# A General Test for Functional Inequalities

Jia Li[*] Zhipeng Liao[†] Wenyu Zhou[‡]

September 17, 2022

**Abstract**

This paper develops a nonparametric test for general functional inequalities that include conditional moment inequalities as a special case. It is shown that the test controls size uniformly over a large class of distributions for observed data, importantly allowing for general forms of time series dependence. New results on uniform growing dimensional Gaussian coupling for general mixingale processes are developed for this purpose, which readily accommodate most applications in economics and finance. The proposed method is applied in a portfolio evaluation context to test for "all-weather" portfolios with uniformly superior conditional Sharpe ratio functions.

**Keywords**: conditional moment inequality; functional inference; Sharpe ratio; series estimation; uniform validity.
**JEL Codes**: C12, C14, C22, C52, G11.

## 1 Introduction

This paper concerns the evaluation of decisions such as forecasting methods, trading strategies, or treatments. Measuring the state-dependent or characteristic-dependent performance of such decisions as functions of certain conditioning state variables or individual characteristics, we propose a

[*]School of Economics, Singapore Management University, Singapore; e-mail: jiali@smu.edu.sg.
[†]Department of Economics, UCLA, Los Angeles, CA 90095; e-mail: zhipeng.liao@econ.ucla.edu.
[‡]International Business School, Zhejiang University, Haining, Zhejiang 314400, China; e-mail: wenyuzhou@intl.zju.edu.cn.

general functional inequality test to carry out the formal conditional evaluation. The proposed test is uniformly valid across a large class of data distributions, importantly accommodating general forms of time series dependence.

Our analysis builds on the large and burgeoning literature in econometrics on moment inequality testing; see Canay and Shaikh (2017) for a comprehensive review. Such tests are routinely used to compare the performance of asset portfolios and forecasting methods (see, e.g., White (2000) and Hansen (2005)). More generally, inference on moment inequalities also arises from partially identified models that are commonly seen in many areas of applied economics; see Molinari (2020) for a comprehensive review. As noted by Imbens and Manski (2004), it is important to ensure that the inference procedure is uniformly valid for large classes of data distributions. The related theoretical analysis is pioneered by Romano and Shaikh (2008) and Andrews and Guggenberger (2009), who establish the first uniformity results for unconditional moment inequality tests, which are further improved by Andrews and Soares (2010), Andrews and Barwick (2012), and Romano, Shaikh, and Wolf (2014), among others.

A more challenging problem pertains to the testing of conditional moment inequalities, which is a special case of functional inequalities, and so, more directly related to our study. The leading methods for testing conditional moment inequalities include the instrumental variable approach of Andrews and Shi (2013) and the nonparametric approach of Chernozhukov, Lee, and Rosen (2013). Chernozhukov, Lee, and Rosen (2013) do not discuss the uniformity issue.[1] Andrews and Shi prove that their test is uniformly valid under general primitive conditions for random samples. They also argue that a similar result holds for the time series setting under a high-level condition, but it is unclear whether that condition can be verified under general forms of dependence; see Section 2.3 below for a more detailed discussion. Moreover, the instrumental variable approach is largely limited to the conditional moment inequality setting, because integrals of general unknown functions (that are not instrumented conditional moments) cannot be directly estimated using unconditional moments. Andrews and Shi's uniformity theory is thus not directly applicable for the general functional inequality setting considered here.

To the best of our knowledge, the present paper is the first that establishes a uniformly valid test for general functional inequalities while accommodating general forms of time series dependence. Our proposal is inspired by the seminal work of Romano, Shaikh, and Wolf (2014), who develop a uniformly valid, practical, yet powerful two-step test for unconditional moment inequalities. We consider a maximum test statistic that estimates the largest violation of the functional

[1] The time series extension developed by Li, Liao, and Quaedvlieg (2022) only concerns a fixed distribution, and so, is also silent on uniform validity.

inequalities implied by the null hypothesis. As is typical for inequality testing, the main challenge for pinning down a uniformly valid critical value stems from the presence of a (functional) nuisance "slackness" parameter. To address this issue, we construct in a first step a uniform upper confidence bound for the functional nuisance parameter and use it to construct a bound for (a coupling-based approximation of) the test statistic. The quantile of the bounding variable can be consistently estimated in a second step, which we use as the critical value.

