{\partial vec\left(\hat{\Omega}\right)}\Xi_T^{-1} =& -\Xi_T^{-2}\frac{\partial}{\partial vec\left(\hat{\Omega}\right)}\Xi_T \\ d\Xi_T^{-1} =& -\Xi_T^{-2}d\Xi_T \\ =& -\Xi_T^{-2}d\left[e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\Omega^{1/2}e_T\right] \\ =& -\Xi_T^{-2}e_T'\Omega^{1/2}d\hat{\Omega}_T^{-1}\Omega^{1/2}e_T.\end{aligned}$$

We can use the matrix chain rule to observe that: $-d\hat{\Omega}^{-1} = \hat{\Omega}_T^{-1}(d\hat{\Omega}_T)\hat{\Omega}_T^{-1}$. Hence,

$$d\Xi_T^{-1} = \Xi_T^{-2}e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}(d\hat{\Omega}_T)\hat{\Omega}_T^{-1}\Omega^{1/2}e_T.$$

Furthermore, observe that for some matrix $\underset{(d\times d)}{Y}$ with components $[Y]_{i,j} = y_{ij}$, and vector $\underset{(1\times d)}{\mathbf{r}} = [r_1, \ldots, r_p]$ we have,

$$\begin{aligned}
\mathbf{r}Y\mathbf{r}' &= \begin{bmatrix} r_1 & \ldots & r_p \end{bmatrix} \begin{bmatrix} y_{11} & & \\ & \ddots & \\ & & y_{pp} \end{bmatrix} \begin{bmatrix} r_1 \\ \vdots \\ r_p \end{bmatrix} \\
&= \begin{bmatrix} \sum_j r_j y_{j1} & \ldots & \sum_j r_j y_{jp} \end{bmatrix} \begin{bmatrix} r_1 \\ \vdots \\ r_p \end{bmatrix} \\
&= \sum_i r_i \left( \sum_j r_j y_{ji} \right).
\end{aligned}$$

We further recall that the Kronecker product $\underset{(1\times pp)}{(\mathbf{r}\otimes\mathbf{r})} \underset{(dd\times 1)}{vec(Y)} = \sum_i \sum_j r_i r_j y_{ij} = \mathbf{r}Y\mathbf{r}'$. Let $\mathbf{r} = e_T'\Omega^{1/2}\hat{\Omega}^{-1}$ and recall $dvec\left(\hat{\Omega}_T\right) = 1$ by definition,

$$\begin{aligned}
d\Xi_T^{-1} =& -\Xi_T^{-2}\mathbf{r}d\hat{\Omega}_T\mathbf{r}' \\
=& \Xi_T^{-2}\left[\mathbf{r}\otimes\mathbf{r}\right] dvec\left(\hat{\Omega}_T\right) \\
=& \Xi_T^{-2}\left[\left(e_T'\Omega^{1/2}\hat{\Omega}^{-1}\right)\otimes\left(e_T'\Omega^{1/2}\hat{\Omega}^{-1}\right)\right] dvec\left(\hat{\Omega}_T\right) \\
=& \Xi_T^{-2}\left[\left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\otimes\left(e_T'\Omega_T^{1/2}\hat{\Omega}^{-1}\right)\right].
\end{aligned}$$

Now we evaluate $d\Xi_T^{-1}$ at $\Omega$. Recall that $\Xi_T(\Omega) = 1$, thus $\Xi_T^{-2}(\Omega) = 1$, and

$$\begin{aligned}
d\Xi_T^{-1}(\Omega) =& \left[\left(e_T'\Omega^{-1/2}\right)\otimes\left(e_T'\Omega^{-1/2}\right)\right] \\
:=& D.
\end{aligned}$$

We continue with the second derivative. Let $f(x)$ and $g(x) = g_1(x) + g_2(x)$ be arbitrary differentiable functions. Using the chain rule and product rule derive the following expression,

$$\begin{aligned}\frac{\partial}{\partial x} f^{-2}(x)g(x) =& (-2)f^{-3}(x)\left(\frac{\partial}{\partial x}f(x)\right)g(x) + f^{-2}(x)\left(\frac{\partial}{\partial x}g(x)\right) \\ =& (-2)f(x)^{-3}\left(\frac{\partial}{\partial x}f(x)\right)g(x) + f^{-2}(x)\left(\frac{\partial}{\partial x}g_1(x)\right) + f^{-2}(x)\left(\frac{\partial}{\partial x}g_2(x)\right) \\ =& D_1 + D_2 + D_3.\end{aligned}$$

The above notation is a general expression, for our case we have the following,

$$d\frac{\partial \Xi^{-1}}{\partial \left[vec(\hat{\Omega})\right]'} = D_1 + D_2 + D_3$$

where $f(x) = \Xi_T$ and $g(x) = \left[\left(e_T'\Omega^{-1/2}\hat{\Omega}_T^{-1}\right) \otimes \left(e_T'\Omega^{-1/2}\hat{\Omega}_T^{-1}\right)\right]$

$$\begin{aligned}D_1 =& -2\Xi_T^{-3}d\Xi_T\left[\left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right) \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right] \\ D_2 =& -\Xi_T^{-2}\left[\left(e_T'\Omega^{1/2}(d\hat{\Omega}_T^{-1})\right) \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}^{-1}\right)\right] \\ D_3 =& -\Xi_T^{-2}\left[\left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right) \otimes \left(e_T'\Omega^{1/2}(d\hat{\Omega}_T^{-1})\right)\right] \\ =& D_2\mathbb{K}_{dd}.\end{aligned}$$

We can find alternative representations for these terms. For example, observe that

$$\begin{aligned}d\Xi_T =& \left[\left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right) \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right] dvec\left(\hat{\Omega}_T\right) \\ =& dvec\left(\hat{\Omega}_T\right)'\left[\left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right) \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right]',\end{aligned}$$

a relationship observed during the derivation for the first derivative of $\Xi_T^{-1}$. We can substitute this term inside of $D_1$,

$$\begin{aligned} D_1 = &- 2\Xi_T^{-3} dvec\left(\hat{\Omega}_T\right)' \left[\left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right) \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right]' \\ &\times \left[\left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right) \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right]. \end{aligned}$$

We can also find an alternative representation for $D_2$ and $D_3$. In this alternative representation we can use a result from earlier, that $-d\hat{\Omega}_T^{-1} = \hat{\Omega}_T^{-1}(d\hat{\Omega}_T)\hat{\Omega}_T^{-1}$, and observe

$$\begin{aligned} D_2 =& - \Xi_T^{-2}\left[\left(e_T'\Omega^{1/2} d\hat{\Omega}_T^{-1}\hat{\Omega}_T\hat{\Omega}_T^{-1}\right) \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right] \\ =&\Xi_T^{-2}\left[\left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}(d\hat{\Omega}_T)\hat{\Omega}_T^{-1}\right) \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right] \\ =&\Xi_T^{-2}\left[\left(\hat{\Omega}_T^{-1}(d\hat{\Omega}_T)\hat{\Omega}_T^{-1}\Omega^{1/2}e_T\right)' \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right] \\ =&\Xi_T^{-2}\left[\left(\hat{\Omega}_T^{-1}(d\hat{\Omega}_T)\left[e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right]'\right)' \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right] \\ =&\Xi_T^{-2}\left[\left(ZY\mathbf{r}'\right)' \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right] \end{aligned}$$

where $Z = \hat{\Omega}_T^{-1}$, $Y = d\Omega_T$, and $r = \left[e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right]$. Notice vector $ZY\mathbf{r}' = vec(ZY\mathbf{r}')$ and $vec(ZYr') = (r \otimes Z)vec(Y)$, which gives us

$$\begin{aligned} D_2 =&\Xi_T^{-2}\left[\left(vec(ZY\mathbf{r}')\right)' \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}^{-1}\right)\right] \\ =&\Xi_T^{-2}\left[\left((r \otimes Z)vec(Y)\right)' \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right]. \end{aligned}$$

Substituting values back in we obtain,

$$\begin{aligned} D_2 =&\Xi_T^{-2}\left[\left(\left(\left[e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right] \otimes \hat{\Omega}_T^{-1}\right) dvec\left(\hat{\Omega}_T\right)\right)' \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right] \\ =&\Xi_T^{-2} dvec\left(\hat{\Omega}_T\right)'\left[\left(\left[e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right] \otimes \hat{\Omega}_T^{-1}\right)' \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right] \\ =&\Xi_T^{-2} dvec\left(\hat{\Omega}_T\right)'\left[\left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)' \otimes \hat{\Omega}_T^{-1} \otimes \left(e_T'\Omega^{1/2}\hat{\Omega}_T^{-1}\right)\right]. \end{aligned}$$

Now we can calculate the second derivative of $\Xi_T^{-1}$ at $\Omega$. Recall $dvec\left(\hat{\Omega}\right) = 1$, $\Xi_T(\Omega) = 1$, and $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$. Therefore,

$$\begin{aligned}
\frac{\partial^2 \Xi_T(\Omega)}{\partial vec\left(\hat{\Omega}\right) \partial vec\left(\hat{\Omega}\right)} =& 2(1)\left(\left[e_T'\Omega^{-1/2}\right]' \otimes \left[e_T'\Omega^{-1/2}\right]'\right)\left(\left[e_T'\Omega^{-1/2}\right] \otimes \left[e_T'\Omega^{-1/2}\right]\right) \\
& - \left(\left[e_T'\Omega^{-1/2}\right]' \otimes \Omega^{-1} \otimes \left[e_T'\Omega^{-1/2}\right]\right)\left(\mathbb{I}_{d^2} + \mathbb{K}_{dd}\right) \\
=& 2\left[\Omega^{-1/2} e_T e_T' \Omega^{-1/2}\right] \otimes \left[\Omega^{-1/2} e_T e_T' \Omega^{-1/2}\right] \\
& - \left[e_T'\Omega^{-1/2}\right]' \otimes \Omega^{-1} \otimes \left[e_T'\Omega^{-1/2}\right]\left(\mathbb{I}_{d^2} + \mathbb{K}_{dd}\right) \\
=& 2\left[\Omega^{-1/2} e_T e_T' \Omega^{-1/2}\right] \otimes \left[\Omega^{-1/2} e_T e_T' \Omega^{-1/2}\right] \\
& - \left[\Omega^{-1/2} e_T e_T' \Omega^{-1/2} \otimes \Omega^{-1}\right] \mathbb{K}_{dd}\left(\mathbb{I}_{d^2} + \mathbb{K}_{dd}\right) \\
:=& J_1 + J_2.
\end{aligned}$$

■

*Begin proof of Lemma 5b.*

Recall Proposition 2, which implies $e_T \perp \hat{\Omega}_T$. Therefore, under Assumption 6 we have,

$$\begin{aligned}
E(L) =& E[Dvec(\hat{\Omega}_T - \Omega)] \\
=& E e_T' \Omega^{-1/2}(\hat{\Omega}_T - \Omega)\Omega^{-1/2} e_T \\
=& -(bT)^{-q} E e_T' \Omega^{-1/2}(g_q h_q)\Omega^{-1/2} e_T - (c_1 + c_b) E e_T' \Omega^{-1/2}\Omega\Omega^{-1/2} e_T + o\left((bT)^{-q}\right) \\
\approx& -(bT)^{-q} E e_T' \Omega^{-1/2}(g_q h_q)\Omega^{-1/2} e_T - (c_1 + c_b) + o\left((bT)^{-q}\right).
\end{aligned}$$

For a symmetric matrix $\underset{(d\times d)}{A}$ and row vector $\underset{(d\times 1)}{\mathbf{r}}$ : $\mathbf{r}'A\mathbf{r} = \sum_j r_j \left(\sum_i r_i a_{ij}\right)$ and $[A\mathbf{r}\mathbf{r}']_{ij} = r_j \left(\sum_i a_{ij} r_i\right)$. Therefore $trace(A\mathbf{r}\mathbf{r}') = \mathbf{r}'A\mathbf{r} = \sum_j r_j \left(\sum_i a_{ij} r_i\right)$. We further observer that $E[e_T e_T'] = \mathbb{I}_d E\left(\frac{1}{||\Omega_{GLS}^{-1/2}\sqrt{T}(R_0\hat{\theta}_{GLS} - r_0)||}\right) = \frac{1}{d}$. With these properties we obtain,

$$\begin{aligned}
E(L) \approx & -(bT)^{-q} g_q E\left[trace\left(\Omega^{-1/2} h_q \Omega^{-1/2} e_T e_T'\right)\right] - (c_1 + c_b) + o\left((bT)^{-q}\right) \\
= & -(bT)^{-q} g_q E\left[trace\left(h_q \Omega^{-1} e_T e_T'\right)\right] - c_1 - c_b + o\left((bT)^{-q}\right) \\
= & -(bT)^{-q} \frac{g_q}{d} trace\left(h_q \Omega^{-1}\right) - c_1 - c_b + o\left((bT)^{-q}\right) \\
= & -(bT)^{-q} g_q w_q - c_1 - c_b + o\left((bT)^{-q}\right).
\end{aligned}$$

■

*Begin proof of Lemma 5c.*

Let $V, B, C$ be arbitrary $d \times d$ square matrices, and $(a, b)$ be arbitrary $d \times 1$ vectors. We will rely on the following properties:

$$(V \otimes B) vec(C) = vec\left(VCB'\right) \tag{4.7}$$

$$trace(A'B) = trace(AB') \tag{4.8}$$

$$a'b = trace(ba') \tag{4.9}$$

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD). \tag{4.10}$$

Furthermore, if $(A, B, C, D)$ are all symmetric matrices,

$$trace(ABCD) = trace(ACDB). \tag{4.11}$$

We start by finding an expression for $E(Q) = \frac{1}{2}E\left[vec\left(\hat{\Omega}_T - \Omega\right)(J_1 + J_2)\left(\hat{\Omega}_T - \Omega\right)\right]$. We first study just the $J_1$ term. Using property (4.7),

$$\begin{aligned}
\frac{1}{2}E\left[vec\left(\hat{\Omega}_T - \Omega\right)' J_1 vec\left(\hat{\Omega}_T - \Omega\right)\right] =& \frac{2}{2}E\left[vec\left(\hat{\Omega}_T - \Omega\right)'\left[\Omega^{-1/2}e_T e_T'\Omega^{-1/2}\right]\right. \\
& \left.\otimes\left[\Omega^{-1/2}e_T e_T'\Omega^{-1/2}\right] vec\left(\hat{\Omega}_T - \Omega\right)\right] \\
=& E\left[vec\left(\hat{\Omega} - \Omega\right)' vec\left\{\Omega^{-1/2}e_T e_T'\Omega^{-1/2}\right]\right. \\
& \left.\times\left(\hat{\Omega} - \Omega\right)\Omega^{-1/2}e_T e_T'\Omega^{-1/2}\right\}\Big] \\
=& E\left[\left(e_T'\Omega^{-1/2}\left(\hat{\Omega} - \Omega\right)\Omega^{-1/2}e_T\right)^2\right] \\
=& E[(X'AX)^2]
\end{aligned}$$

where $A = \Omega^{-1/2}\left(\hat{\Omega} - \Omega\right)\Omega^{-1/2}$, $[A]_{ij} = A_{ij}$,$X = e_T$. Using the tower property of expectations and Proposition 2 we can rewrite this as the following,

$$\begin{aligned}
E[(X'AX)^2] =& E\left[\sum_{i,j}\sum_{m,l} A_{ij}A_{lm}X_iX_jX_lX_m\right] \\
=& E\left[\sum_{i,j}\sum_{m,l} A_{ij}A_{lm}E[X_iX_jX_lX_m|A]\right] \\
=& \sum_{i,j}\sum_{m,l} E[A_{ij}A_{lm}]E[X_iX_jX_lX_m].
\end{aligned}$$

Without loss of generality, if $X_i$ is different from $(X_j, X_l, X_m)$, then $E[X_iX_jX_lX_m] = 0$ [23, 52]. Thus, we only care about four cases: $(i = j = l = m)$, $(i = j, l = m)$, $(i = l, m = j)$, and $(i = m, j = l)$. Furthermore, $E[X_i^4] = \frac{3}{d(d+2)}$ and $E[X_i^2X_j^2] = \frac{1}{d(d+2)}$. We break up the expectation into these four cases,

$$\begin{aligned}
\sum_{i,j}\sum_{m,l} A_{ij}A_{lm}E[X_iX_jX_lX_m] &= \sum_{i=j=l=m} A_{ii}^2E(X_i^4) + \sum_{i=j}\sum_{l=m} A_{ii}A_{jj}E[X_i^2X_m^2] \\
&\quad + \sum_{i=l}\sum_{j=m} A_{ij}^2E[X_i^2X_j^2] + \sum_{j=l}\sum_{j=l} A_{ij}^2E[X_i^2X_j^2] \\
&= \sum_{i=j=l=m} A_{ii}^2E(X_i^4) + \sum_{i=j}\sum_{l=m} A_{ii}A_{jj}E[X_i^2X_m^2] \\
&\quad + 2\sum_{i=l}\sum_{j=m} A_{ij}^2E[X_i^2X_j^2] \\
&= \sum_{i=j=l=m} A_{ii}^2\left(\frac{3}{d(d+2)}\right) + \sum_{i=j}\sum_{l=m} A_{ii}A_{jj}\left(\frac{1}{d(d+2)}\right) \\
&\quad + 2\sum_{i=l}\sum_{j=m} A_{ij}^2\left(\frac{1}{d(d+2)}\right).
\end{aligned}$$

We still need the expected value of the $A$ terms. Notice that the off-diagonals of $A$ squared are 0, i.e. $E[A_{ij}^2] = 0$ for $i \neq j$. Therefore,

$$\begin{aligned}
\sum_{i,j}\sum_{m,l} E\left[A_{ij}A_{lm}\right]E[X_iX_jX_lX_m] &= \sum_{i=j=l=m} E[A_{ii}^2]\left(\frac{3}{d(d+2)}\right) + \sum_{i=j}\sum_{l=m} E[A_{ii}A_{jj}]\left(\frac{1}{d(d+2)}\right) \\
&= \left(\frac{3}{d(d+2)}\right)E\left[trace\left(AA\right)\right] \\
&\quad + \left(\frac{1}{d(d+2)}\right)E\left[trace^2\left(A\right) - trace(AA)\right] \\
&= \left(\frac{1}{d(d+2)}\right)2\left[E\left[trace(AA)\right] + E\left[trace^2(A)\right]\right].
\end{aligned}$$

Next we find $E[trace(AA)]$. We use the trace property illustrated in equation (4.8) to rearrange,

$$
\begin{aligned}
E[trace(AA)] =& E\left[trace\left(vec'(A)vec(A)\right)\right] \\
=& E\left[trace\left(vec'(\Omega^{-1/2}(\hat{\Omega}_T-\Omega)\Omega^{-1/2})vec(\Omega^{-1/2}(\hat{\Omega}_T-\Omega)\Omega^{-1/2})\right)\right] \\
=& E\left[trace\left(vec'(\hat{\Omega}_T-\Omega)(\Omega^{-1/2}\otimes\Omega^{-1/2})(\Omega^{-1/2}\otimes\Omega^{-1/2})vec(\hat{\Omega}_T-\Omega)\right)\right] \\
=& E\left[trace\left((\Omega^{-1/2}\otimes\Omega^{-1/2})vec(\hat{\Omega}_T-\Omega)vec'(\hat{\Omega}_T-\Omega)(\Omega^{-1/2}\otimes\Omega^{-1/2})\right)\right].
\end{aligned}
$$

To further simplify we use Theorem 1, a trace property of symmetric matrices illustrated in equation (4.11), and a multiplication property of matrices in equation (4.10),

$$
\begin{aligned}
E[trace(AA)] =& trace\left((\Omega^{-1/2}\otimes\Omega^{-1/2})Var\left(vec(\hat{\Omega}_T\right)(\Omega^{-1/2}\otimes\Omega^{-1/2})\right) \\
\approx& bc_2 trace\left((\Omega^{-1/2}\otimes\Omega^{-1/2})\left(\mathbb{I}_{d^2}+\mathbb{K}_{dd}\right)(\Omega\otimes\Omega)(\Omega^{-1/2}\otimes\Omega^{-1/2})\right)+o(b) \\
=& bc_2 trace\left((\Omega^{-1/2}\otimes\Omega^{-1/2})(\Omega\otimes\Omega)(\Omega^{-1/2}\otimes\Omega^{-1/2})\left(\mathbb{I}_{d^2}+\mathbb{K}_{dd}\right)\right)+o(b) \\
=& bc_2 trace\left((\Omega^{0}\otimes\Omega^{0})\left(\mathbb{I}_{d^2}+\mathbb{K}_{dd}\right)\right)+o(b) \\
=& bc_2 trace\left(\mathbb{I}_{d^2}+\mathbb{K}_{dd}\right)+o(b) \\
=& bc_2\left[trace\left(\mathbb{I}_{d^2}\right)+trace\left(\mathbb{K}_{dd}\right)\right]+o(b) \\
=& bc_2\left[d^2+d\right]+o(b).
\end{aligned}
$$

Next we derive $E\left[trace^2(A)\right]$. To further simplify we use a trace property for symmetric matrices illustrated in equation (4.11), and observe that $(\sum_i A_{ii})^2=\sum_i\sum_j A_{ii}Ajj$,

$$
\begin{aligned}
E\left[trace^2(A)\right] =& E\left[trace^2\left(\Omega^{-1/2}(\hat{\Omega}_T-\Omega)\Omega^{-1/2})\right)\right] \\
=& E\left[trace^2\left((\hat{\Omega}_T-\Omega)\Omega^{-1})\right)\right] \\
=& E\left[trace\left((\hat{\Omega}_T-\Omega)\Omega^{-1})\right)trace\left((\hat{\Omega}-\Omega)\Omega^{-1})\right)\right] \\
=& E\left[vec'\left((\hat{\Omega}_T-\Omega)\Omega^{-1})\right)vec(\left((\hat{\Omega}_T-\Omega)\Omega^{-1})\right)\right].
\end{aligned}
$$

Let $[\hat{\Omega}_T - \Omega]_{ij} = \phi_{ij}$, and $[\Omega^{-1}]_{ij} = \omega_{ij}$. For $i \neq j$, $E(\phi_{ij}^2) = 0$, and for $(i,j) = (l,k)$ we have that $E(\phi_{ij}\phi_{lk}) = 0$. Hence $vec'\left((\hat{\Omega}_T - \Omega)\Omega^{-1})\right) vec\left((\hat{\Omega}_T - \Omega)\Omega^{-1})\right) = \sum_l (\phi\omega)_l^2 = \sum_l \phi_{ll}^2\omega_{ll}^2$. We also observe that

$$\begin{aligned} vec'(\Omega^{-1})vec(\hat{\Omega}_T - \Omega)vec'(\hat{\Omega}_T - \Omega)vec(\Omega^{-1}) &= \left(\sum_l (\omega\phi)_l\right)\left(\sum_h (\omega\phi)_h\right) \\ &= \sum_l \sum_h (\omega\phi)_l \sum_l (\omega\phi)_h \\ &= \sum_l (\omega\phi)_l^2 \\ &= \sum_l \omega_{ll}^2\phi_{ll}^2. \end{aligned}$$

Therefore $vec'\left((\hat{\Omega}_T - \Omega)\Omega^{-1})\right) vec(\left((\hat{\Omega}_T - \Omega)\Omega^{-1})\right) = vec'(\Omega^{-1})vec(\hat{\Omega}_T - \Omega)vec'(\hat{\Omega}_T - \Omega)vec(\Omega^{-1})$, and we have the following,

$$E\left[trace^2(A)\right] = E\left[vec'(\Omega^{-1})vec(\hat{\Omega}_T - \Omega)vec'(\hat{\Omega}_T - \Omega)vec(\Omega^{-1})\right].$$

Furthermore by Theorem 1 and property (4.7),

$$\begin{aligned} E\left[trace^2(A)\right] =& vec'(\Omega^{-1})Var(vec(\hat{\Omega}_T))vec(\Omega^{-1}) \\ \approx& bc_2 vec'(\Omega^{-1})\left(\Omega \otimes \Omega\right)\left(\mathbb{I}_{d^2} + \mathbb{K}_{dd}\right)vec(\Omega^{-1}) + o(b) \\ =& 2bc_2 vec'(\Omega^{-1})\left(\Omega \otimes \Omega\right)vec(\Omega^{-1}) + o(b) \\ =& 2bc_2 vec'(\Omega^{-1})vec(\Omega\Omega^{-1}\Omega) + o(b) \\ =& 2bc_2 vec'(\Omega^{-1})vec(\Omega) + o(b) \\ =& 2bc_2 d + o(b). \end{aligned}$$

We now have what we need for $\frac{1}{2}E\left[vec'\left(\hat{\Omega}_T-\Omega\right)J_1 vec\left(\hat{\Omega}_T-\Omega\right)\right]$,

$$\begin{aligned}\frac{1}{2}E\left[vec'\left(\hat{\Omega}_T-\Omega\right)J_1 vec\left(\hat{\Omega}_T-\Omega\right)\right] =& \frac{1}{d(d+2)}\left[2E\left(trace(AA)\right)+E\left(trace^2(A)\right)\right]\\ \approx& \frac{1}{d(d+2)}\left[2(bc_2)(d^2+d)+2bc_2 d\right]+o(b)\\ =& \frac{2bc_2 d}{d(d+2)}\left[(d+1)+1\right]+o(b)\\ =& \frac{2bc_2}{(d+2)}\left[d+2\right]+o(b)\\ =& 2bc_2+o(b).\end{aligned}$$

Next we derive $\frac{1}{2}E\left[vec'\left(\hat{\Omega}_T-\Omega\right)J_2 vec\left(\hat{\Omega}_T-\Omega\right)\right]$. Recall for a symmetric commutation matrix $\mathbb{K}_{dd}\mathbb{K}'_{dd}=\mathbb{I}_{dd}$. Utilizing this relationship, property (4.7), and property (4.9) we obtain,

$$\begin{aligned}vec'\left(\hat{\Omega}_T-\Omega\right)J_2 vec\left(\hat{\Omega}_T-\Omega\right) =& -vec'\left(\hat{\Omega}_T-\Omega\right)\left[\Omega^{-1/2}e_T e_T'\Omega^{-1/2}\otimes\Omega^{-1}\right]\mathbb{K}_{dd}\\ & \times\left(\mathbb{I}_{d^2}+\mathbb{K}_{dd}\right)vec\left(\hat{\Omega}_T-\Omega\right)\\ =& -vec'\left(\hat{\Omega}_T-\Omega\right)\left[\Omega^{-1/2}e_T e_T'\Omega^{-1/2}\otimes\Omega^{-1}\right]\mathbb{K}_{dd}vec\left(\hat{\Omega}_T-\Omega\right)\\ & -vec'\left(\hat{\Omega}_T-\Omega\right)\left[\Omega^{-1/2}e_T e_T'\Omega^{-1/2}\otimes\Omega^{-1}\right]\mathbb{I}_{dd}vec\left(\hat{\Omega}_T-\Omega\right)\\ =& -2vec'\left(\hat{\Omega}_T-\Omega\right)\left[\Omega^{-1/2}e_T e_T'\Omega^{-1/2}\otimes\Omega^{-1}\right]vec\left(\hat{\Omega}_T-\Omega\right)\\ =& -2vec'\left(\hat{\Omega}_T-\Omega\right)vec\left(\Omega^{-1}(\hat{\Omega}_T-\Omega)\Omega^{-1/2}e_T e_T'\Omega^{-1/2}\right)\\ =& -2vec'\left(\hat{\Omega}_T-\Omega\right)vec\left(\Omega^{-1/2}Ae_T e_T'\Omega^{-1/2}\right)\\ =& -2trace\left(vec\left(\Omega^{-1/2}Ae_T e_T'\Omega^{-1/2}\right)vec'\left(\hat{\Omega}_T-\Omega\right)\right)\\ =& -2trace\left(\Omega^{-1/2}Ae_T e_T'\Omega^{-1/2}(\hat{\Omega}_T-\Omega)\right).\end{aligned}$$

We can permute the arrangements of symmetric matrices within the trace operator, and use trace property (4.9) to simplify,

$$\begin{aligned}
&= -2trace\left(Ae_Te_T'\Omega^{-1/2}(\hat{\Omega}_T - \Omega)\Omega^{-1/2}\right) \\
&= -2trace\left(AXX'A\right) \\
&= X'AAX.
\end{aligned}$$

Next we find the expected value of $X'AAX$. By the tower property and Proposition 2,

$$\begin{aligned}
E[X'AAX] &= E\left[\sum_{i,j,l} A_{il}A_{lj}X_iX_j\right] \\
&= E\left[\sum_{i,j,l} A_{il}A_{lj}E[X_iX_j|A]\right] \\
&= \sum_{i,j,l} E\left[A_{il}A_{lj}\right]E\left[X_iX_j\right]
\end{aligned}$$

Recall: $E[X_iX_j] = 0$ if $i \neq j$, and $A$ is symmetric. Therefore,

$$\begin{aligned}
E[X'AAX] &= \sum_{l}\sum_{i=j} E\left[A_{il}A_{li}\right]E\left[X_i^2\right] \\
&= \sum_{l}\sum_{i=j} E\left[A_{il}A_{li}\right]\frac{1}{d} \\
&= \sum_{l}\sum_{i=j} E\left[A_{il}^2\right]\frac{1}{d} \\
&= \sum_{i=j=l} E\left[A_{ii}^2\right]\frac{1}{d} \\
&= E\left[trace(AA)\right]\frac{1}{d} \\
&= \frac{1}{d}\left[c_2b(d^2+d)\right] \\
&= c_2b(d+1).
\end{aligned}$$

Hence,

$$\begin{aligned}\frac{1}{2}E\left[vec'\left(\hat{\Omega}_T-\Omega\right)J_2vec\left(\hat{\Omega}_T-\Omega\right)\right] =&\frac{1}{2}(-2)c_2b(d+1)\\ =&-c_2b(d+1).\end{aligned}$$

We now have the components needed to obtain the expected value of $Q$,

$$\begin{aligned}E[Q] =&\frac{1}{2}E\left[vec'\left(\hat{\Omega}_T-\Omega\right)(J_1+J_2)vec\left(\hat{\Omega}_T-\Omega\right)\right]\\ \approx&2bc_2-c_2b(d+1)+o(b)\\ =&c_2b[2-d-1]+o(b)\\ =&-c_2[d-1]b+o(b).\end{aligned}$$

■

*Begin proof of Lemma 5d.*

We next derive the expected value of $L^2$,

$$\begin{aligned}E[L^2] =&E\left[Dvec\left(\hat{\Omega}_T-\Omega\right)Dvec\left(\hat{\Omega}_T-\Omega\right)\right]\\ =&E\left[Dvec\left(\hat{\Omega}_T-\Omega\right)vec'\left(\hat{\Omega}_T-\Omega\right)D'\right]\\ =&E\left[\left(e_T'\Omega^{-1/2}\otimes e_T'\Omega^{-1/2}\right)vec\left(\hat{\Omega}_T-\Omega\right)vec'\left(\hat{\Omega}_T-\Omega\right)\left(e_T'\Omega^{-1/2}\otimes e_T'\Omega^{-1/2}\right)'\right]\\ =&E\left[\left(e_T'\Omega^{-1/2}\otimes e_T'\Omega^{-1/2}\right)vec\left(\hat{\Omega}_T-\Omega\right)vec'\left(\hat{\Omega}_T-\Omega\right)\left(e_T'\Omega^{-1/2}\otimes e_T'\Omega^{-1/2}\right)'\right].\end{aligned}$$

By Proposition 2 and Theorem 1, and multiplication property in equation (4.10) we obtain,

$$\begin{aligned}
E[L^2] \approx & c_2 b E\left[\left(e_T'\Omega^{-1/2}\otimes e_T'\Omega^{-1/2}\right)\left(\mathbb{I}_{d^2}+\mathbb{K}_{dd}\right)\left(\Omega\otimes\Omega\right)\left(e_T'\Omega^{-1/2}\otimes e_T'\Omega^{-1/2}\right)'\right]+o(b)\\
=& 2c_2 b E\left[\left(e_T'\Omega^{-1/2}\otimes e_T'\Omega^{-1/2}\right)\left(\Omega\otimes\Omega\right)\left(e_T'\Omega^{-1/2}\otimes e_T'\Omega^{-1/2}\right)'\right]+o(b)\\
=& c_2 b E\left[\left(e_T'\Omega^{-1/2}\otimes e_T'\Omega^{-1/2}\right)\left(\Omega^{1/2}e_T\otimes\Omega^{1/2}e_T\right)\right]+o(b)\\
=& 2c_2 b E\left[\left(e_T'e_T\otimes e_T'e_T\right)\right]+o(b)\\
=& 2c_2 b E\left[\left(e_T'e_T\right)^2\right]+o(b).
\end{aligned}$$

Let $X = e_T$ as before, then

$$\begin{aligned}
E\left[\left(e_T'e_T\right)^2\right] =& E\left[\left(\sum_i X_i^2\right)^2\right]\\
=& E\left[\left(\sum_i\sum_j X_i^2X_j^2\right)\right]\\
=& E\left[\left(\sum_{i=j} X_i^4+\sum_{i\neq j} X_i^2X_j^2\right)\right]\\
=& d\left(\frac{3}{d(d+2}\right)+(d^2-d)\left(\frac{1}{d(d+2)}\right)\\
=& \frac{1}{d+2}\left(3+(d-1)\right)\\
=& 1.
\end{aligned}$$

Therefore, $E(L^2) \approx 2c_2 b + o(b)$ as desired.

■

The assumptions of this Lemma are required because of the reliance on Proposition 2, the expression for the variance in Theorem 1, and the expression for the bias in Assumption 6.

The proofs for Lemma 4 and 5 closely follow Sun [52, 54], except it incorporates the finite sampling bias through Assumption 6 and it is specific to SV estimators with $\kappa \in \mathscr{K}_2 \cup \mathscr{K}_3$, instead of series estimators and mother kernels. In Theorem 1 and Assumption 6 we observe that the asymptotic bias of $\hat{\Omega}_T$ attributed to the kernel can be represented through $(bT)^{-q} g_q h_q$; however, in most contexts $(bT)^{-q} g_q w_q$ comes up more regularly when finding expressions for probabilities. The value $w_q$ is dependent on the kernel only through $q$. The kernels in class $\mathscr{K}_1$ are positive-semi definite, and have guaranteed $q \leq 2$. Kernels in class $\mathscr{K}_2$ largely inherit the same $q$ as their generating mother kernel with the exception of the zero lugsail kernels that have $g_\infty = 0$. Estimating the exact $w_q$ for each scenario is not practical so practitioners often use a rudimentary but effective estimate of $w_q$ by relying on an AR(1) model with auto regressive coefficient $\rho$ [1, 27, 2]. This gives us $w_1 = \frac{2\rho}{1-\rho^2}$, and $w_2 = \frac{2\rho}{(1-\rho)^2}$. With this standard approach we can readily obtain approximations for $(bT)^{-q} g_q w_q$ for all kernels $\kappa \in \mathscr{K}_2 \cup \mathscr{K}_3$.