In connection to prior work, our test may be viewed as a functional version of the two-step test of Romano, Shaikh, and Wolf (2014). The present study concretely demonstrates that Romano et al.'s testing strategy is broadly applicable for a substantively wider range of applications than what is recognized in prior work. When specialized in the setting with conditional moment inequalities, the proposed test offers an alternative way of handling the nuisance parameter problem, which is distinct from the generalized moment selection employed by Andrews and Shi (2013) and, in a similar spirit, the adaptive inequality selection used by Chernozhukov, Lee, and Rosen (2013). The said distinction is analogous to that between Romano, Shaikh, and Wolf (2014) and Andrews and Soares (2010) in the context of testing unconditional moment inequalities.

The key technical ingredient for establishing the uniform validity of the proposed two-step test is to characterize the asymptotic type I error probability in each step. The underlying non-Donsker type functional inference generally relies on a growing dimensional Gaussian coupling theory and the uniformity statement particularly requires the said coupling to hold uniformly across relevant classes of data distributions. Chernozhukov, Chetverikov, and Kato (2013, 2017) and Chernozhukov, Chetverikov, Kato, and Koike (2022) develop a general coupling theory for independent data, and Chernozhukov, Chetverikov, and Kato (2019) develop similar results allowing for a specific $\beta$-mixing type of serial dependence. It is well known, however, that the mixing concept is rather restrictive, particularly for applications in economics and finance. Indeed, it generally cannot accommodate the "simple" martingale difference sequence (MDS), which has zero autocorrelation but may be highly persistent through higher order conditional moments; this is undesirable because the MDS is the key building block for modeling asset prices and also routinely arises from dynamic rational expectation models. In addition, Andrews (1984, 1985) show that a simple nearly i.i.d. autoregressive process can fail to be $\alpha$-mixing, and so, cannot fulfill the more restrictive $\beta$, $\rho$, or $\phi$-mixing requirements, either. In order to accommodate a broad range of economic applications, we thus develop a uniform Gaussian coupling theory for the so-called mixingale class (McLeish, 1975; Andrews, 1988), which, as suggested by its name, captures the distinct martingale-type and mixing-type dependencies as special cases and is in fact substantially more general. This novel result strengthens the pointwise (i.e., for a fixed distribution) coupling theory developed in a prior paper (Li and Liao, 2020). It may also be useful for establishing uniformity results in the other inferential settings for time series applications; see Fang and Seo (2021) for an interesting example on testing shape restrictions.[2]

An important empirical motivation for studying general functional inequalities—beyond the special case with conditional moment inequalities—is to evaluate the performance of asset portfolios. The state-dependent performance of a portfolio may be measured by its conditional Sharpe ratio, namely the ratio between the conditional mean and standard deviation functions of the portfolio's excess return. Portfolio evaluation is of fundamental importance for both academic finance and investment practice. The proposed functional inference complements the conventional unconditional evaluation methods from a conditional perspective. The associated econometric setting is also pedagogically interesting because the econometric analysis for the conditional Sharpe ratio function naturally involves shape-restricted nonparametric estimation and nonparametrically generated variables, which are not commonly encountered in the baseline problem of testing conditional moment inequalities. For these reasons, we study this special case in greater details in Section 3, examine the resulting test's finite-sample performance via simulations in Section 4, and provide an empirical application using twelve distinct datasets from the U.S. equity market in Section 5.

The remainder of this paper is organized as follows. In Section 2, we describe the econometric setting for the general functional inequality testing problem, propose the two-step test, and establish its uniform asymptotic validity. In Section 3, we further focus on a more concrete setting for testing functional inequalities defined by conditional Sharpe ratios. Simulation and empirical results are presented in Sections 4 and 5, respectively. Section 6 concludes. The Appendix collects the proofs for our main theoretical results. The Supplemental Appendix collects additional technical results including particularly our new uniform Gaussian coupling theory for mixingale processes. The Supplemental Appendix also contains additional simulation results and empirical robustness checks.

[2]Zhang and Wu (2017) and Zhang and Cheng (2018) also develop a coupling theory for a class of time series models proposed by Wu (2005) under a notion of "physical dependence." A key restriction for this class of models, however, is that the observed time series has to be represented as a nonlinear function of i.i.d. shocks. Although assuming shocks to be i.i.d. might be reasonable in certain physical systems, which motivates Wu's (2005) original proposal, it appears to be unnatural in an economic context because the notion of (unexpected) economic shocks pertains to the economic agent's information filtration, and so, is better captured by an MDS. As a simple example, consider stock returns modeled as increments of the stochastic integral of a stochastic volatility process with respect to a Brownian motion; this is a workhorse MDS model in financial econometrics but does not fulfill the representation required by Wu (2005). Moreover, as acknowledged in Wu (2005), the said representation requirement does not nest mixing processes, as the latter do not require such a representation.

The following notation will be adopted throughout the paper. All limits are for $n \to \infty$, with $n$ denoting the sample size. We use $\|\cdot\|$ and $\|\cdot\|_S$ to denote the matrix Frobenius norm and spectral norm, respectively. For any real symmetric matrix $A$, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote its smallest and largest eigenvalues, respectively. We use $\mathcal{P}_n$ to denote a class of distributions that govern the observed data and use $P$ to denote a generic element in $\mathcal{P}_n$. For any random sequence $X_n$ and positive real sequence $v_n$, we write $X_n = O_{\mathcal{P}}(v_n)$ if $\lim_{M\to\infty} \limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} P\left(|X_n| \geq M v_n\right) = 0$, and $X_n = o_{\mathcal{P}}(v_n)$ if for any $\epsilon > 0$, $\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} P\left(|X_n| \geq \epsilon v_n\right) = 0$.

# 2 A Test for General Functional Inequalities

This section presents our main theory for testing functional inequalities. Section 2.1 introduces the testing problem and discusses its empirical relevance. Section 2.2 details the testing procedure, with its theoretical properties presented in Section 2.3.

## 2.1 Hypotheses of interest and motivating examples

We start with introducing the hypothesis to be tested. Our development is mainly motivated by the task of evaluating the relative state-dependent or characteristic-dependent performance of competing *decisions*, which may be forecasts, trading strategies, or medical/economic "treatments" depending on the applied scenario. We index the decisions of interest by a finite set $\mathcal{J} \equiv \{1, \ldots, J\}$. For each $j \in \mathcal{J}$, the corresponding decision's state-dependent performance is summarized by a (possibly vector-valued) function $h_j(\cdot) : \mathcal{X} \mapsto \mathbb{R}^d$, where the set $\mathcal{X}$ is assumed to be compact for technical convenience. The evaluation inference is to test a null hypothesis with the form

$$H_0 : \max_{j\in\mathcal{J}} \sup_{x\in\mathcal{X}} \ell(h_j, x) \leq 0, \tag{2.1}$$

where the mapping $h_j \mapsto \ell(h_j, \cdot)$ specifies how the information encoded in $h_j(\cdot)$ may be further transformed into an evaluation criterion. The $h_j(\cdot)$ functions are unknown but nonparametrically estimable. The mapping $\ell(\cdot,\cdot)$ is known, as it represents the evaluation criterion specified by the user.