**Theorem 3:** *Let $\kappa \in \mathscr{K}_2 \cup \mathscr{K}_3$ and assume that $bT \to \infty$ as $T \to \infty$, but at a slower rate. Under Assumptions 4, 5, and 6,*

*(a)* $P(dF_{OLS} \leq z) = G_d(z) + G_d''(z) z^2 c_2 b - G_d'(z) z [c_1 + c_2(d-1)] b$

$- G_d'(z) z c_b - (bT)^{-q} G_d'(z) z g_q w_q + o(b) + o\left((bT)^{-q}\right)$

*(b)* $P_\delta(dF_{OLS} \leq z) = G_{d,\delta^2}(z) + G_{d,\delta^2}''(z) z^2 c_2 b - G_{d,\delta^2}'(z) z [c_1 + c_2(d-1)] b$

$- G_{d,\delta^2}'(z) z c_b - (bT)^{-q} G_{d,\delta^2}'(z) z g_q w_q + o(b) + o\left((bT)^{-q}\right).$

*Begin proof of Theorem 3a.*

Use similar steps as in [54]. Using Lemma 4 and Lemma 5,

$$\begin{aligned} P(dF_{OLS} \leq z) =& E\left[G_d\left(z\Xi_T^{-1}\right)\right] \\ \approx& E\left[G_d\left(z(1+L+Q)\right)\right] + remainder. \end{aligned}$$

We next take the Taylor Series expansion of $G_d\left(z(1+L+Q)\right)$ around $G_d(z)$,

$$\begin{aligned} P(dF_{OLS} \leq z) \approx& E\left\{G_d(z) + \frac{G_d(z)}{1!}z(L+Q) + \frac{G_d''(z)}{2}z^2(L+Q)^2\right\} + remainder \\ =& G_d(z) + \frac{G_d(z)}{1!}zE(L+Q) + \frac{G_d''(z)}{2}z^2E\left((L+Q)^2\right) + remainder \\ \approx& G_d(z) + \frac{G_d(z)}{1!}zE(L+Q) + \frac{G_d''(z)}{2}z^2E\left(L^2\right) + remainder. \end{aligned}$$

We next use the results of Lemma 5. We further observe that the derivatives of $G_d'(x)$ and $G_d''(x)$ are bounded, and use a similar argument from Lemma 4 to simplify,

$$\begin{aligned} P(dF_{OLS} \leq z) \approx& G_d(z) + \frac{G_d(z)}{1!}z\left(-(bT)^q g_q w_q - c_1 b - c_b - bc_2(d-1)\right) + \frac{G_d''(z)}{2}z^2\left(2c_2 b\right) \\ &+ o(b) + o\left((bT)^{-q}\right) \\ =& G_d(z) + G_d^{''}(z)z^2 bc_2 - G_d'(z)z[c_b + c_1 b + c_2(d-1)b] - (bT)^{-q}G_d'(z)g_q w_q \\ &+ o(b) + o\left((bT)^{-q}\right). \end{aligned}$$

■

*Proof of Theorem 3b.*

Define

$$\begin{aligned} \underset{(d\times 1)}{e_\delta} :=& \frac{\Omega_{GLS}^{-1/2}\left[R\sqrt{T}(\hat{\theta}_{GLS} - \theta_0)\right] + \delta}{||\Omega_{GLS}^{-1/2}\left[R\sqrt{T}(\hat{\theta}_{GLS} - \theta_0)\right] + \delta||} \\ \underset{(1\times 1)}{\Upsilon_\delta} :=& ||\Omega_{GLS}\left[R\sqrt{T}(\hat{\theta}_{GLS} - \theta_0)\right] + \delta||^2 \\ \underset{(1\times 1)}{\Xi_\delta} :=& e_\delta'\Omega_{GLS}^{1/2}\hat{\Omega}_T^{-1}\Omega_{GLS}^{1/2}e_\delta. \end{aligned}$$

Using similar arguments as in Lemma 4, under $H_A$ with non-centrality parameter $\delta^2$

$$dF_{GLS} = \Upsilon_\delta \Xi_\delta$$

which implies

$$\begin{aligned} P_\delta\left(dF_{OLS} \leq z\right) =& P_\delta\left(\Upsilon_\delta \Xi_\delta \leq z\right) + O(T^{-1}) \\ =& E G_{d,\delta^2}\left(z \Xi_T^{-1}\right) + O(T^{-1}) \end{aligned}$$

for some $z \in \mathbb{R}$. Use similar steps as in Lemma 5 and Theorem 3a,

$$\begin{aligned} P_\delta(dF_{OLS} \leq z) =& G_{d,\delta^2}(z) + G''_{d,\delta^2}(z)\, b(z)^2\, c_2 \\ & - G'_{d,\delta^2}(z) z [c_b + c_1 b + c_2 (d-1) b] \\ & - (bT)^{-q} G'_{d,\delta^2}(z)\, z g_q w_q + o(b) + o\left((bT)^{-q}\right). \end{aligned}$$

■

The first three terms of Theorem 3a are the same as the first terms as the fixed-$b$ random variable $F_\infty$ in Theorem 2. These terms capture the variability of $\hat{\Omega}_T$, but also contain the de-meaning bias term $c_1 b$ which is not dependent on sample size. The discrepancy between Theorem 3a and Theorem 2 comes from the bias components that are specific to the sample size, i.e. the dominate kernel bias term and the finite samping bias. The expression in Theorem 3b largely shares the same general properties.

# Chapter 5

# Optimal Bandwidth for the Lugsail Estimator

## 5.1 Alternative Loss Functions and Bandwidth Rules

Classical inference procedures often have the strict but common set of assumptions that the error terms are approximately normal, iid, and homoskedastic. In contrast robust inference procedures utilize SV estimators which are designed to handle a wider variety of error structures that may not meet these assumptions. The robust SV estimator depends on two tuning parameters, a bandwidth and a kernel function. Research has largely focused on optimizing the bandwidth.

When searching for a bandwidth for robust estimation the goal is to select $b$ such that for all $|s| > bT$ we have that $\Gamma(s) \approx 0$ [49], which we refer to as the fundamental principal of bandwidth selection. Practitioners can arbitrarily pick such a $b$ for each scenario

by studying the autocovariance function, but this is usually highly subjective, variable, and tedious. Hence, bandwidth rules emerged that are data driven and adaptive. These rules allow practitioners to automatically procure a reasonable $b$ without much discernment. Politis [47] proposed a such a rule for the class of flat top kernels with the fundamental principle in mind. Let $\hat{\rho}(s) = \hat{\Gamma}(s)/\hat{\Gamma}(0)$,

$$m = \underset{m \in \mathbb{N}^+}{\operatorname{argmin}} \left\{ \hat{\rho}(\hat{m}+k) < c\sqrt{\log(T)/T} \text{ for } k = 1, \ldots, K_T \right\},$$

$c \in \mathbb{R}^+$, $K_T = O(\log(T))$, and $1 \leq K_T \leq T$. For a Bartlett flat top kernel it is recommended to use $c = 2$, $K_T = \max(5, \lfloor \log(T) \rfloor)$, which results in the bandwidth rule

$$b_{ft} = \frac{2m}{T}. \tag{5.1}$$

A limitation of this rule is that it will never attain zero, even in the presence of no correlation. It is also highly variable and computationally burdensome as it relies on the estimation of several random components.

Most other bandwidth rules are motivated by loss functions, and historically MSE has been a popular loss function for picking an optimal bandwidth for LRV estimators. Andrews [1] proposed a rule based on AMSE that has been commonly used; for the Bartlett mother kernel with VAR(1) random errors and correlation coefficient matrix $\rho \mathbf{1}_d$ we can simplify this rule to

$$b_{mse} = 1.1447 \left( \frac{2\hat{\rho}}{(1-\hat{\rho})^2} \frac{1}{T} \right)^{2/3}. \tag{5.2}$$

The AMSE rule is intuitive, relies on classical statistical principals and is easy to use. However, it accounts for correlation in only how it relates to the asymptotic kernel bias. It also induces a paradox for kernels with infinite characteristic components: the typical

dominate asymptotic bias term is zero, thus we optimize according to asymptotic variance alone and obtain $b_{mse} = 0$, resulting in a clearly non-optimal biased estimate.

Although practical, bandwidth rules often rely on proxy objectives, and in the quest of generality miss nuances of individual finite samples. Recent efforts have been made on loss functions and rules designed around the prime objective of the estimator, inference. To achieve statistical validity it is well known that inference procedures should have an observed Type 1 error rate less than or equal to $\alpha$. Most these rules focus on this goal, with only minimal consideration of power. Let $\tau \in [0, 1]$, $e_I(b)$ denote the Type 1 error rate, $e_{II}(b)$ denote the Type 2 error rate, $\Delta_s = e_I(b) - \alpha$ denote the size distortion, $\Delta_p^{max}$ denote the maximum size-adjusted power loss, and $\Delta_p^I = \int \Delta_p(\delta) d \prod_\delta(\delta)$ be the integrated density function over the non-centrality parameter $\delta$. As summarized by Lazarus et al. [27], following testing centered rules have been considered for kernels in class $\mathscr{K}_1$,

(a) $\tau|\Delta_s| + (1-\tau)\Delta_p^{max}$

(b) $\tau|\Delta_s| + (1-\tau)\Delta_p^{I}$

(c) $\tau\Delta_s^2 + (1-\tau)\left(\Delta_p^{max}\right)^2$

(d) $\tau e_I(b) + (1-\tau)e_{II}(b)$

(e) $\underset{b}{\text{argmax}}\,(1 - e_{II}(b))$ such that $e_I(b) \leq \alpha(1+\tau)$.

The loss functions above require a tuning parameter $\tau$, with little guidance on how to select $\tau$ except for the general suggestion that $\tau \approx 0.9$ for functions (a)-(d), and $\tau \approx 0.1$ for (e) to emphasize the importance of Type 1 error [27, 29, 54]. The loss functions (a), (b), and (d) can be thought of as a weighted average of bias and variance due to size distortion

being dominated by the bias of $\hat{\Omega}_T$, and power being heavily influenced by variability of $\hat{\Omega}_T$ [27, 56]. The function $(c)$ instead resembles a modified version of MSE, because it can be viewed as a weighted average of bias squared and variance. The loss function (e) yields a different bandwidth depending on the direction of correlation, accounting for the fact that a negatively correlated data set generally produces a less biased estimate for LRV [54]. Coverage probability error has also been considered for series estimators [53].

We follow a spirit similar to the loss function (e) proposed by Sun [54], and the general construction of Neyman-Pearson like testing procedures. We also account for finite sampling bias as a way to incorporate the fundamental principle.

## 5.2 Approximating Error Rates

### 5.2.1 Finite Sample Bias

To obtain an expression for Type 1 and Type 2 error we must approximate $c_b$, and we use an auto-regressive model of order 1. Consider the Gaussian location model in equation (4.2), let $\epsilon_t \overset{iid}{\sim} N_p(0, \Omega_\epsilon)$ and $A$ be a known absolutely summable non-random matrix. We can model the error terms $w_t$ as

$$w_t = Aw_{t-1} + \epsilon_t$$

and refer to $\epsilon_t$ as the disturbances. With this construction we can calculate the true LRV for the random variable $y_t$, whose variability is contained in the error $w_t$ [34, 62]. We start by deriving $\Gamma(h)$,

$$\begin{aligned}
\Gamma(h) =& Cov(w_t, w_{t-h}) \\
=& E\left[w_t w_{t-h}'\right] \\
=& E\left[(Aw_{t-1} + \epsilon_{t-1})w_{t-h}'\right] \\
=& E\left[Aw_{t-1}w_{t-h}'\right] + E\left[\epsilon_{t-1}w_{t-h}'\right].
\end{aligned}$$

This helps us recognize that

$$\Gamma(h) = \begin{cases} A\Gamma(-1) + \Omega_\epsilon & \text{if } h = 0 \\ A\Gamma(h-1) & \text{if } h > 0. \end{cases} \tag{5.3}$$

The equations (5.3) are known as the Yule-Walker equations. With this information we recognize that if we know $\Gamma(0)$, $A$, and $\Omega_\epsilon$ then $\Gamma(h)$ can be solved recursively for all $h$. As a first step we must find $\Gamma(0)$ utilizing $\Omega_\epsilon$ and $A$,

$$\begin{aligned}
\Gamma(0) =& A\Gamma(1)' + \Omega_\epsilon \\
=& A\Gamma(0)A' + \Omega_\epsilon \\
vec(\Gamma(0)) =& vec\left[A\Gamma(0)A' + \Omega_\epsilon\right] \\
=& (A \otimes A)vec(\Gamma(0))) + vec(\Omega_\epsilon) \\
vec(\Gamma(0))(I - A \otimes A) =& vec(\Omega_\epsilon) \\
vec(\Gamma(0)) =& (I - A \otimes A)^{-1} vec(\Omega_\epsilon).
\end{aligned}$$

We now derive the the true value of $\Omega$,

$$\begin{aligned}
\Omega =& \sum_{j=-\infty}^{\infty} \Gamma(j) \\
=& \Gamma(0) + \sum_{j=1}^{\infty} R(j) + \sum_{j=1}^{\infty} \Gamma(j)' \\
=& \Gamma(0) + \sum_{j=1}^{\infty} A^j \Gamma(0) + \sum_{j=1}^{\infty} \left( A_1^j \Gamma(0) \right)' \\
=& \sum_{j=0}^{\infty} A^j \Gamma(0) + \sum_{j=0}^{\infty} \left( A^j \Gamma(0) \right)' - \Gamma(0) \\
=& (I - A)^{-1} \Gamma(0) + \Gamma(0)(I - A')^{-1} - \Gamma(0).
\end{aligned}$$

If $A = \rho \mathbf{1}_p$ for $|\rho| < 1$, then $\Omega_{ii} = \frac{1}{(1-\rho)^2} [\Omega_\epsilon]_{i,i}$. Concentrating on one diagonal element of the LRV we obtain an estimate for the finite sampling bias,

$$\begin{aligned}
\Omega_{ii} c_b \approx & \sum_{|s| > \lfloor bT \rfloor} \Gamma_{ii}(s) \\
=& \sum_{s=\infty}^{\infty} \Gamma_{ii}(s) - \sum_{s=-\lfloor bT \rfloor}^{\lfloor bT \rfloor} \Gamma_{ii}(s) \\
=& 2 \sum_{s=\lfloor bT \rfloor + 1}^{\infty} \Gamma_{ii}(s)
\end{aligned}$$

We plug in $\Gamma_{ii}(s)$ by using the Yule-Walker equations, use the geometric series identity $\sum_{k=n}^{\infty} r^k = \frac{r}{1-r} r^n$, and substitution with $\Omega_{ii} = \frac{1}{(1-\rho)^2} [\Omega_\epsilon]_{i,i}$,

$$
\begin{aligned}
\Omega_{ii} c_b \approx & 2\Gamma_{ii}(0) \left( \sum_{s=\lfloor bT \rfloor + 1}^{\infty} \rho^s \right) \\
= & 2\Gamma_{ii}(0) \left( \frac{\rho}{1-\rho} \rho^{\lfloor bT \rfloor + 1} \right) \\
= & 2 \frac{1}{(1-\rho)(1+\rho)} [\Omega_\epsilon]_{i,i} \frac{\rho}{1-\rho} \rho^{\lfloor bT \rfloor + 1} \\
= & \Omega_{ii} \frac{2\rho}{1+\rho} \rho^{\lfloor bT \rfloor + 1} \\
= & \Omega_{ii} \frac{2\rho^2}{1+\rho} \rho^{\lfloor bT \rfloor}.
\end{aligned}
$$

This leads us to the following proposition.

**Proposition 3:** *Under Assumption 6, we approximate the finite sampling bias as*

$$
c_b \approx \frac{2\hat{\rho}^2}{1+\hat{\rho}} \hat{\rho}^{bT}. \tag{5.4}
$$

### 5.2.2 Asymptotic Error Rates

With Theorem 2, Theorem 3, and Proposition 3 we can derive accessible expressions for Type 1 and Type 2 error when utilizing fixed-$b$ critical values that incorporate the finite sampling bias.

**Corollary 2:** *Let $\tilde{\kappa} \in \mathscr{K}_3$ and assume that $bT \to \infty$ as $T \to \infty$, but at a slower rate. Using test statistic $dF_{OLS}$ and fixed-b critical values, under Assumptions 4, 5, and 6,*

*(a)* $e_I(b) \approx \alpha + G'_d(\chi_d^\alpha)\chi_d^\alpha \frac{2\rho^2}{1+\rho}\rho^{bT} + (bT)^{-q} G'_d(\chi_d^\alpha)\chi_d^\alpha g_q w_q$

*(b)* $1 - e_{II}(b) \approx 1 - G_{d,\delta^2}(\chi_d^\alpha) - \left[ \frac{\delta^2}{2} G'_{(d+2),\delta^2}(\chi_d^\alpha) - G'_{d,\delta^2}(\chi_d^\alpha)(d-1) \right] \chi_d^\alpha c_2 b$

$+ G'_{d,\delta^2}(\chi_d^\alpha)\chi_d^\alpha \frac{2\rho^2}{1+\rho}\rho^{bT} + G'_{d,\delta^2}(\chi_d^\alpha)\chi_d^\alpha c_1 b + (bT)^{-q} G'_{d,\delta^2}(\chi_d^\alpha)\chi_d^\alpha g_q w_q.$

*Begin proof of Corollary 2a.*

We begin with the Type 1 error expression. By Theorem 2,

$$\begin{aligned}1 - P(dF_\infty \leq z) =& 1 - G_d(z) - G_d^{''}(z)\,(z)^2\, c_2 b + G_d(z) z[c_1 + c_2(d-1)]b + o(b) \\ =& 1 - G_d(z) - G_d^{''}(z)\,(z)^2\, c_2 b + G_d(z) z[c_1 + c_2(d-1)]b + o(b) \\ \overset{set}{=}& \alpha + o(b).\end{aligned}$$

By Theorem 3,

$$\begin{aligned}1 - P(dF_{T,OLS} \leq z) =& 1 - G_d(z) - G_d^{''}(z)\,(z)^2\, c_2 b + G_d'(z) z[c_b + c_1 + c_2(d-1)b] \\ & - (bT)^{-q} G_d'(z) z g_q w_q + o\left((bT)^{-q}\right) \\ =& 1 - G_d(z) - G_d^{''}(z)\,(z)^2\, c_2 b + G_d'(z) z[c_1 + c_2(p-1)]b + G_d'(z) z[c_b] \\ & - (bT)^{-q} G_d'(z) z g_q w_q + o\left((bT)^{-q}\right).\end{aligned}$$

Recognizing components from Theorem 2 we can rewrite the above expression,

$$1 - P(dF_{T,OLS} \leq z) = \alpha + G_d'(z) z[c_b] - (bT)^{-q} G_d'(z) z g_q w_q + o(b) + o\left((bT)^{-q}\right).$$

Observing $dc_d^\alpha(b) \approx \chi_d^\alpha + o(b)$, and using the approximation in Proposition 3 we have,

$$e_I(b) \approx \alpha + G_d'(\chi_d^\alpha)\chi_d^\alpha \frac{2\rho^2}{1+\rho}\rho^{bT} + (bT)^{-q} G_d'(\chi_d^\alpha)\chi_d^\alpha g_q w_q.$$

■

*Begin proof of Corollary 2b.*

Let $z = dc_d^\alpha(b)$ for ease of notation. Recall from Theorem 3,

$$\begin{aligned}P_\delta(dF_{OLS} \leq z) =& G_{d,\delta^2}(z) + G_{d,\delta^2}^{''}(z)\,(z)^2\, c_2 \\ & - G_{d,\delta^2}'(z) z[c_b + c_1 + c_2(d-1)b] \\ & - (bT)^{-q} G_{d,\delta^2}'(z)\, z g_q w_q + o(b) + o\left((bT)^{-q}\right). \qquad (5.5)\end{aligned}$$