An important special case of the problem under study pertains to moment inequality testing, which is the dominant econometric method for the empirical analysis of partially identified models; see Canay and Shaikh (2017) and Molinari (2020) for comprehensive reviews. When $h_j(\cdot)$ is a conditional mean function and $\ell(h_j, x) = h_j(x)$, the null hypothesis in (2.1) reduces to what has been extensively studied in the conditional moment inequality literature (Andrews and Shi, 2013, 2017; Chernozhukov, Lee, and Rosen, 2013; Lee, Song, and Whang, 2013; Armstrong, 2015;

Chetverikov, 2018; Li, Liao, and Quaedvlieg, 2022). If the conditioning information set is further assumed to be trivial, the testing problem pertains to unconditional moment inequalities (Romano and Shaikh, 2008; Andrews and Guggenberger, 2009; Andrews and Soares, 2010; Andrews and Barwick, 2012; Romano, Shaikh, and Wolf, 2014; Cox and Shi, 2019).

While moment inequalities routinely arise in partially identified economic models, we recognize that focusing solely on them may be restrictive for the purpose of evaluating decisions, in that a researcher's evaluation criterion may not be based on moment conditions per se. To better appreciate this point, we consider some examples.

Example 1 (Portfolio Evaluation via State-Dependent Sharpe Ratio). Consider a collection $\mathcal{J}$ of trading strategies. Let the excess return of strategy $j \in \mathcal{J}$ over the period $[t, t+1]$ be denoted by $r_{j,t+1}$. It is standard to gauge the performance of a trading strategy by its Sharpe ratio, defined as the ratio between the expected excess return (i.e., risk premium) and the standard deviation. Since the conditional mean and standard deviation of excess returns are generally state-dependent, so is the Sharpe ratio. Given a conditioning state variable $X_t$, we denote the conditional mean and variance functions as $\mu_j(x) = \mathbb{E}[r_{j,t+1}|X_t = x]$ and $\sigma_j^2(x) = Var(r_{j,t+1}|X_t = x)$. The conditional Sharpe ratio for strategy $j$ at a given state $x \in \mathcal{X}$ may then be written as

$$\mathrm{SR}_j(x) \equiv \frac{\mu_j(x)}{\sigma_j(x)}. \tag{2.2}$$

It is empirically relevant to test whether a "benchmark" strategy, indexed below by $j = 0$, weakly dominates strategy $j$ across all conditioning states, namely,

$$\mathrm{SR}_0(x) \geq \mathrm{SR}_j(x) \text{ for all } x \in \mathcal{X}. \tag{2.3}$$

Setting $h_j(\cdot) = \mathrm{SR}_j(x) - \mathrm{SR}_0(x)$ and $\ell(h_j, x) = h_j(x)$, we may rewrite the uniform weak dominance condition displayed in (2.3) equivalently as

$$\sup_{x \in \mathcal{X}} \ell(h_j, x) \leq 0. \tag{2.4}$$

The hypothesis that the benchmark uniformly weakly dominates all competing strategies in $\mathcal{J}$ can then be succinctly stated as (2.1). □

Example 2 (Portfolio Evaluation via Modified Sharpe Ratio). While the classical Sharpe ratio considered in Example 1 is mainly motivated in a Gaussian model with risk measured by standard deviation, a practitioner may pay more attention to downside risk that is commonly measured by the Value-at-Risk (VaR), say, at the 95% confidence level. The VaR is formally defined as the conditional quantile of the trading loss (i.e., $-r_{j,t+1}$) given the state variable. Using the state-dependent VaR as a measure of risk, one may modify the Sharpe ratio of strategy $j$ as

$$\mathrm{SR}'_j(x) \equiv \frac{\mu_j(x)}{\mathrm{VaR}_j(x)}.$$

Portfolio evaluation according to this modified Sharpe ratio may be cast as testing (2.1), with $h_j(\cdot) = \mathrm{SR}'_j(x) - \mathrm{SR}'_0(x)$ and $\ell(h_j, x) = h_j(x)$. $\square$

In these examples, the functional objects of direct inferential interest are not conditional mean functions, in that the classic Sharpe ratio and its VaR-modified version involve nonlinear transformations of conditional moment functions and/or conditional quantile functions. We also note that for this type of applications, it is essential to accommodate *general* forms of serial dependence in the data, particularly including the martingale-type behavior in asset prices and the high persistence in stochastic volatility. These considerations lead to nontrivial theoretical challenges for establishing the uniform validity of the proposed test, which is a key component of our technical contribution.