We wish to rewrite this expression so that it has more accessible terms. We will utilize a Cornish-Fisher type expansion as in Sun [54, Theorem 4] and Lazarus et al. [27, Theorem 1]. Cornish-Fisher expansions are used to approximate quantiles of a random variable using the first few cumulants of the Gaussian distribution [15, 31], provided the random variable being approximated has a Gaussian limiting distribution. Cornish-Fisher *type* expansions broaden this result to include non-Gaussian limiting distributions [13]. For our purposes we use the $\chi^2$ limiting distribution to approximate the fixed-$b$ critical value $z$ as $b \to 0$,

$$\begin{aligned} z = dc_d^\alpha(b) \approx & \chi_d^\alpha - \frac{G_d^{''}\left(\chi_d^\alpha\right)}{G_d^{'}\left(\chi_d^\alpha\right)}\left(\frac{\gamma_1}{6}\right) \\ \approx & \chi_d^\alpha - \frac{G_d^{''}\left(\chi_d^\alpha\right)}{G_d^{'}\left(\chi_d^\alpha\right)}\left(c_2\left(\chi_d^\alpha\right)^2\right) b + O(b) \end{aligned}$$

where $\gamma_1$ is the skewness of the $\mathscr{F}_\infty$ random variable. As already mentioned, $d\mathscr{F}_\infty(b,d) \to \chi_d^2$; the Cornish-Fisher type expansion above gives a structure to this approximation that helps us rearrange terms. We plug in the expansion into the first term of (5.5) and further approximate with a Taylor series expansion around $\chi_d^2$,

$$\begin{aligned} G_{d,\delta^2}(z) \approx & G_{d,\delta^2}\left(\chi_d^\alpha - \frac{G_d^{''}(\chi_d^\alpha)c_2\left(\chi_d^\alpha\right)^2}{G_d^{'}(\chi_d^\alpha)} b\right) \\ \approx & G_{d,\delta^2}(\chi_d^\alpha) - G_{d,\delta^2}^{'}(\chi_d^\alpha)\left(\frac{G_d^{''}(\chi_d^\alpha)c_2\left(\chi_d^\alpha\right)^2}{G_d^{'}(\chi_d^\alpha)}\right) b + O(b). \end{aligned}$$

We plug this new representation back into (5.5),

$$\begin{aligned} P_\delta(dF_{OLS} \le z) = & G_{d,\delta^2}(\chi_d^\alpha) - G_{d,\delta^2}^{'}(\chi_d^\alpha)\left(\frac{G_d^{''}(\chi_d^\alpha)c_2\left(\chi_d^\alpha\right)^2}{G_d^{'}(\chi_d^\alpha)}\right) b \\ & + G_{d,\delta^2}^{''}(z)(z)^2 c_2 - G_{d,\delta^2}^{\prime}(z) z[c_b + c_1 + c_2(d-1)b] \\ & - (bT)^{-q} G_{d,\delta^2}^{\prime}(z)\, z g_q w_q + O(b) + o\left((bT)^{-q}\right). \end{aligned}$$

We further substitute $z = dc_d^\alpha(b) \approx \chi_d^\alpha + O(b)$ to rearrange terms,

$$\begin{aligned} P_\delta(dF_{OLS} \leq z) \approx & G_{d,\delta^2}(\chi_d^\alpha) - G'_{d,\delta^2}(\chi_d^\alpha)\left(\frac{G''_d(\chi_d^\alpha)c_2\,(\chi_d^\alpha)^2}{G'_d(\chi_d^\alpha)}\right) b \\ & + G''_{d,\delta^2}(\chi_d^\alpha)(\chi_d^\alpha)^2 c_2 - G'_{d,\delta^2}(\chi_d^\alpha)\chi_d^\alpha[c_b + c_2(d-1)b] \\ & - (bT)^{-q} G'_{d,\delta^2}(\chi_d^\alpha)\chi_d^\alpha g_q w_q + O(b) + o\left((bT)^{-q}\right). \end{aligned}$$

Using the result from Sun [54, Theorem 5],

$$G''_{d,\delta^2}(\chi_d^\alpha) - G'_{d,\delta^2}(\chi_d^\alpha)\left(\frac{G''_d(\chi_d^\alpha)}{G'_d(\chi_d^\alpha)}\right) = \frac{\delta^2}{2\chi_d^\alpha} G'_{(d+2),\delta^2}(\chi_d^\alpha),$$

we can further simplify terms

$$\begin{aligned} P_\delta(dF_{OLS} \leq dc_d^\alpha(b)) \approx & G_{d,\delta^2}(\chi_d^\alpha) + \frac{\delta^2}{2} G'_{(d+2),\delta^2}(\chi_d^\alpha)\chi_d^\alpha c_2 b \\ & - G'_{d,\delta^2}(\chi_d^\alpha)\chi_d^\alpha[c_b + c_1 + c_2(d-1)b] \\ & - (bT)^{-q} G'_{d,\delta^2}(\chi_d^\alpha)\chi_d^\alpha g_q w_q + O(b) + o\left((bT)^{-q}\right). \end{aligned}$$

We use Proposition 3 to reach our final expression:

$$\begin{aligned} e_{II}(b) \approx & G_{d,\delta^2}(\chi_d^\alpha) + \frac{\delta^2}{2} G'_{(d+2),\delta^2}(\chi_d^\alpha)\chi_d^\alpha c_2 b - G'_{d,\delta^2}(\chi_d^\alpha)\chi_d^\alpha \frac{2\rho^2}{1+\rho}\rho^{bT} \\ & - G'_{d,\delta^2}(\chi_d^\alpha)\chi_d^\alpha[c_1 + c_2(d-1)b] - (bT)^{-q} G'_{d,\delta^2}(\chi_d^\alpha)\chi_d^\alpha g_q w_q. \end{aligned}$$

■

Expressions in Corollary 2 are similar to [27, 29, 54]. We obtain the expression in Corollary 2a by recognizing the behavior that a fixed-$b$ critical value captures for a finite data set. That is, the Type 1 error rate and the distortion thereof. The distortion arises

from the kernel bias and the finite sample bias, the only bias terms in Assumption 6 that explicitly disappear as the sample size goes to infinity.

In Corollary 2b we essentially obtain another representation for Theorem 3b. Here we applied a Cornish-Fisher type expansion, a Taylor series approximation, an asymptotic representation of a fixed-$b$ critical value, and a distribution approximation in order to reconstruct the expression in Theorem 3b to have more practical terms to better understand the behavior of $F_T$ under the alternative hypothesis. We observe that the finite sampling bias and kernel bias decrease power as $b$ increases, as expected. The term with coefficient $c_2$ increases or decreases power as a function $b$ depending on the context. This term largely captures the asymptotic distributional effects of $\hat{\Omega}_T$ on the inference procedure. Lastly, the de-meaning bias term increases power as a function of $b$, an expected property that has been shown in other contexts [10].

As alluded by the testing optimal loss functions in the previous sections, quantifying Type 2 error (and power) in a robust inference setting is not as straight forward as size distortion. In general, we expect overall power to decrease as variability increases. This is because distributions under the null and alternative are difficult to distinguish if the tails are very heavy. In contrast, the negative bias of the LRV has the opposite effect on power as it does for Type 1 error in that a negative bias improves power.

To further illustrate these effects consider the following scenario motivated by a thought experiment from Rubin [51]. Suppose you generate a confidence interval for a univariate data set using two different approaches. Confidence bounds are generated via the formula $\hat{\theta} \pm \left(\frac{\Omega}{T}\right)^{1/2} z_\alpha$, where we refer to $\left(\frac{\Omega}{T}\right)^{1/2} z_\alpha$ as the margin of error. Suppose for

procedure $A$ that $\Omega$ is known (the oracle test), thus the $z_\alpha$ only needs to account for the distribution of $\hat{\theta}$ and the margin of error is the same each time the study is repeated.

Further suppose for procedure $B$ that $\Omega$ is unknown and estimated with a moderately variable and negatively biased SV estimator $\hat{\Omega}_T$, and that the bandwidth is given. SV estimators are mean invariant, so for both procedures shifts along the real line are attributed to the estimated mean and not $\Omega$. The negative bias indicates that $E\left(\hat{\Omega}_T\right)$ is less than $\Omega$, thus the confidence intervals generated under procedure $B$ are on average narrower than procedure $A$. We further note that the SV estimator $\hat{\Omega}_T$ has a positively skewed distribution (see Chapter 3 for connections to $\chi^2$ distribution). Because of this asymmetry the majority of estimates will be smaller than $E\left(\hat{\Omega}_T\right)$, with outliers being much larger. This is the cascading effect of SV estimators. The asymmetric density indicates the majority of estimates will be smaller than $E(\hat{\Omega}_T)$, which is already less than the true value $\Omega$. Hence, if procedures $B$ uses the $z_\alpha$ intended for procedure $A$ then the confidence intervals for procedure $B$ will typically be much narrower than procedure $A$ because the bias and distributional shape of $\hat{\Omega}_T$. When creating confidence intervals the role of the term $z_\alpha$ is to capture the distributional properties of the random elements in the procedure. The $z_\alpha$ from procedure $A$ only accounts for the randomness of estimating $\theta$ and not $\Omega$.

Previously, we have said that the fixed-$b$ distribution accounts for the variability of $\hat{\Omega}_T$, but that is a bit simplistic. Further suppose that procedure $B$ now has a different LRV estimator that is infeasibly symmetrically distributed, but with the same bias and variance as before. Then the performance of procedure $B$ would become closer to the performance

of procedure $A$, despite the bias and variance being the same. It is more accurate to say the fixed-$b$ distribution accounts for the distribution of $\hat{\Omega}_T$, instead of simply saying variability.

It was previously mentioned that there are two types of corrections that can be made when there is a discrepancy between an inference procedure and limiting distribution: adjustments to estimates so they converge faster to the limiting distribution, and selecting a different limiting distribution that captures more nuanced finite sample behavior [3]. When we use fixed-$b$ limiting theory for procedure $B$ we account for the asymptotic distributional properties of estimating both $\theta$ and $\Omega$. However, the confidence intervals under procedure $B$ are still expected to have sub-optimal Type 1 error because $\hat{\Omega}_T$ is negatively biased. Using the lugsail estimator addresses the issue of bias. These effects are observed in Corollary 2a, where we see the higher level terms from Theorem 3a disappear when utilizing fixed-b critical values, but terms intrinsic to bias persist. The lugsail estimator has a zero asymptotic kernel bias which further decreases the size distortion in comparison to the classical mother kernels.

For our thought experiment the only way to decrease power was to increase the sample size, as we were given a bandwidth. However, the ability to control sample size is limited in real world applications, so instead we rely on our choices for the bandwidth and kernel function to control testing behavior. As discussed throughout, and illustrated in Theorem 1, the bandwidth is directly related to the behavior of $\hat{\Omega}_T$. Therefore, it is reasonable to use different bandwidths between the different lugsail settings. Although in an ideal setting we would select a bandwidth to minimize the size distortion, i.e. $b = 1$, the overall variability would suffer and the distribution under the null and alternative hypothesis would be nearly indistinguishable for most practical cases under the alternative

hypothesis. To control testing mechanics it is necessary to obtain a parsimonious estimator that incorporates as many relevant autocovariance matrices but no more.

We observe that the Type 1 error rate converges to $\alpha$ as the bandwidth increases. We further observe that the Type 2 error rate is a concave function. If we pick $b$ in order to optimize the Type 1 error then this always results in a bandwidth of 1, which is impractical. When picking $b$ to minimize Type 2 error we obtain a bandwidth of 0 or 1, depending on the context.

Instead we select $b$ a where Type 1 error rate begins to stabilize. That is, when the improvement on Type 1 error from increasing $b$ becomes increasingly marginal. When the Type 1 error stabilizes, increasing $b$ is largely decreasing power with minimal effect on the Type 1 error rate.

## 5.3 Zero-Lugsail Bandwidth Rule

We choose the smallest bandwidth where the Type 1 error is in the neighborhood of $\alpha$. The optimization rule can be summarized as follows,

$$b_{opt} = \min \left\{ b \in [0,1] : e_I(b) \leq \alpha + \tau \right\}. \tag{5.6}$$

For zero lugsail kernels simplifies to

$$b_{opt} = min \left\{ b \in [0,1] : b \geq \log \left( \frac{\tau}{G_d'(\chi_d^\alpha)\chi_d^\alpha} \frac{1+\rho}{2\rho^2} \right) \frac{1}{\log(\rho)} \frac{1}{T} \right\}. \tag{5.7}$$

The curve of the Type 1 error rate in Corollary 2a as a function of $b$ in the presence of correlation results in a shape of an 'L' or an arm, which is illustrated in Figure 5.1. This

arm like shape is commonly observed in loss functions, and the general consensus is to select a tuning parameter that is located near the 'elbow' (the bend of the curve), or just past it [19]. Observe that the curve converges to origin as the correlation decreases or sample size increases. It also shifts both horizontally and vertically as the significance level changes, generally to the origin as $\alpha$ decreases.

If we select an arbitrary constant $\tau \in \mathbb{R}^+$ we risk the possibility that the Type 1 error curve never falls in the neighborhood, or obtaining a bandwidth parameter that is not 'past the elbow' and too far 'up the arm' in the wrong direction. We instead use

$$\tau = -\frac{\alpha^{1/(2d)}}{T\log(\rho)}, \tag{5.8}$$

to select a bandwidth where the the Type 1 error rate begins to stabilize. The $\tau$ in equation (5.8) is related to the derivative of the Type 1 error rate. When the Type 1 error rate begins to stabilize, the derivative of the Type 1 error rate also begins to stabilize and becomes increasing close to 0. Instead of finding a bandwidth that minimizes Type 1 error rate, i.e. where the derivative is 0, we are picking a point where the derivative is *near* 0. By picking $\tau$ according to (5.8) we are picking the smallest bandwidth $b$ at the point at which the Type 1 error rate becomes mostly flat instead of mostly vertical regardless of the strength of the correlation, sample size, or significance level. We get this effect for any $\tau_0 \in [0, 1]$ such that $\tau = -\frac{\tau_0}{T\log(\rho)}$, with values closer to zero having a farther clearance from the elbow. We define the 'elbow' as the point where $t_0 = 1$. However, using the value $\tau_0 = \alpha^{1/(2d)}$ creates a small clearance from the elbow that naturally tightens for higher significance levels and accounts for the dimensionality bias which greatly effects the behavior of the testing procedure. For example, with $\alpha = 0.05$ and $d = 1$, the bandwidth

rule selected is at the point were a 1% increase in $b$ is associated with a 0.22% decrease in Type 1 error rate. Meaning we observe an increasingly marginal decrease in error compared to the increase in bandwidth. With this rule it is possible to obtain bandwidths of 0 when the data set has low correlation and a large sample size.

Figure 5.1 shows the asymptotic Type 1 error rate under various sample sizes, significance levels, and correlations. The bandwidth $b_{opt}$ in equation (5.7) is indicated with a circle, and adjusts to the curve as expected. Highly correlated data is expected to digress from the origin, and thus the $b_{opt}$ rule is larger in these settings. Similarly, smaller data sets also digress from the origin.

| $\rho \setminus \alpha$ | 0.100 | 0.050 | 0.025 | 0.010 |
|---|---|---|---|---|
| 0.15 | 0.0055 | 0.0053 | 0.0051 | 0.0045 |
| 0.20 | 0.0076 | 0.0074 | 0.0071 | 0.0065 |
| 0.25 | 0.0097 | 0.0096 | 0.0092 | 0.0084 |
| 0.30 | 0.0120 | 0.0118 | 0.0113 | 0.0105 |
| 0.35 | 0.0144 | 0.0142 | 0.0136 | 0.0126 |
| 0.40 | 0.0170 | 0.0167 | 0.0161 | 0.0150 |
| 0.45 | 0.0199 | 0.0196 | 0.0189 | 0.0176 |
| 0.50 | 0.0232 | 0.0228 | 0.0220 | 0.0205 |
| 0.55 | 0.0269 | 0.0265 | 0.0256 | 0.0239 |
| 0.60 | 0.0314 | 0.0309 | 0.0298 | 0.0278 |
| 0.65 | 0.0367 | 0.0362 | 0.0349 | 0.0325 |
| 0.70 | 0.0434 | 0.0427 | 0.0412 | 0.0383 |
| 0.75 | 0.0519 | 0.0511 | 0.0492 | 0.0456 |
| 0.80 | 0.0635 | 0.0625 | 0.0600 | 0.0553 |
| 0.85 | 0.0803 | 0.0789 | 0.0755 | 0.0691 |
| 0.90 | 0.1075 | 0.1053 | 0.1001 | 0.0902 |

Table 5.1: Various $b_{opt}$ Values for the Zero Lugsail Kernel with $T = 200$.

Table 5.1 lists values for $b_{opt}$ according to (5.7) for a sample size of $T = 200$ under various settings. We note that the optimal bandwidth decreases as $\alpha$ decreases despite

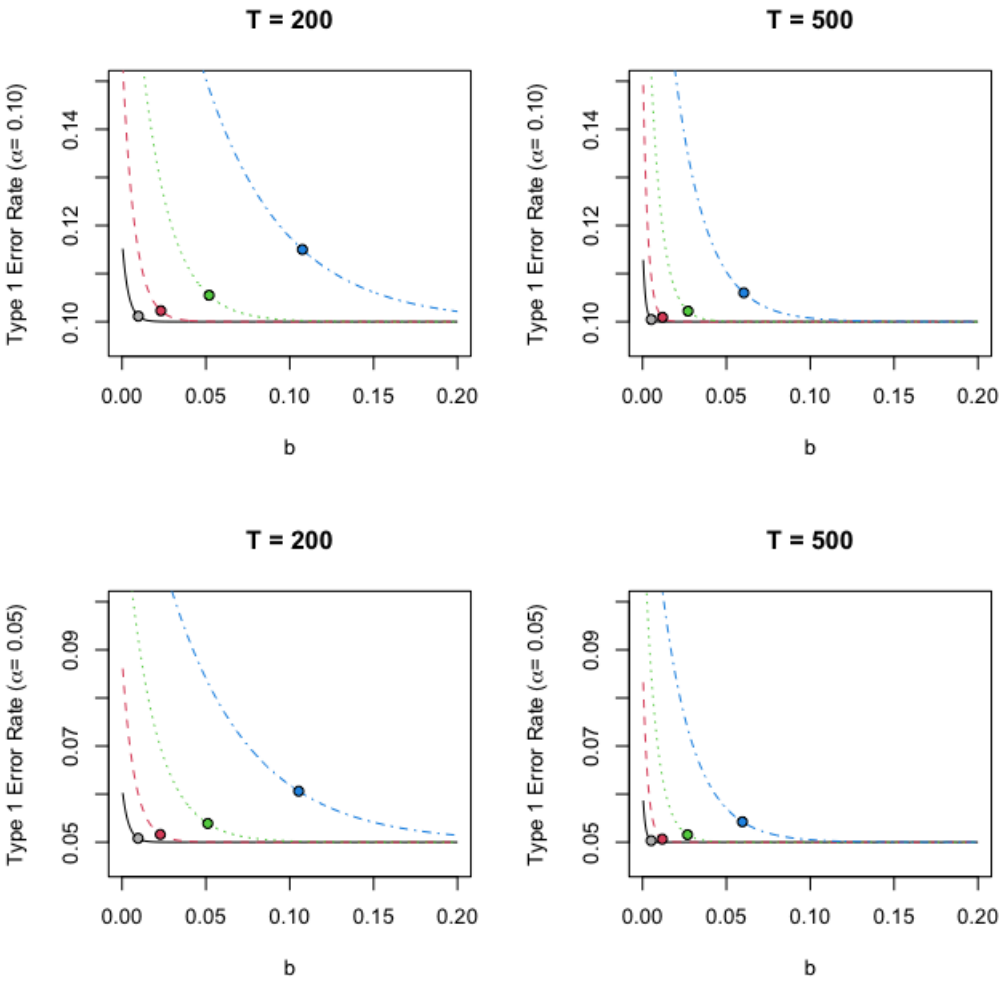


Figure 5.1: Expected Type 1 error rate of zero lugsail kernels as a function of $b$ for $d = 1$, $\alpha = (0.05, 0.01)$, $T = (200, 500)$, and $\rho = .25$ (solid black), 0.50 (dashed red), 0.75 (dotted green), 0.90 (dot-dashed blue). Circles indicate Type 1 error rate at $b_{opt}$.

enforcing a tighter neighborhood, this is because the curve almost always shifts horizontally to the origin as the significance level increases.