Deviating from the conditional moment inequality framework has a direct implication on the choice of testing strategy. There are two parallel strategies in the existing literature on conditional moment inequality testing. The nonparametric approach of Chernozhukov, Lee, and Rosen (2013) is based on nonparametric estimators of the unknown functions and relies on non-Donsker-type functional inference. In contrast, Andrews and Shi (2013) transform conditional moment inequalities into an infinite collection of unconditional moment inequalities via "many" instruments (also see Chernozhukov, Chetverikov, and Kato (2019) and Bai, Santos, and Shaikh (2022)); they then test the unconditional ones without nonparametrically estimating the original conditional moment functions. However, the instrumental variable approach is highly specific to the conditional moment inequality setting, because integrals of general unknown functions (that are not instrumented conditional moments) cannot be directly estimated using unconditional moments. Therefore, the nonparametric approach is a more natural choice for the general functional inequality testing problem studied here.

While the main theme of the present paper is to address functional inequality testing under general forms of data dependence, we recognize that studying the general null hypothesis (2.1) may shed new light on certain applied microeconomic problems. The following example illustrates such a possibility.

Example 3 (Risk-Adjusted Treatment Effect). Consider a collection $\mathcal{J}$ of treatments and index the benchmark treatment by $j = 0$. Let $Y_{j,i}$ be the potential outcome of treatment $j$ on individual $i$. Since the expected outcome may depend on observed individual characteristics (e.g., age), it is natural to gauge the effectiveness of a treatment using the conditional mean function $\mu_j(x) = \mathbb{E}[Y_{j,i}|X_i = x]$, where $X_i$ denotes the individual characteristic of interest. In a baseline setting with $h_j(\cdot) = \mu_j(\cdot) - \mu_0(\cdot)$ and $\ell(h_j, x) = h_j(x)$, the hypothesis (2.1) is equivalent to $\mu_0(x) \geq \mu_j(x)$ for all $x \in \mathcal{X}$ and $j \in \mathcal{J}$, which asserts that the benchmark treatment weakly dominates all competitors across all subpopulations "categorized" by the conditioning individual characteristic. This baseline inference problem may be further generalized in the directions discussed in Examples 1 and 2. Indeed, if a treatment is "risky," such as a vaccine with potentially severe side effects, it appears natural to gauge its riskiness using the conditional standard deviation. A risk-adjusted average treatment might be constructed in a similar way as the Sharpe ratio discussed in Example 1 and the treatments can be evaluated accordingly. □

In summary, these illustrative examples demonstrate the empirical relevance of testing the functional inequality hypothesis (2.1), which naturally generalizes the important problem of testing moment inequalities. They also highlight the necessity of accommodating general serial dependence in the data so as to afford a broad range of empirical applications particularly for macroeconomics and finance. We have explained why the nonparametric approach suits the present general inference problem better than the instrumental variable approach. We next propose a nonparametric test.

## 2.2 The test

We now describe our nonparametric test and provide some intuition for its construction. For concreteness, we suppose that the unknown $h_j(\cdot)$ functions can be estimated using a series approach, including but not restricted to the series regression (Andrews, 1991; Newey, 1997). Specifically, each $h_j(\cdot)$ function admits a series approximation $h_j^*(\cdot) \equiv \beta_j^{*\top}\psi(\cdot)$, where $\psi(\cdot)$ is a $k$-dimensional vector of approximating functions (e.g., orthogonal polynomials, splines, or wavelets) and $\beta_j^*$ is the unknown "population" coefficient.[3] Note that the $h_j(\cdot)$ functions may be vector-valued and the $\beta_j^*$ coefficients are matrices in general. The nonparametric interpretation under the series approach is obtained by letting the number of series terms $k \to \infty$ as the sample size $n \to \infty$. The series estimator for $h_j(\cdot)$ has the form $\hat{h}_j(\cdot) \equiv \hat{\beta}_j^\top \psi(\cdot)$.[4] For ease of our subsequent discussion, we stack the

[3]For ease of exposition, we suppose that a common approximation basis $\psi(\cdot)$ is used for nonparametrically estimating all $h_j(\cdot)$ functions. This setting can be trivially extended so as to allow the approximating basis to differ across $j$'s only at the cost of more complicated notation.

[4]For any $j \in \mathcal{J}$, $h_j(\cdot)$ could be a vector of unknown functions. In this case, both $\beta_j^*$ and $\hat{\beta}_j$ are real matrices of the same dimensions. The series estimator presented here allows for general semi/nonparametric multi-step estimation, where the components in $h_j(\cdot)$ are estimated in different steps.

estimated and population coefficients into "long" $m$-dimensional vectors as $\hat{\beta} \equiv (\text{vec}(\hat{\beta}_j))_{j\in\mathcal{J}}$ and $\beta^* \equiv (\text{vec}(\beta_j^*))_{j\in\mathcal{J}}$, respectively, and define $L_j$ as the selection matrix such that $\text{vec}(\hat{\beta}_j) = L_j\hat{\beta}$. We use $\Omega$ to denote the variance-covariance matrix of $n^{1/2}(\hat{\beta} - \beta^*)$ and assume that it can be estimated by a symmetric positive definite matrix $\widehat{\Omega}$. The requisite technical conditions are detailed in the theory Section 2.3 below.

Series estimation is easy to implement in practice as it typically only involves running linear least-squares or quantile regressions and the estimator $\widehat{\Omega}$ can be constructed post estimation using standard "textbook" formulas. Of course, an important theoretical complication for justifying those formulas stems from the fact that the dimensionality of the approximating function $\psi(\cdot)$ diverges asymptotically, and so, the associated asymptotic analysis requires probabilistic tools for growing dimensional Gaussian approximation (Chernozhukov, Chetverikov, and Kato, 2013, 2014, 2019; Chernozhukov, Chetverikov, Kato, and Koike, 2022; Chernozhukov, Lee, and Rosen, 2013; Li and Liao, 2020; Li, Liao, and Quaedvlieg, 2022). The theoretical analysis in this paper builds on and nontrivially extends this line of results.

We construct the test statistic as follows. Let $\dot{\ell}$ denote the derivative of the mapping $\beta \mapsto \ell(\psi(\cdot)^\top \beta, x)$ with respect to $\text{vec}(\beta)$. For each $j \in \mathcal{J}$, we estimate the standard deviation of $n^{1/2}(\ell(\hat{h}_j, x) - \ell(h_j, x))$ by

$$\hat{s}_j(x) \equiv \sqrt{\dot{\ell}(\hat{h}_j, x)^\top L_j \widehat{\Omega} L_j^\top \dot{\ell}(\hat{h}_j, x)}. \tag{2.5}$$

The test statistic for the null hypothesis in (2.1) is defined as

$$\widehat{T}_n \equiv \max\left\{\max_{j\in\mathcal{J}} \sup_{x\in\mathcal{X}} \frac{n^{1/2}\ell(\hat{h}_j, x)}{\hat{s}_j(x)}, 0\right\}, \tag{2.6}$$

which is a one-sided sup-t statistic truncated at zero. The null hypothesis will be rejected when $\widehat{T}_n$ is "sufficiently large" relative to some critical value.

The key step is to properly determine the critical value. It should ensure that the test controls size uniformly across a large class of data generating processes for the observed data and, subject to this requirement, exhibits good power properties. In the context of testing unconditional moment inequalities with independent data, one of the state-of-the-art methods that achieve these goals is the two-step Bonferroni-type procedure developed by Romano, Shaikh, and Wolf (2014). Our test proposed below extends this approach by further accommodating (i) dependent data and (ii) functional inequalities. Since the present functional inequality setting includes the conditional moment inequality setting as a special case, a by-product of our extension is a new test for conditional moment inequalities, which complements the existing approaches of Andrews and Shi (2013) and Chernozhukov, Lee, and Rosen (2013).