The bandwidth rule $b_{opt}$ in equation (5.7) with $\tau$ in (5.8) is ideal for zero lugsail kernels. Many of the other bandwidth rules discussed depended on the kernel function through $q$. Similarly, this rule relies on the fact that zero lugsail estimators have $q = \infty$, and no other kernel information. The class of flat top kernels proposed by Politis [48, 47] also have $q = \infty$ but are constructed in a different fashion. Under the right settings some flat top kernels are equivalent to zero lugsail kernels, e.g. the Bartlett kernel. Although this rule in equation (5.7) does not immediately apply to the class flat top kernels generally, it is reasonable to presume with minor modifications the results can be extended to include the them.

Determining the best parameter for an 'L' shaped function is not a particularly new problem. Functions with similar behaviors are sometimes referred to as scree, knee, or elbow plots in clustering problems [19]. There are different methods used in clustering to select a tuning parameter based on an elbow plot. One method is to select the tuning parameter at the point of maximum curvature [66]. The curvature at a point is the radius of a circle that best approximates the curve of that point, with a curvature of zero indicting the point is on a straight line. Optimization methods sometimes select tuning parameter for elbow plots by using maximum curvature which can be determined via a grid search. A numerical investigation found that the $b_{opt}$ rule proposed with recommended $\tau$ is typically between the elbow and the point of maximum curvature, and the three values converge as

the sample size increases. An advantage of the $b_{opt}$ rule is that it does not require a grid search, automatically adjusts to the significance level, and accounts for dimensionality bias.

Although we can select a smaller $\tau_0$ to obtain an even smaller size distortion, we found this rule of thumb suffers only a mild power loss and typically falls between the conservative elbow rule ($\tau_0 = 1$) and aggressive rule of selecting a bandwidth that maximizes curvature.

## 5.4 Extensions to the Zero-Lugsail Bandwidth Rule

The bandwidth rule $b_{opt}$ in (5.7) differs from most other rules because of its reliance on finite sampling bias instead of kernel bias. It is not immediately applicable to non-zero lugsail settings, i.e. mother, adaptive, or over. To use a comparable rule with adaptive or over lugsail settings we recommend substituting a candidate $b$ value for the kernel bias term in the Type 1 error expression in Corollary 1a and continue using the loss function in (5.6) with the recommended $\tau$ in (5.8). This results in a slightly modified version of (5.7). The effective bandwidth $b_{opt}$ becomes smaller than the bandwidth rule for zero lugsail kernels. This is expected since over and adaptive lugsail kernels induce a positive bias which makes the overall net bias negligible at a faster rate than zero lugsail kernels, requiring a smaller bandwidth. With positively biased kernels it may even be possible to neutralize the finite sampling bias. We recommend using the zero lugsail rule in (5.7) as the candidate $b$ value.

For mother kernels we cannot use the same procedure without modifying $\tau$ because the negative kernel bias cause the the curve to shift vertically away from the origin and therefore there is no longer a $b$ that meets the criteria in (5.6) using the recommended

$\tau$. Instead we recommend using the zero bandwidth rule described in (5.7) as a rudimentary and effective approach. This obtains an optimal bandwidth for mother kernels according to their finite sampling bias exclusively, instead of the kernel bias exclusively which contrasts what most other bandwidth rules consider. Presumably a slightly larger bandwidth would be more desired for mother kernels, but this criteria would still result in something reasonably close.

There is also the possibility of utilizing a loss function of $\tau$ that considers the behavior of the derivative of the size distortion for both the finite sampling bias and kernel bias, again finding a $b$ where the derivative is *near* 0. However, in most instances no closed form solution exists, so one must resort to a numerical method. Additional investigation on ways to incorporate both sources of biases is warranted.

Furthermore, there is also the consideration of negatively correlated data. Throughout this document we have been under the premise that the data is positively correlated. With negatively correlated data the issue of bias is less problematic because of the oscillatory nature of the autocovariance function. Although one may substitute $\rho$ with $|\rho|$ in the bandwidth rule (5.7) with a zero lugsail kernel, it would result in an unnecessarily large loss in power. The mother kernel with the same bandwidth rule would be a more reasonable approach. The major issue for inference with negatively correlated data is not bias but the distributional properties of $\hat{\Omega}_T$.

Lastly, the bandwidth rule we suggested is built with an estimated $\rho$. For multivariate data multiple values for $\rho$ can be considered. We suggest using a weighted average of

the estimated autocorrelation coefficients, or selecting the largest value that was estimated. Both techniques are common practice [1, 33].

# Chapter 6

# Robust Testing In Practice

## 6.1 Estimate Corrections

Lugsail kernels do not guaranteed a positive semi-definite estimate which we can verify by observing that there exists some $\omega \in [0, 2\pi]$ such that $K(\omega) \leq 0$. Although the true LRV is positive semi-definite, our estimates for $\hat{\Omega}_T$ need to be positive definite so the test statistic $F_T$ is defined and the inverse of $\hat{\Omega}_T$ exists. The lack of positive semi-definiteness is not new to LRV estimators. The truncated kernel, and flat top kernels also suffer from this problem, where several corrections have been purposed [35, 36, 37, 61]. We suggest implementing a simple correction that still generates a consistent estimator as $b \to 0$ at a slower rate than $T \to \infty$. Let $\hat{\Omega}_T^{(L)}$ be the unadjusted lugsail estimator, $\hat{\Omega}_T^{(M)}$ be an SV estimator generated with $\kappa^* \in \mathscr{K}_1$, and $I$ be the set of indices such that $\left[\hat{\Omega}_T^{(L)}\right]_{i,i} \leq 0$. We define the new positive semi-definite corrected lugsail estimator as

$$\left[\hat{\Omega}_T^+\right]_{i,j} = \begin{cases} \left[\hat{\Omega}_T^{(M)}\right]_{i,j} & \text{if } (i=j) \text{ and } i \in I \\ \left[\hat{\Omega}_T^{(L)}\right]_{i,j} & \text{if otherwise} \end{cases}. \tag{6.1}$$

The estimator $\hat{\Omega}_T^+$ simply replaces values that in the estimate that are non-positive using a mother a kernel.

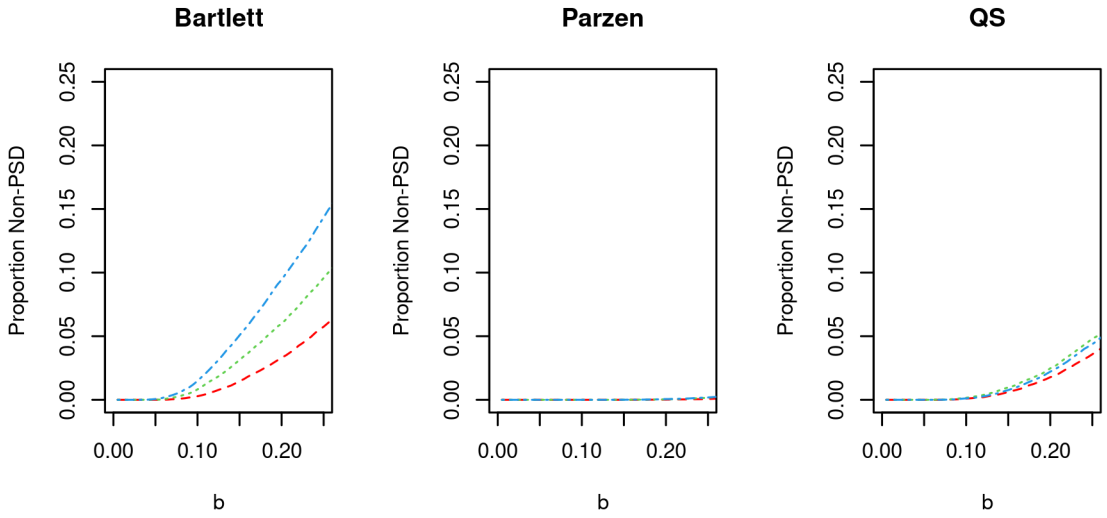


Figure 6.1: Proportion of zero (dashed red), adaptive (dotted blue), and over (dot dashed green) lugsail estimators corrected for lack of positive definiteness using 50,000 simulations.

In Figure 6.1 we observe the proportion of times a lugsail estimator with $d = 1$ for the Bartlett, Parzen, and quadratic spectral kernels was corrected using the zero (dashed red), adaptive (dotted blue), and over (dot dashed green) lugsail settings. Simulations were generated under the settings described in Section 7.2 using 50,000 simulations. A correction becomes more likely as $b$ increases as expected because autocovariance matrices with higher lags are more variable and utilize less data. Larger data sets are expected to need fewer corrections because the estimator is asymptotically positive semi-definite.

There is also the issue of an estimator being computationally positive definite. Despite being mathematically positive semi-definite sometimes estimates have exceptionally small or large elements that are beyond the scope of the software to handle. This happens as $d$ and $b$ increase, and is typically only observed under simulation experiments with extreme and unlikely settings. In these instances the `nearPD()` function is used from the `R` package `Matrix` to correct estimates to be more manageable.

Because positive semi-definiteness is no longer guaranteed for lugsail estimators the motivation for using the biased autocovariance estimator $\hat{\Gamma}(s)$ diminishes. We can instead revert back to the unbiased autocovariance estimator $\frac{T}{T-s}\hat{\Gamma}(s)$ to further decrease bias [48, 49]. However, we will continue to use the conventional estimator $\hat{\Gamma}(s)$ to be consistent with current software and practices.

## 6.2 Procedure

The heteroskadistic and autocorrelation robust inference procedure with a testing optimal zero lugsail estimator can be outlined as follows.

1. Specify the the significance level $\alpha$, and the hypotheses

$$H_0: \quad R(\theta - \theta_0) = 0$$

$$H_A: \quad R(\theta - \theta_0) \neq 0$$

   where $R$ is a $(d \times p)$ known matrix of interest specific to the application.

2. Estimate $\theta$ by ordinary least squares (OLS) and obtain the residuals $\hat{u}_t$. Center the residuals, $\tilde{u}_t = \hat{u}_t - \hat{\bar{u}}$.

3. Fit an VAR(1) model with autocorrelation matrix $\rho\mathbf{1}_p$ to the centered residuals and compute $\hat{\rho}$.

4. Specify the kernel $\kappa \in \mathscr{K}_2$.

5. Compute the automatic bandwidth using the rule in equation (5.6).

6. Calculate $\hat{\Omega}_T^{(L)}$ using equation (2.5) with the centered errors and selected kernel from the previous steps.

7. Check for positive definiteness, and obtain the estimator $\hat{\Omega}_T^+$ as described in equation (6.1).

8. Calculate the test statistic $F_{OLS}$ in equation (4.4) with the LRV estimator $\hat{\Omega}_T^+$.

9. Obtain the critical value $c_d^\alpha(b)$ using an appropriate fixed-$b$ table.

10. Compare test statistic to fixed-$b$ critical value and reject the null hypothesis if $F_{OLS} > c_d^\alpha(b)$.

The utilization of centered errors, lugsail kernels, fixed-$b$ critical values, and the data driven bandwidth rule are all minor modifications that add little to no practical complexity for the practitioner compared to similar procedures in heteroskedastic and autocorrelation robust settings [40, 56]. The procedure is robust in the sense that it is asymptotically valid for a wide variety of error structures. Although there are several steps that require some discernment, we can also think of this procedure as being robust to the choices of the practitioner as most choices will still yield a favorable performance.

The first choice is that of $R$, which is typically $\mathbf{1}_p$ unless there is another application specific relation of interest. There is no statistical motivation for the choice of $R$.

The next choice comes from the estimation of the finite sample bias and $\rho$. We have recommended a simple, popular, and effective but crude approach in Proposition 3. If there is reasonable motivation for approximating the behavior of the error terms with another structure, then the procedure can be modified to do so but it may not result in a closed form solution for $b_{opt}$. For multivariate data sets a few other modifications to step 4 were discussed in Section 5.3. This estimation step is not novel, and some semblance of it appears in most automatic bandwidth procedures [1, 2, 27, 29, 53, 54, 56].

Next the practitioner must determine which kernel function to select. We recommend considering the broad class of lugsail kernels $\mathscr{K}_2$, recall $\mathscr{K}_1 \subset \mathscr{K}_2$. Historically, robust inference procedures restricted the possible kernels to class $\mathscr{K}_1$. Among the kernels in class $\mathscr{K}_1$ two are especially pertinent, the quadratic spectral kernel and the Bartlett kernel. The quadratic spectral kernel has been shown to have optimal asymptotic MSE [1], achieves the envelope of the size-adjusted power, is optimal among other testing focused bandwidth rules with fixed-$b$ critical values [57], and is equivalent to a weighted orthonormal series estimator [27]. In contrast, the Bartlett kernel is computationally efficient and is asymptotically equivalent to a batch means estimator [4, 62]. In general mother kernels with $q = 2$ tend to preform marginally better according to a variety of metrics [49, Sec 7.5]; however, there typically is not a large discrepancy between kernels in $\mathscr{K}_1$. We have chosen the Bartlett kernel for the simulation study in the subsequent chapter due to its popularity, efficiency,

and its readable connections to the batch means and flat top (with zero lugsail settings) estimators.

In regards to the lugsail settings we recommend selecting the mother lugsail settings when $\hat{\rho} < 0$, and the zero lugsail setting when $\hat{\rho} > 0$; for $\hat{\rho} \approx 0$ the two settings are equivalent. For exceptionally large samples and highly correlated data the over and adaptive lugsail settings can be utilized instead. Highly correlated data negatively impacts power, and the increased variability of the over and adaptive lugsail settings further decrease it. We recommend selecting the over or adaptive lugsail settings according to Table (2.1), but only when the data set is substantially large.

Lastly, there is the tuning parameter $\tau$ which is used to obtain $b_{opt}$. We have suggested picking a $\tau$ that is responsive to the behavior of the Type 1 error by picking a bandwidth where the Type 1 error deteriorates at a faster rate than the bandwidth grows. Following the guidelines in Chapter 5. It is tempting to select some small $\tau_0$ where $\tau = \frac{-\tau_0}{Tlog(\rho)}$ to ensure a bandwidth clears the elbow with room to spare but such a practice without an adequately large sample size is likely to result in a substantial loss of power. It is recommended to not deviate from our recommendation without justification, because although Type 1 error would be improved marginally, the Type 2 error would suffer greatly. Through simulation it was observed that out of all these components the bandwidth is the most critical which follows the results of others [2, 14]. Robustness to bandwidth selection is improved as sample size increases.

# Chapter 7

# Examples

To asses the validity of our findings we might ask ourselves the questions: *How well do the derived expressions represent the truth? How does the testing procedure compare to alternative procedures? How do lugsail settings influence testing performance?* To address these broad prompts we conduct a simulation study with a variety of classic yet common scenarios in time series and econometric applications that are often used as benchmarks [1, 28, 29, 61]. We caution that we cannot address these questions comprehensively due to the range of applicability, but instead rely on fundamental models to showcase general trends.

In Section 7.1 we begin with generating our fixed-$b$ critical values and representations thereof. In Section 7.2 we use a VAR(1) model with $d = 1, 2, 3, 4$ and observe the behavior of the error curves and power function. In Section 7.3 we concentrate on various different model structures in econometrics and compare Type 1 error rates. For Sections 7.2 and 7.3 all results were generated using 1000 simulations, $\alpha = 0.05$ and the Bartlett kernel

with mother and zero lugsail settings. We leave the exploration of adaptive and over lugsail kernels as a future project. In Section 7.4 we showcase our model with an application in finance. Section 7.5 concludes with a discussion.

## 7.1 Fixed-$b$ Critical Values

The fixed-$b$ distribution is not generally tractable, so to obtain critical values we follow the work of Kiefer & Vogelsong [25], Sun et al. [56], and Lazarus et al. [29]. We simulate 50,000 replicates of normalized random vectors with $T = 1000$ from an independent standard normal distribution.

| | *Mother* | | | | *Zero* | | | |
|---|---|---|---|---|---|---|---|---|
| $b \setminus d$ | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 0.005 | 3.846 | 3.009 | 2.635 | 2.417 | 3.884 | 3.043 | 2.684 | 2.475 |
| 0.010 | 3.880 | 3.071 | 2.692 | 2.489 | 3.969 | 3.183 | 2.812 | 2.629 |
| 0.015 | 3.929 | 3.132 | 2.758 | 2.555 | 4.067 | 3.294 | 2.953 | 2.787 |
| 0.020 | 3.975 | 3.189 | 2.827 | 2.635 | 4.164 | 3.429 | 3.115 | 2.972 |
| 0.025 | 4.033 | 3.239 | 2.902 | 2.704 | 4.287 | 3.559 | 3.288 | 3.181 |
| 0.030 | 4.085 | 3.298 | 2.978 | 2.781 | 4.409 | 3.701 | 3.469 | 3.418 |
| 0.035 | 4.143 | 3.364 | 3.037 | 2.863 | 4.485 | 3.862 | 3.673 | 3.696 |
| 0.040 | 4.191 | 3.429 | 3.109 | 2.946 | 4.580 | 4.017 | 3.943 | 4.063 |
| 0.045 | 4.245 | 3.491 | 3.194 | 3.038 | 4.757 | 4.223 | 4.228 | 4.549 |
| 0.050 | 4.310 | 3.557 | 3.273 | 3.122 | 4.865 | 4.414 | 4.622 | 5.233 |
| 0.055 | 4.358 | 3.614 | 3.340 | 3.216 | 5.008 | 4.632 | 5.049 | 6.190 |
| 0.060 | 4.422 | 3.690 | 3.414 | 3.306 | 5.174 | 4.931 | 5.589 | 7.985 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

Table 7.1: Bartlett Critical Value Table for Mother and Zero Lugsail Kernels at $\alpha = 0.05$.

We approximate the fixed-$b$ distribution for the Bartlett, Parzen, and quadratic spectral kernels with mother, zero, adaptive (initial $b = 0.022$), and over lugsail settings for $b = 0.005, 0.010, 0.015, \ldots, 0.990$ and $d = 1$. In addition we further approximated the

distribution for the Bartlett kernel under various lugsail settings for $d = 1, 2, 3, 4$. We use the quantiles of these simulated distributions to obtain fixed-$b$ critical values which are then organized into tables. An example of such a table for the Bartlett kernel using the mother and zero lugsail settings at $\alpha = 0.05$ is provided in Table 7.1.

Some authors extrapolate critical value relationships across bandwidths for a given kernel, $d$, and significance level using the model,

$$cv_d^{\alpha}(b) = a_0 + a_1 b + a_2 b^2 + a_3 b^3. \tag{7.1}$$

This model is fitted via OLS using the critical value tables. It is common to fix the intercept to be the small-$b$ critical value because of the relationship between small-$b$ and fixed-$b$ asymptotics. We have plotted the fixed-$b$ distribution of the test statistic under various settings when $d = 1$ in Figures 7.1 - 7.3. On the left side of the figure we have density functions for a specified bandwidth and denoted the specific critical value across the different settings. On the right side of the figure we have the fitted critical value curves from equation (7.1). The points that intersect the vertical line correspond to the critical values for the specific the bandwidth, kernel and significance level.

When fitting the critical value model in equation (7.1) we observed an $R^2$ of 0.990 or higher for all settings with $d = 1$. However, we found that using the critical value tables directly was more accurate especially for smaller bandwidths and higher dimensions. We use the fixed-$b$ critical values for the remainder of the simulations.

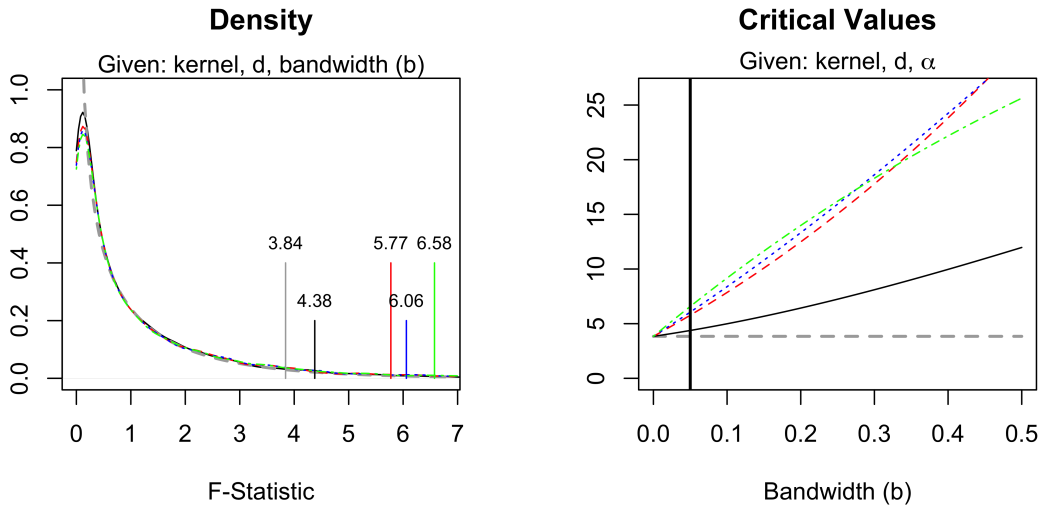


Figure 7.1: Colors indicate the mother (solid black), zero (dashed red), adaptive (dotted blue), and over (dot-dashed green) lugsail settings. The small-$b$ asymptotic values are in dashed grey. *Left*: Fixed-$b$ density for $F_T$ with $d = 1$ and $b = 0.05$ using Bartlett lugsail kernels. Vertical lines indicate the $\alpha = 0.05$ critical values. *Right*: Fitted critical values of $F_T$ using using Bartlett lugsail kernels. Vertical line intersection indicates the critical values for $b = 0.05$ for the various lugsail kernels. Intersection points match vertical lines on the left plot.

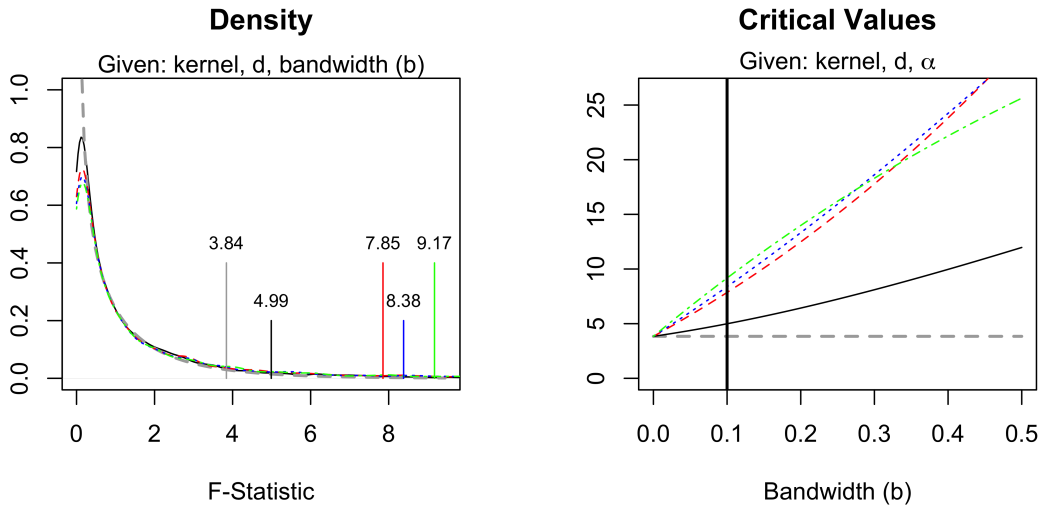


Figure 7.2: Colors indicate the mother (solid black), zero (dashed red), adaptive (dotted blue), and over (dot-dashed green) lugsail settings. The small-$b$ asymptotic values are in dashed grey. *Left*: Fixed-$b$ density for $F_T$ with $d = 1$ and $b = 0.1$ using Bartlett lugsail kernels. Vertical lines indicate the $\alpha = 0.05$ critical values. *Right*: Fitted critical values of $F_T$ using using Bartlett lugsail kernels. Vertical line intersection indicates the critical values for $b = 0.1$ for the various lugsail kernels. Intersection points match vertical lines on the left plot.

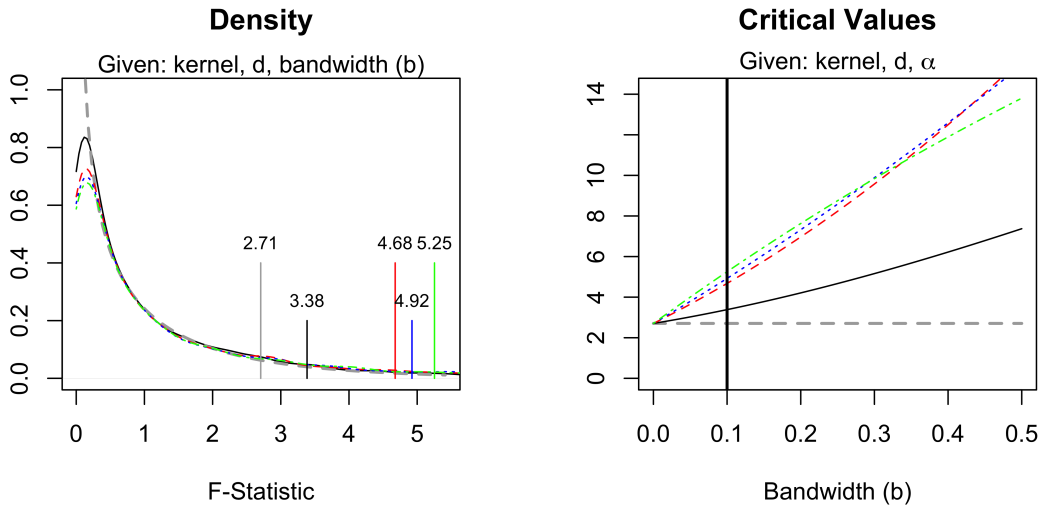


Figure 7.3: Colors indicate the mother (solid black), zero (dashed red), adaptive (dotted blue), and over (dot-dashed green) lugsail settings. The small-$b$ asymptotic values are in dashed grey. *Left*: Fixed-$b$ density for $F_T$ with $d = 1$ and $b = 0.1$ using Bartlett lugsail kernels. Vertical lines indicate the $\alpha = 0.1$ critical values. *Right*: Fitted critical values of $F_T$ using using Bartlett lugsail kernels. Vertical line intersection indicates the critical values for $b = 0.1$ for the various lugsail kernels. Intersection points match vertical lines on the left plot.

## 7.2 Accuracy of Approximation

First consider a Gaussian location model with VAR(1) error terms $w_t$, and independent standard normal disturbances $\epsilon_t$ as described in Section 5.2. The primary goal of this section is to asses if the testing procedure preforms as expected. We simulate results for a VAR(1) process with $T = 500$, $d = (1, 2, 3, 4)$, $\rho = (0, 0.75)$, small-$b$ and fixed-$b$ critical values, and the mother and zero lugsail settings. For each scenario we present the results into three plots: bandwidth versus Type 1 error rate, bandwidth versus Type 2 error rate, and noncentrality parameter versus power.

For Type 2 error rate we must select a point under the alternative hypothesis. It is customary to select a point $\delta^2$ under the alternative hypothesis such that the asymptotic power is around 50% - 75% [6, 54, 57]. In other words, select $\delta^2$ such that $P_{\delta^2}(\varrho \geq \chi^{\alpha}_{d}) = .50$ (or .75) where $\varrho \sim \chi^2_{d,\delta^2}$. We choose to generate data under the alternative hypothesis $H_A : R(\theta - \theta_0) = R\Lambda T^{-1/2}\delta$ where $\delta^2$ corresponds to the point where asymptotic power is approximately 0.66. Lazarus et al. [29] discovered through simulation that maximum size adjusted power loss ($\Delta_p^{max}$) for mother kernels typically occus at this point regardless of $d$. Recall that fixed-$b$ critical values are still asymptotic critical values that may not accurately represent finite data sets and not achieve a test with a rejection rate of size $\alpha$. *Size adjusted power* is the power of the test when using critical values that have been adjusted to be exact. *Size-adjusted power loss* is the difference in power when using asymptotic critical values and exact critical values. Per the discussion in Chapter 5 we observed that when we adjust the critical values from small-$b$ to fixed-$b$ the power of the test suffers; however, the tests do not have the same $\alpha$ level because of the testing mechanics, thus the two scenarios

are not really comparable. The concept of size adjusted power was conceived in order to give a fairer comparison across procedures [29]. The *maximum size adjusted power loss* is largest difference in power between the adjusted and unadjusted critical values. Lazarus et al. [29] determined that for mother kernels with fixed-$b$ critical values this point typically occurs at $\delta^2 = 0.66$. This could also be thought of as the point at which the true power data is the least represented by the asymptotic power. We have selected this point for comparison and Table 7.2 contains the $\delta^2$ values used.

| $d$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\delta^2$ | 2.43 | 3.25 | 3.81 | 4.27 |

Table 7.2: $\delta^2$ Values for $d = 1, 2, 3, 4$.

We further note that the AMSE rule $b_{mse}$ in equation 5.2 was founded for mother kernels using small-$b$ critical values, the classical approach for robust estimation, and $b_{opt}$ in equation 5.7 was founded for zero lugsail kernels with fixed-$b$ critical values. We have indicated the $b_{mse}$ and $b_{opt}$ values for their respective scenarios in the Type 1 and Type 2 error plots.

In addition, we plot the performance for mother kernels using fixed-$b$ critical values and zero lugsail kernels with small-$b$ critical values to better observe the effects of the two mechanics. We have not indicated any specific bandwidth points under these scenarios for the error plots. However, the power function must be plotted using a specified bandwidth, we chose $b_{mse}$ for mother kernels and $b_{opt}$ for zero lugsail kernels for convenience.

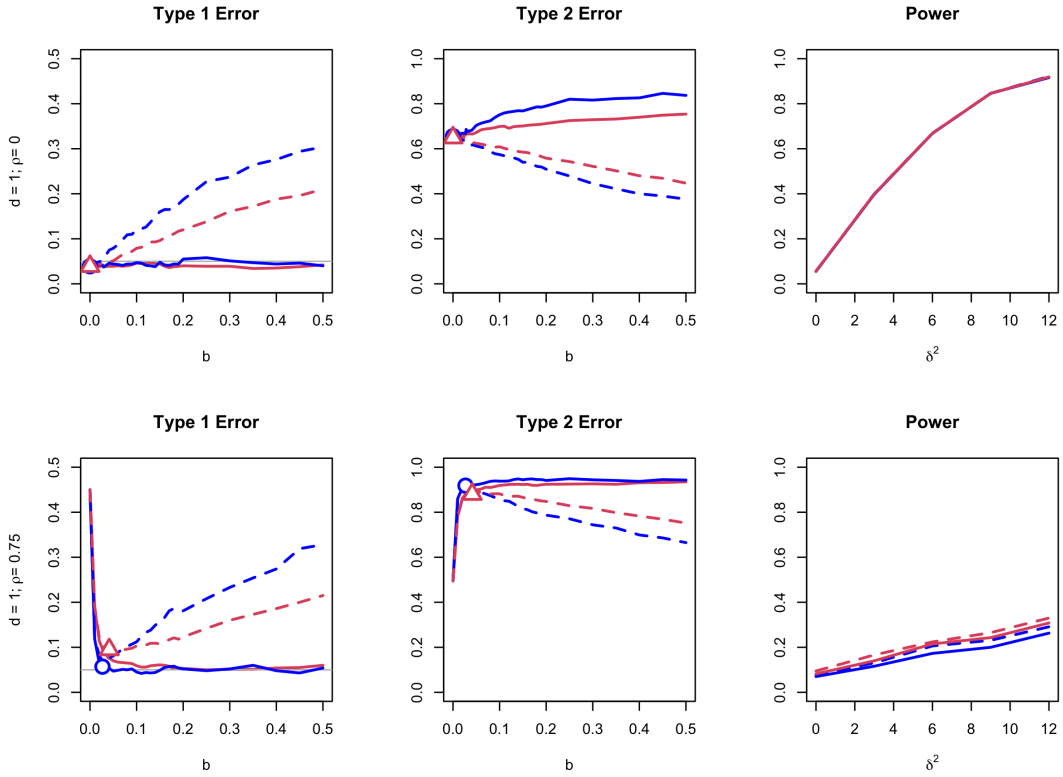


Figure 7.4: Inference performance metrics for a VAR(1) process with $d = 1$. Dashed lines indicate small-$b$ critical values, and solid lines are fixed-$b$ critical values. Red indicates the mother Bartlett kernel, and blue indicates the zero lugsail Bartlett kernel. The triangles and circles indicate $b_{mse}$ and $b_{opt}$ respectively. The power curve has been drawn using $b_{mse}$ for the mother kernels, and $b_{opt}$ for the zero lugsail kernels. The Type 2 error rate is at a fixed point under the alternative hypothesis.

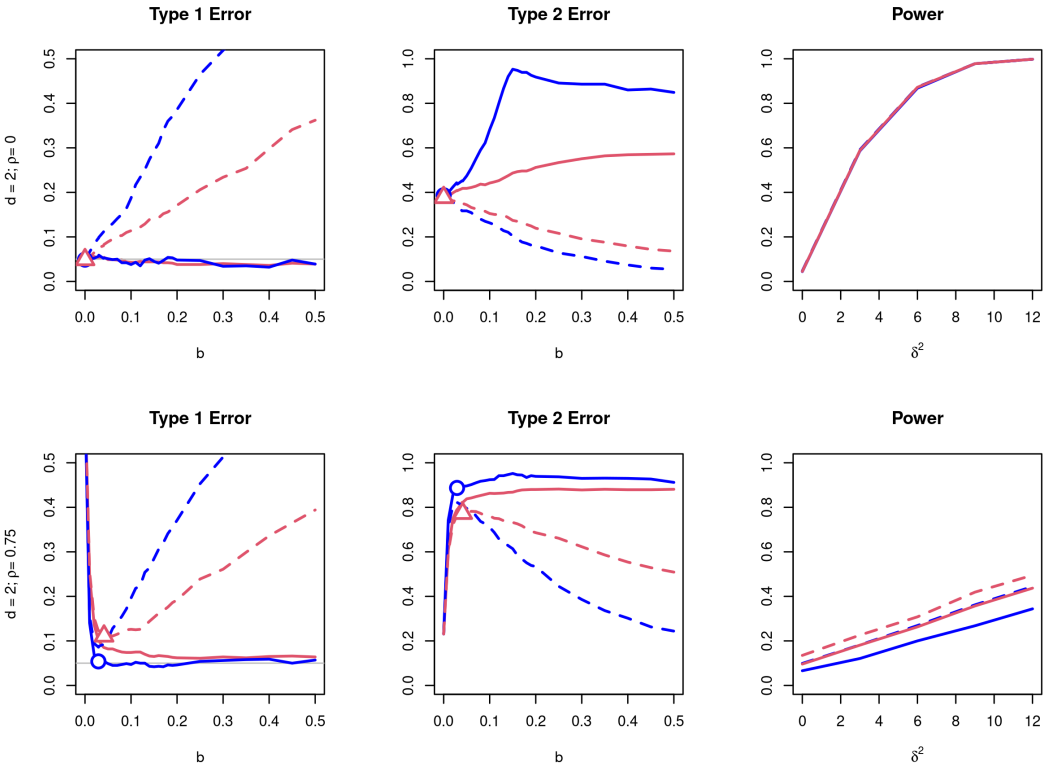


Figure 7.5: Inference performance metrics for a VAR(1) process with $d = 2$. Dashed lines indicate small-$b$ critical values, and solid lines are fixed-$b$ critical values. Red indicates the mother Bartlett kernel, and blue indicates the zero lugsail Bartlett kernel. The triangles and circles indicate $b_{mse}$ and $b_{opt}$ respectively. The power curve has been drawn using $b_{mse}$ for the mother kernels, and $b_{opt}$ for the zero lugsail kernels. The Type 2 error rate is at a fixed point under the alternative hypothesis.