It is instructive to explain heuristically the logic under each step of the test construction. To begin with, we observe that the pointwise (i.e., for a given conditioning state $x$) t-statistic may be decomposed as

$$\frac{n^{1/2}\ell(\hat{h}_j,x)}{\hat{s}_j(x)} = \frac{n^{1/2}(\ell(\hat{h}_j,x)-\ell(h_j,x))}{\hat{s}_j(x)} + \frac{n^{1/2}\ell(h_j,x)}{\hat{s}_j(x)}, \tag{2.7}$$

where the two terms on the right-hand side of (2.7) capture the statistical estimation error and the underlying "signal," respectively. Under proper regularity conditions, we can show that the statistical error can be uniformly "coupled" (in a strong sense) by a Gaussian process, that is,

$$\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}(\ell(\hat{h}_j,x)-\ell(h_j,x))}{\hat{s}_j(x)} - \frac{\mathbb{G}_{j,n}(x)}{s_j(x)}\right| = o_{\mathcal{P}}(\log(n)^{-1}),$$

where the $o_{\mathcal{P}}$ statement holds uniformly across a large class of distributions $\mathcal{P}_n$ and $\mathbb{G}_{j,n}$ is a centered Gaussian process with its standard deviation function given by

$$s_j(x) \equiv \sqrt{\dot{\ell}(h_j^*,x)^\top L_j \Omega L_j^\top \dot{\ell}(h_j^*,x)}. \tag{2.8}$$

It can also be shown that $\hat{s}_j(x)$ is a valid estimator for $s_j(x)$ in a similar uniform sense. These approximations allow us to further approximate the original test statistic $\widehat{T}_n$ with the following easier-to-analyze "coupling" variable

$$\widetilde{T}_n \equiv \max\left\{\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left(\frac{\mathbb{G}_{j,n}(x)}{s_j(x)} + \frac{n^{1/2}\ell(h_j,x)}{s_j(x)}\right), 0\right\}. \tag{2.9}$$

Equipped with this coupling result, the task of determining a "good" critical value essentially boils down to designing a "good" feasible approximation for the distribution of $\widetilde{T}_n$. Since $\mathbb{G}_{j,n}(\cdot)$ is known to be a Gaussian process and its covariance function is consistently estimable, there is no fundamental difficulty in capturing the distributional property of the $\mathbb{G}_{j,n}(\cdot)/s_j(\cdot)$ component in $\widetilde{T}_n$. The main difficulty, however, stems from the $n^{1/2}\ell(h_j,x)/s_j(x)$ term, because the scaled function $n^{1/2}\ell(h_j,x)$ is unknown and cannot be estimated uniformly well. Indeed, while the $\ell(h_j,\cdot)$ function may be estimated at a usual nonparametric rate, the associated estimation error would be magnified substantially by the (large) $n^{1/2}$ scaling factor. The presence of this nontrivial nuisance term is the main obstacle for constructing a feasible distributional approximation for $\widetilde{T}_n$ that holds uniformly across a large class of distributions.

A standard strategy to address this nuisance parameter problem in the context of inequality testing is to instead work with an upper bound for the test statistic (or asymptotically equivalently its coupling variable $\widetilde{T}_n$), so that quantiles of the bounding variable may be used as critical values. Since $\widetilde{T}_n$ is nondecreasing in $\ell(h_j,\cdot)$, bounds for $\widetilde{T}_n$ can be obtained by bounding the $\ell(h_j,\cdot)$

functions. The most straightforward idea, which traces back at least to White (2000), is to invoke the least favorable deterministic bound $\ell(h_j, \cdot) \le 0$ directly implied by the null hypothesis, yielding

$$\widetilde{T}_n \le \max\left\{\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\frac{\mathbb{G}_{j,n}(x)}{s_j(x)}, 0\right\}.$$

Quantiles of this least favorable majorant are valid critical values for size control, but they may be "excessively large" in practice when $\ell(h_j, x)$ is nontrivially below zero for many $(j, x)$'s, which in turn would hurt the power of the test.