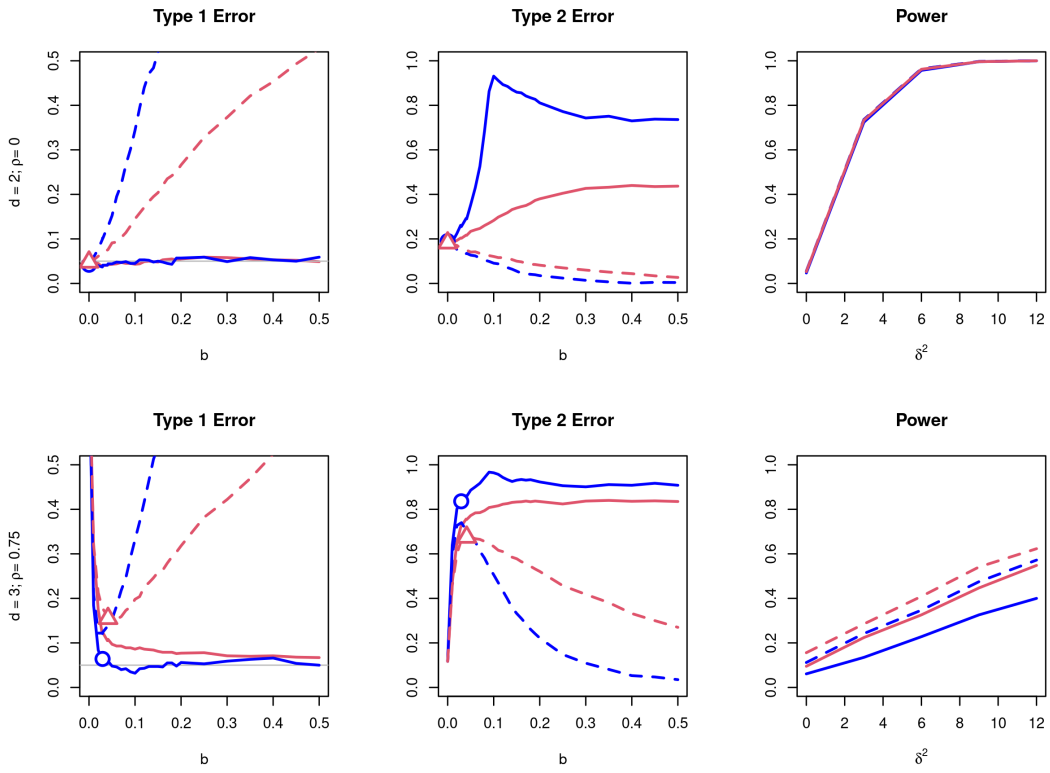


Figure 7.6: Inference performance metrics for a VAR(1) process with $d = 3$. Dashed lines indicate small-$b$ critical values, and solid lines are fixed-$b$ critical values. Red indicates the mother Bartlett kernel, and blue indicates the zero lugsail Bartlett kernel. The triangles and circles indicate $b_{mse}$ and $b_{opt}$ respectively. The power curve has been drawn using $b_{mse}$ for the mother kernels, and $b_{opt}$ for the zero lugsail kernels. The Type 2 error rate is at a fixed point under the alternative hypothesis.

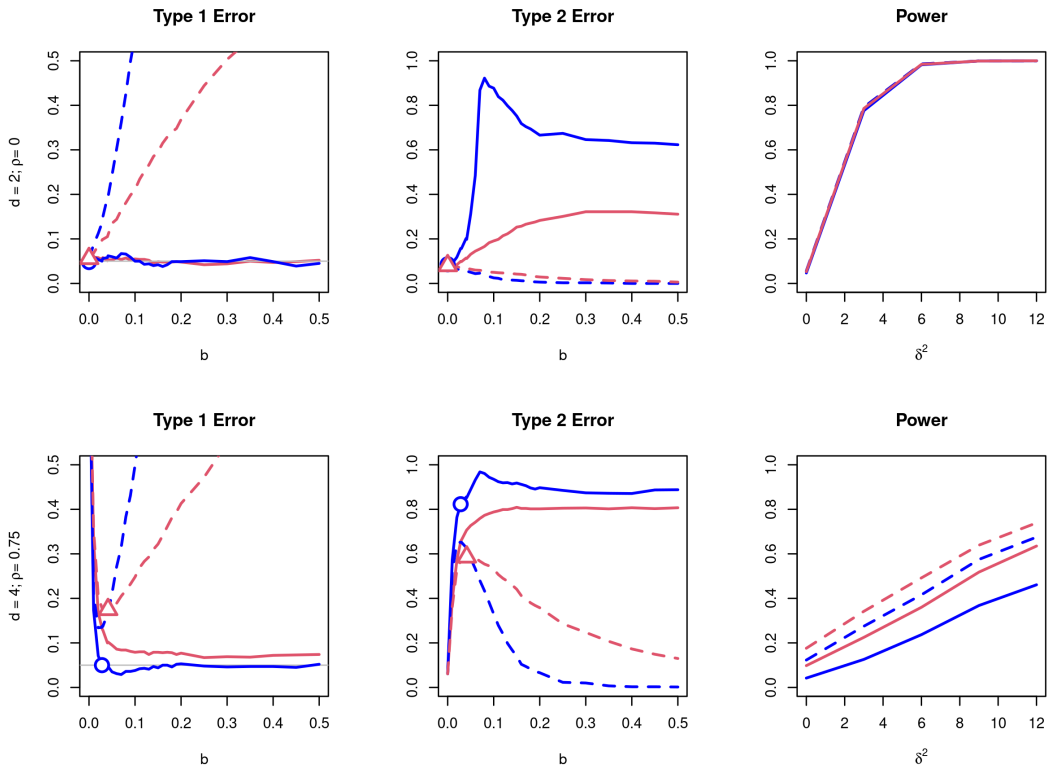


Figure 7.7: Inference performance metrics for a VAR(1) process with $d = 4$. Dashed lines indicate small-$b$ critical values, and solid lines are fixed-$b$ critical values. Red indicates the mother Bartlett kernel, and blue indicates the zero lugsail Bartlett kernel. The triangles and circles indicate $b_{mse}$ and $b_{opt}$ respectively. The power curve has been drawn using $b_{mse}$ for the mother kernels, and $b_{opt}$ for the zero lugsail kernels. The Type 2 error rate is at a fixed point under the alternative hypothesis.

When $\rho = 0$ the optimal bandwidth is $b_{mse} = b_{opt} = 0$, and the LRV estimators are the same for all four scenarios. The discrepancy between the performance of the Type 1 error and Type 2 error plots is largely due to variability. The large bandwidths increase variability which is why the Type 1 error rate begins to inflate using the small-$b$ critical values. The fixed-$b$ critical values are bandwidth specific and account for the distributional effects of $\hat{\Omega}_T$. Type 1 error does not suffer from increasing the bandwidth while using fixed-$b$ critical values, but the Type 2 error is worse for fixed-$b$ critical values then for small-$b$. Accounting for the increased variability inflates the critical values and the distributions become hard to distinguish. The power is the same for all scenarios as each curve was plotted using $b = 0$ at which both the zero lugsail and mother kernels are equivalent, and the fixed-$b$ and small-$b$ critical values converge.

When $\rho = .75$ we observe an arm like shape for Type 1 error as expected. Higher correlation tends to be more notorious to model. We observe that zero lugsail kernel with fixed-$b$ critical value essentially attains the prescribed error rate. We further observe using fixed-$b$ critical values the zero lugsail kernel is closer to the prescribed error rate for nearly all values of $b$. The power curve also changes in the presence of correlation. Difference between the dashed lines and solid lines are due to variability, and the difference between the colors are due to bias. The larger the bandwidth the more the curves diverge from each other as expected.

## 7.3 Comparing Procedures Mechanics

In this section we compare the Type 1 error rates for a variety of econometric models. Let $\{w_t\}$ be an unknown stationary process with mean zero and finite variance that is possibly correlated or heteroskedastic. We further define

$$\begin{aligned} \tilde{x}_t' &= (\tilde{x}_{t,1}, \ldots, \tilde{x}_{t,p}) & \beta' &= (\gamma, \theta_1, \ldots, \theta_p) \\ x_t' &= (1, \tilde{x}_t') & &= (\gamma, \theta') \\ X &= [x_1', x_2', \ldots, x_T'] & Y' &= [y_1, \ldots, y_T] \end{aligned}$$

where $\beta \in \mathbb{R}^{p+1}$ is a vector of parameters to be estimated, $\{y_t\}$ is a sequence of observed random variables, and $X$ is an observed $(T \times (p+1))$ matrix. For our simulation study we use the common linear model structure:

$$y_t = x_t'\beta + w_t. \tag{7.2}$$

where $t = 1, \ldots, T$. To estimate $\beta$ we use ordinary least square (OLS) estimators.

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'Y \\ &= \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\sum_{t=1}^{T} x_t y_t\right) \end{aligned}$$

To conduct inference on the parameters $\beta$, we require the covariance matrix for $\beta$. We first construct $\hat{\Omega}_T$ from equation (2.14) using $\hat{u}_t = x_t y_t$ and $\kappa \in \mathscr{K}_2$. Let $M = (\frac{1}{T}X'X)^{-1}$, then the covariance matrix of $\beta$ is $M\hat{\Omega}_T M$ where the $i^{th}$ diagonal element is

the variance for $\beta_i$. Let $e_t \overset{iid}{\sim} N_p(0, \mathbf{1}_p)$. We generate $X$ using an auto-regressive process of the following form,

$$\tilde{x}_t = \rho \tilde{x}_{t-1} + e_t. \tag{7.3}$$

That is, each $\tilde{x}_{t,i}$ is an independent AR(1) process. We consider the following different structures for the error term $w_t$.

- **AR1-HOMO**: An auto-regressive model of order 1 with homoskedastic disturbances [1, 54, 61]. Let $\epsilon_t \overset{iid}{\sim} N(0, 1)$, and

$$w_t = \rho_w w_{t-1} + \epsilon_t. \tag{7.4}$$

- **AR1-HET**: An auto-regressive process of order 1 with heteroskedastic disturbances [1, 11, 65]. Let $a_0 = 5$, $a_1 = .25$, $\epsilon_t \overset{iid}{\sim} N(0, 1)$, and

$$w_t = \sqrt{a_0 + a_1 w_{t-1}} v_t$$

$$v_t = \rho_v v_{t-1} + \epsilon_t.$$

- **ARMA-G**: A simple auto regressive moving average model, ARMA(1,1), with standard Gaussian disturbances. Let $Z_t \overset{iid}{\sim} N(0, 1)$, $\epsilon_t \overset{iid}{\sim} N(0, 1)$, $\psi_z = 0.5$, and

$$w_t = Z_t + \psi_z Z_{t-1} + \rho_w w_{t-1} + \epsilon_t. \tag{7.5}$$

- **ARMA-L**: A simple auto regressive moving average model, ARMA(1,1), with standard Laplace disturbances. Let $Z_t$ and $\psi_z = 0.5$ be as defined in equation (7.5), $\epsilon_t \overset{iid}{\sim} Laplace(0, 1)$, and

$$w_t = Z_t + \psi_z Z_{t-1} + \rho_w w_{t-1} + \epsilon_t.$$

In general we consider a range of values and let $0 < \rho = \rho_w = \rho_v < 1$. The first model AR1-HOMO is self evident, it is used as the backbone of the bandwidth rule for our procedure, and of most procedures in this space [1, 61]. We next consider the model AR1-HET, a commonly used model for errors with serial correlation and heteroskedasticy [1, 11, 65]. The last two models are ARMA-L and ARMA-G are used to explore different error processes with Gaussian and non-Gaussian disturbances. The ARMA models considered are based off of McElroy & Politis [35] and Lazarus et al. [28].

Our test is invariant to our choice of $\beta$ so we generate the data with $\gamma = 0$ and $\theta_i = 0$ for $i = 1, \ldots, p$. We test the hypotheses that the first coefficient after the intercept is zero, i.e. $H_0 : \theta_1 = 0$, using $\alpha = 0.05$ and the Bartlett kernel. Our test statistic can be expressed as,

$$\begin{aligned} F_{OLS} =& \left[\sqrt{T}\left(\hat{\theta}_{1,T} - \theta_{1,0}\right)\right]' \left[M\hat{\Omega}_T^{-1}M\right]_{22} \left[\sqrt{T}\left(\hat{\theta}_{1,T} - \theta_{1,0}\right)\right] /d \\ =& \frac{T\hat{\theta}_{1,T}^2}{Var(\hat{\theta}_1)} \end{aligned}$$

Different sample sizes, various $\rho$'s, and four different procedures are considered. We use the procedure outlined in Section 6.2 with zero lugsail kernel and the absolute value of $\hat{\rho}$ to obtain $b_{opt}$. We also consider the bandwidth rule $b_{ft}$ with a zero lugsail kernel using both fixed-$b$ and small-$b$ critical values as this procedure depends on the behavior of the autocovariance function alone and not testing mechanics. Lastly, we consider the classical procedure with the mother kernel and small-$b$ critical values. The results are presented in Tables 7.3, 7.4, 7.5 and 7.6.

| | | Fixed-$b$ | | Small-$b$ | |
|---|---|---|---|---|---|
| | $\rho$ | $b_{opt}$ | $b_{ft}$ | $b_{ft}$ | $b_{mse}$ |
| | 0.00 | 0.044* | 0.045 | 0.049+ | 0.045 |
| | 0.25 | 0.079 | 0.068*+ | 0.072 | 0.069 |
| *T=200* | 0.50 | 0.089 | 0.076*+ | 0.079 | 0.086 |
| | 0.75 | 0.104*+ | 0.148 | 0.159 | 0.117 |
| | 0.90 | 0.151*+ | 0.189 | 0.212 | 0.193 |
| | 0.00 | 0.062*+ | 0.066 | 0.067 | 0.063 |
| | 0.25 | 0.067 | 0.061*+ | 0.063 | 0.063 |
| *T=500* | 0.50 | 0.061 | 0.063 | 0.065 | 0.057*+ |
| | 0.75 | 0.087*+ | 0.104 | 0.106 | 0.091 |
| | 0.90 | 0.111*+ | 0.119 | 0.123 | 0.124 |
| | 0.00 | 0.035* | 0.037 | 0.038+ | 0.035 |
| | 0.25 | 0.069 | 0.058*+ | 0.060 | 0.058 |
| *T=1000* | 0.50 | 0.060 | 0.063 | 0.064 | 0.058*+ |
| | 0.75 | 0.061 | 0.072 | 0.077 | 0.057*+ |
| | 0.90 | 0.089*+ | 0.096 | 0.104 | 0.098 |

Table 7.3: Results with AR(1)-Homoskadiscity error terms. The $*$ denotes the smallest Type 1 error rate, and the $+$ denotes the setting closes to the prescribed $\alpha$.

| | | Fixed-$b$ | | Small-$b$ | |
|---|---|---|---|---|---|
| | $\rho$ | $b_{opt}$ | $b_{ft}$ | $b_{ft}$ | $b_{mse}$ |
| | 0.00 | 0.056 | 0.054 | 0.058 | 0.053*+ |
| | 0.25 | 0.086 | 0.074*+ | 0.080 | 0.075 |
| *T=200* | 0.50 | 0.087 | 0.073*+ | 0.077 | 0.078 |
| | 0.75 | 0.087*+ | 0.119 | 0.123 | 0.101 |
| | 0.90 | 0.157*+ | 0.198 | 0.218 | 0.214 |
| | 0.00 | 0.067 | 0.070 | 0.072 | 0.066*+ |
| | 0.25 | 0.054 | 0.050+ | 0.051 | 0.049* |
| *T=500* | 0.50 | 0.058 | 0.060 | 0.061 | 0.055*+ |
| | 0.75 | 0.070*+ | 0.099 | 0.103 | 0.080 |
| | 0.90 | 0.097*+ | 0.108 | 0.116 | 0.118 |
| | 0.00 | 0.047 | 0.046* | 0.050+ | 0.049 |
| | 0.25 | 0.072 | 0.061*+ | 0.062 | 0.063 |
| *T=1000* | 0.50 | 0.068*+ | 0.071 | 0.072 | 0.071 |
| | 0.75 | 0.071*+ | 0.082 | 0.083 | 0.076 |
| | 0.90 | 0.072*+ | 0.082 | 0.090 | 0.083 |

Table 7.4: Results with AR(1)-heteroskedasticy error terms. The $*$ denotes the smallest Type 1 error rate, and the $+$ denotes the setting closes to the prescribed $\alpha$.

| | | Fixed-$b$ | | Small-$b$ | |
|---|---|---|---|---|---|
| | $\rho$ | $b_{opt}$ | $b_{ft}$ | $b_{ft}$ | $b_{mse}$ |
| | 0.00 | 0.052*+ | 0.057 | 0.063 | 0.054 |
| | 0.25 | 0.084 | 0.067*+ | 0.068 | 0.069 |
| *T=200* | 0.50 | 0.062 | 0.057*+ | 0.066 | 0.062 |
| | 0.75 | 0.092*+ | 0.132 | 0.140 | 0.113 |
| | 0.90 | 0.142*+ | 0.175 | 0.203 | 0.197 |
| | 0.00 | 0.056 | 0.051 | 0.056 | 0.05*+ |
| | 0.25 | 0.075 | 0.068 | 0.069 | 0.067*+ |
| *T=500* | 0.50 | 0.069*+ | 0.075 | 0.076 | 0.071 |
| | 0.75 | 0.062*+ | 0.070 | 0.075 | 0.071 |
| | 0.90 | 0.093*+ | 0.105 | 0.126 | 0.128 |
| | 0.00 | 0.051+ | 0.057 | 0.057 | 0.049* |
| | 0.25 | 0.048+ | 0.037* | 0.039 | 0.039 |
| *T=1000* | 0.50 | 0.049*+ | 0.061 | 0.062 | 0.054 |
| | 0.75 | 0.061 | 0.071 | 0.072 | 0.06*+ |
| | 0.90 | 0.082*+ | 0.091 | 0.097 | 0.109 |

Table 7.5: Results with ARMA(1,1) error terms with Gaussian disturbances. The $*$ denotes the smallest Type 1 error rate, and the + denotes the setting closes to the prescribed $\alpha$.

| | | Fixed-$b$ | | Small-$b$ | |
|---|---|---|---|---|---|
| | $\rho$ | $b_{opt}$ | $b_{ft}$ | $b_{ft}$ | $b_{mse}$ |
| | 0.00 | 0.056 | 0.060 | 0.066 | 0.053*+ |
| | 0.25 | 0.067 | 0.061 | 0.065 | 0.060*+ |
| *T=200* | 0.50 | 0.076 | 0.069*+ | 0.075 | 0.072 |
| | 0.75 | 0.084*+ | 0.132 | 0.136 | 0.106 |
| | 0.90 | 0.148*+ | 0.181 | 0.209 | 0.199 |
| | 0.00 | 0.081 | 0.078 | 0.079 | 0.077*+ |
| | 0.25 | 0.064 | 0.053*+ | 0.053 | 0.055 |
| *T=500* | 0.50 | 0.063*+ | 0.068 | 0.069 | 0.067 |
| | 0.75 | 0.084*+ | 0.102 | 0.108 | 0.093 |
| | 0.90 | 0.116*+ | 0.129 | 0.143 | 0.146 |
| | 0.00 | 0.054 | 0.057 | 0.058 | 0.052*+ |
| | 0.25 | 0.066 | 0.053*+ | 0.054 | 0.054 |
| *T=1000* | 0.50 | 0.059*+ | 0.069 | 0.072 | 0.060 |
| | 0.75 | 0.066*+ | 0.073 | 0.077 | 0.069 |
| | 0.90 | 0.065*+ | 0.072 | 0.079 | 0.079 |

Table 7.6: Results with ARMA(1,1) error terms with Laplace disturbances. The $*$ denotes the smallest Type 1 error rate, and the + denotes the setting closes to the prescribed $\alpha$.

Across the different procedures, sample sizes, and $\rho$'s the observed Type 1 error rate closest to the prescribed rate was typically the procedure that relied on $b_{opt}$. In particular, for $\rho = 0.9$ the bandwidth rule $b_{opt}$ prevailed in all instances. When correlation was low the rates where more comparable as expected. Interestingly, we observe a larger difference between the procedures for the models AR1-HET and ARMA-L, with the $b_{opt}$ prevailing. This suggests the $b_{opt}$ procedure is more robust to complicated error structures. In instances when $b_{opt}$ is not the optimal choice it is typically near it.

## 7.4 Application

Following the work of Pellatt & Sun [43] and Chang et al. [2] we investigate how interest rates of securities with varying maturities move together. We consider the 10 year United States Treasury bond (T-bond) yield as a response variable and the 3 month United State Treasury bill (T-bill) yield as a explanatory variable. The data is recorded monthly from January 1962 and December 2007 for a total sample size of 512, and was collected from Federal Reserve Economic Data (FRED) of the St. Louis FED. The range was chosen because beginning 1962 the 10 year T-bond was recorded more frequently, and the Federal reserve system adopted drastic changes in response to the Global Financial Crisis (GFC) of 2008 [2, 43]. Figure 7.8 contains a plot of the two sequences overlaid. It is believed that in general the rates move together with a 'parallel shift'. That is, a a change in the explanatory variable corresponds to a change in the response variable in the same direction with equal strength. We construct the following model,

$$y_t = \beta_0 + \beta_1 x_t + w_t,$$

where $w_t$ is an unknown error process. We test the hypothesis $H_0 : \beta_1 = 1$ against the alternative hypothesis that $H_A : \beta_1 \neq 1$.

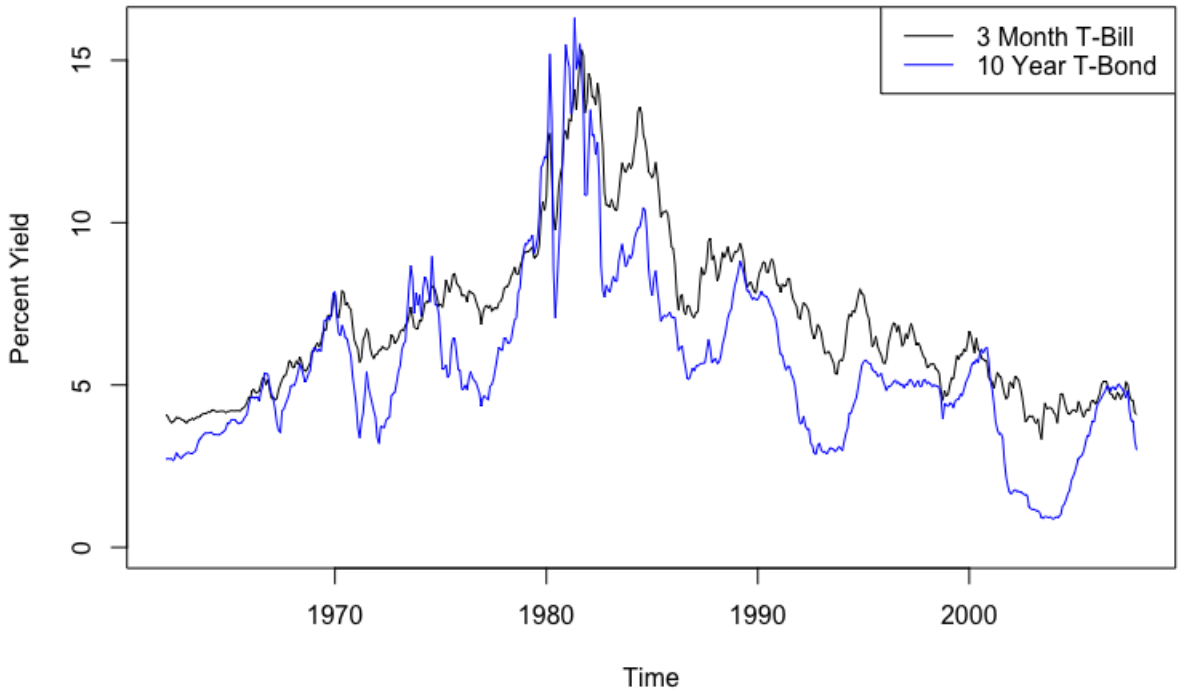


Figure 7.8: 10 Year United States Treasury Bond Yield and 3 Month United States Treasury Bill Yield from January 1962 to December 2007.

We present our results in a similar manor as McElroy & Politis [35]. The test statistic, calculations, and testing procedures are similar to those in Section 7.3. We also continue to use the Bartlett kernel and a significance level of $\alpha = 0.05$.

| Lugsail Type | CV Type | Bandwidth | $F_{OLS}$ | Lower | Upper |
|---|---|---|---|---|---|
| Zero | Fixed-$b$ | $b_{opt} = 0.0682$ | 3.7545 | 0.6329 | 1.0117 |
| Zero | Fixed-$b$ | $b_{ft} = 0.0833$ | 3.7846 | 0.6195 | 1.0252 |
| Zero | Small-$b$ | $b_{ft} = 0.0833$ | 3.7846** | 0.6610 | 0.9837 |
| Mother | Small-$b$ | $b_{mse} = 0.0919$ | 4.6339** | 0.6606 | 0.9841 |

Table 7.7: Various testing statistics of interest for four different testing procedures. Columns indicate the respective kernels and critical values used. Under the given constructions we produce the generated bandwidth, test statistic, and bounds of a 95% confidence interval for $\beta_1$. Test statistics denoted with * indicate significance at the 5% level, and ** indicate significance at the 1% level. The estimated autocorrelation coefficient is $\hat{\rho} = 0.924$, and coefficient is $\hat{\beta}_1 = 0.822$.

The testing results are consistent with the simulation study in the previous section. The largest test statistic corresponds to the mother kernel which is caused by the negative bias of the LRV estimator. Furthermore, the confidence interval is narrow under small-$b$ asymptotics, as it does not account for the estimation of the LRV. The zero lugsail estimator under fixed-$b$ asymptotics fail to reject the null hypotheses which corresponds to expert opinion [2, 43]. Although this cannot be used as validation for the procedure, it is encouraging.

## 7.5 Discussion

The beginning of this chapter posed three questions. We attempted to address them all throughout, albeit to different degrees. In Section 7.2 we observe the behavior of Type 1 error, Type 2 error, and power. The fixed-$b$ asymptotic framework accounts for the variability that causes the Type 1 error rate to inflate as $b$ increases despite $\hat{\Omega}_T$ being less biased. This is evident by the dashed lines in Figures 7.4 - 7.7. The zero lugsail estimator in contrasts addresses the asymptotic bias and has smaller size distortion in comparison

to the mother lugsail settings at the same bandwidth, as expected. The testing mechanics work well as they address different issues. The Type 2 error curve is not strictly concave like the asymptotic expression, but has a concave like behavior. The power is also less for zero lugsail estimators then mother estimators as expected.

In regards to procedure mechanics our recommended procedure was typically optimal or near optimal as illustrated in Section 7.3. The largest differences in performance typically occurred at higher correlations, and for the AR1-HET and ARMA-L models. This suggests that the recommended procedure is comparatively more robust to complex error structures. In Section 7.2 we observed that for nearly any set bandwidth the recommended procedure yields a smaller size distortion. However, we select $b$ specific to the estimator and setting so a comparison of this sort is not as informative. In Sections 7.2 and 7.3 we presented the $b_{ft}$ optimal value for comparison. This rule is intended for flat top estimators and generally yields similar results to $b_{opt}$, with $b_{opt}$ having slightly better performance. The bandwidth rule $b_{opt}$ has an advantage that it is designed to work for all zero lugsail kernels, versus $b_{ft}$ is only immediately applicable to the Bartlett zero lugsail kernel.

We only considered the zero and mother kernel lugsail settings and leave additional settings for future work. There is a distinct and noteworthy advantage in regards to Type 1 error when using a zero lugsail estimator. The cost of this improvement is felt in power loss, which makes selecting an appropriate bandwidth imperative. The bandwidth rule we propose lands at a conservative point just past the elbow and is specific to the testing scenario.

# Bibliography


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

[2] Yoosoon Chang, Ye Lu, and Joon Park. Understanding regressions with observations collected at high frequency over long span. 2021.

[3] Francisco Cribari-Neto and Gauss M Cordeiro. On bartlett and bartlett-type corrections francisco cribari-neto. *Econometric reviews*, 15(4):339–367, 1996.

[4] Halim Damerdji. Strong consistency and other properties of the spectral variance estimator. *Management Science*, 37(11):1424–1440, 1991.

[5] Wouter J Den Haan and Andrew T Levin. 12 a practitioner's guide to robust covariance matrix estimation. *Handbook of statistics*, 15:299–342, 1997.

[6] Graham Elliott, Thomas J Rothenberg, and James H Stock. Efficient tests for an autoregressive unit root, 1992.

[7] James M Flegal and Galin L Jones. Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38(2):1034–1070, 2010.

[8] Gerald B Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.

[9] Ulf Grenander, Murray Rosenblatt, and Nelson Blachman. Statistical analysis of stationary time series. *Physics Today*, 10(6):47, 1957.

[10] Alastair R Hall. Covariance matrix estimation and the power of the overidentifying restrictions test. *Econometrica*, 68(6):1517–1527, 2000.

[11] James Douglas Hamilton. *Time series analysis*. Princeton university press, 1994.

[12] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society*, pages 1029–1054, 1982.

[13] Geoffrey W Hill and AoW Davis. Generalized asymptotic expansions of cornish-fisher type. *The Annals of Mathematical Statistics*, 39(4):1264–1273, 1968.

[14] Masayuki Hirukawa. Robust covariance matrix estimation in time series: A review. *Econometrics and Statistics*, 2021.

[15] Glyn Holton. Cornish-fisher-expansion-value-at-risk: Theory and practice, Jul 2016.

[16] Jungbin Hwang and Yixiao Sun. Asymptotic f and t tests in an efficient gmm setting. *Journal of econometrics*, 198(2):277–295, 2017.

[17] Rustam Ibragimov and Ulrich K Müller. t-statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics*, 28(4):453–468, 2010.

[18] Jack Indritz. *Methods in analysis*. Macmillan, 1963.

[19] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[20] Roger Jang. Matrix inversion lemma, May 2023. Online Notes.

[21] Richard Arnold Johnson, Dean W Wichern, et al. *Applied multivariate statistical analysis*. Prentice hall Upper Saddle River, NJ, 2002.

[22] Galin L Jones. On the Markov chain central limit theorem. *Probability surveys*, 1:299–320, 2004.

[23] VI Khokhlov. The uniform distribution on a sphere in \bfr^s. properties of projections. i. *Theory of Probability & Its Applications*, 50(3):386–399, 2006.

[24] Nicholas M Kiefer and Timothy J Vogelsang. Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory*, 18(6):1350–1366, 2002.

[25] Nicholas M Kiefer and Timothy J Vogelsang. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21(6):1130–1164, 2005.

[26] Gregory F Lawler. *Introduction to stochastic processes*. CRC Press, 2018.

[27] Eben Lazarus, Daniel J Lewis, and James H Stock. The size-power tradeoff in har inference. *Econometrica*, 89(5):2497–2516, 2021.

[28] Eben Lazarus, Daniel J Lewis, James H Stock, and Mark W Watson. Har inference: kernel choice, size distortions, and power losses, 2016.

[29] Eben Lazarus, Daniel J Lewis, James H Stock, and Mark W Watson. Har inference: Recommendations for practice. *Journal of Business & Economic Statistics*, 36(4):541–559, 2018.

[30] Wei-Ming Lee and Chung-Ming Kuan. Testing over-identifying restrictions without consistent estimation of the asymptotic covariance matrix. *Available at SSRN 1555827*, 2009.

[31] Yoong-Sin Lee and Ting-Kwong Lin. Algorithm as 269: High order cornish-fisher expansion. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(1):233–240, 1992.

[32] Ying Liu and J Flegal. Optimal mean squared error bandwidth for spectral variance estimators in MCMC simulations. *ArXiv e-prints*, 2018.

[33] Ying Liu, Dootika Vats, and James M Flegal. Batch size selection for variance estimators in MCMC. *Methodology and Computing in Applied Probability*, pages 1–29, 2021.

[34] Helmut Lütkepohl. *Introduction to multiple time series analysis*. Springer Science & Business Media, 1991.

[35] Tucker McElroy and Dimitris N Politis. Estimating the spectral density at frequencies near zero. *Journal of the American Statistical Association*, pages 1–29, 2022.

[36] Timothy L McMurry and Dimitris N Politis. Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis*, 31(6):471–482, 2010.

[37] Timothy L McMurry and Dimitris N Politis. High-dimensional autocovariance matrices and optimal linear prediction. 2015.

[38] Marc S Meketon and Bruce Schmeiser. Overlapping batch means: Something for nothing? Technical report, Institute of Electrical and Electronics Engineers (IEEE), 1984.

[39] Ulrich K Müller. A theory of robust long-run variance estimation. *Journal of Econometrics*, 141(2):1331–1352, 2007.

[40] Ulrich K Müller. Hac corrections for strongly autocorrelated time series. *Journal of Business & Economic Statistics*, 32(3):311–322, 2014.

[41] Whitney K Newey and Kenneth D West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, 1986.

[42] Emanuel Parzen. On consistent estimates of the spectrum of a stationary time series. *The Annals of Mathematical Statistics*, pages 329–348, 1957.

[43] Daniel F Pellatt and Yixiao Sun. Asymptotic f test in regressions with observations collected at high frequency over long span. *Journal of Econometrics*, 2022.

[44] Peter CB Phillips. Hac estimation by automated regression. *Econometric Theory*, 21(1):116–142, 2005.

[45] Peter CB Phillips and Steven N Durlauf. Multiple time series regression with integrated processes. *The Review of Economic Studies*, 53(4):473–495, 1986.

[46] Peter CB Phillips, Yixiao Sun, and Sainan Jin. Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. *International Economic Review*, 47(3):837–894, 2006.

[47] Dimitris N Politis. Adaptive bandwidth choice. *Journal of Nonparametric Statistics*, 15(4-5):517–533, 2003.

[48] Dimitris N Politis and Joseph P Romano. Bias-corrected nonparametric spectral estimation. *Journal of time series analysis*, 16(1):67–103, 1995.

[49] Maurice Bertram Priestley. *Spectral analysis and time series*. Academic Press, 1981.

[50] Vijay K Rohatgi and AK Md Ehsanes Saleh. *An introduction to probability and statistics*. John Wiley & Sons, 2015.

[51] Donald B Rubin. Multiple imputation after 18+ years. *Journal of the American statistical Association*, 91(434):473–489, 1996.

[52] Yixiao Sun. Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *Journal of Econometrics*, 164(2):345–366, 2011.

[53] Yixiao Sun. A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator. *The Econometrics Journal*, 16(1):1–26, 2013.

[54] Yixiao Sun. Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. *Journal of Econometrics*, 178:659–677, 2014.

[55] Yixiao Sun and Min Seong Kim. Simple and powerful gmm over-identification tests with accurate size. *Journal of Econometrics*, 166(2):267–281, 2012.

[56] Yixiao Sun and Peter CB Phillips. Optimal bandwidth choice for interval estimation in GMM regression. *Cowles Foundation Discussion Paper*, 2008.

[57] Yixiao Sun and Jingjing Yang. Testing-optimal kernel choice in har inference. *Journal of Econometrics*, 219(1):123–136, 2020.

[58] K Tanaka. Time series analysis–nonstationary and noninvertible distribution theory wiley, new york. *Elliott and Müller, Minimizing the Impact of the Initial Condition on Testing for Unit*, 1996.

[59] George B Thomas, Maurice D Weir, and Joel Hass. Thomas' calculus: Multivariable. 2010.

[60] JM TUKEY. Sampling theory of power spectrum estimates. ln: Symi'. on appl. autocorre. analysis to physical problems. washinton, dc, 1950. *VS Nary Research, NAVE XOS-P-735*, pages 47–67, 1949.

[61] Dootika Vats and James M Flegal. Lugsail lag windows for estimating time-average covariance matrices. *arXiv preprint arXiv:1809.04541*, 2020.

[62] Dootika Vats, James M Flegal, Galin L Jones, et al. Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *Bernoulli*, 24(3):1860–1909, 2018.

[63] Carlos Velasco and Peter M Robinson. Edgeworth expansions for spectral density estimates and studentized sample mean. *Econometric Theory*, pages 497–539, 2001.

[64] Halbert White. *Asymptotic theory for econometricians*. Academic press, 1984.

[65] Jeffrey M Wooldridge. *Introductory econometrics: A modern approach*. Cengage learning, 2015.

[66] Yaqian Zhang, Jacek Mańdziuk, Chai Hiok Quek, and Boon Wooi Goh. Curvature-based method for determining the number of clusters. *Information Sciences*, 415:414–428, 2017.

[67] Eric Zivot and Jiahui Wang. *Modeling financial time series with S-Plus®*, volume 191. Springer Science & Business Media, 2007.