It is thus desirable to consider tighter data-driven bounds for $\widetilde{T}_n$. Inspired by the analysis of Romano, Shaikh, and Wolf (2014) and Bai, Santos, and Shaikh (2022) for unconditional moment inequalities, we develop a two-step method in the present functional inequality setting. For a given significance level $\alpha \in (0, 0.5)$, we first pick some "small" constant $\rho \in (0, \alpha)$ and construct a uniform (w.r.t. $j \in \mathcal{J}$, $x \in \mathcal{X}$, and $P \in \mathcal{P}_n$) upper confidence bound, denoted $\hat{u}_j(x)$, for $\ell(h_j, x)$ at confidence level $1 - \rho$. In restriction to the event $\{\ell(h_j, x) \le \hat{u}_j(x)$ for all $j \in \mathcal{J}$ and $x \in \mathcal{X}\}$, we are able to bound $\widetilde{T}_n$ as

$$\widetilde{T}_n \le \max\left\{\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\frac{\mathbb{G}_{j,n}(x) + n^{1/2}\hat{u}_j(x)}{s_j(x)}, 0\right\}. \tag{2.10}$$

To adjust for the fact that $(\hat{u}_j(\cdot))_{j\in\mathcal{J}}$ may fail to bound $(\ell(h_j, \cdot))_{j\in\mathcal{J}}$ up to probability $\rho$ asymptotically, we set the critical value as an estimate of the $\alpha - \rho$ (rather than $\alpha$) upper quantile of the bounding variable in (2.10). As such, the overall false rejection probability from the two steps is bounded by $\rho + (\alpha - \rho) = \alpha$ in large samples. For ease of application and reference, we summarize the testing procedure as the following algorithm.

**Algorithm 1 (Testing for Generic Functional Inequalities)**.
Step 1. Set $\widehat{\mathbb{G}}^*_{j,n}(x) \equiv \dot{\ell}(\hat{h}_j, x)^\top L_j \widehat{\Omega}^{1/2}\mathcal{N}^*_m$, where $\mathcal{N}^*_m$ denotes an $m$-dimensional standard normal random vector independent of the data. Let $\rho \in (0, \alpha)$. Compute $\hat{q}_{1-\rho}$ as the $1-\rho$ conditional quantile of $\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\widehat{\mathbb{G}}^*_{j,n}(x)/\hat{s}_j(x)$ given data and set $\hat{u}_j(x) = \min\{\ell(\hat{h}_j, x) + n^{-1/2}\hat{s}_j(x)\hat{q}_{1-\rho}, 0\}$.
Step 2. Set the critical value $\hat{c}_n$ as the $1 - \alpha + \rho$ data-conditional quantile of

$$\widehat{T}^*_n \equiv \max\left\{\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left(\frac{\widehat{\mathbb{G}}^*_{j,n}(x) + n^{1/2}\hat{u}_j(x)}{\hat{s}_j(x)}\right), 0\right\}. \tag{2.11}$$

Reject the null hypothesis (2.1) when $\widehat{T}_n > \hat{c}_n$. □

We close this subsection with a few remarks on the proposed test. First, the testing procedure described in Algorithm 1 takes the $\hat{h}_j(\cdot)$ and $\widehat{\Omega}$ estimators as given, the construction of which depends on the specific application. Under the baseline setting with $h_j(\cdot)$ being a conditional mean or quantile function, these estimators can be obtained as standard outputs from a least-squares or quantile series regression, and we refer the reader to the broader series estimation literature for further details.[5] Under the more general settings described in Section 2.1, the construction of $\hat{h}_j(\cdot)$ may involve shape restrictions, sequential estimation, and/or generated variables, and $\widehat{\Omega}$ needs to be tailored accordingly so as to account for these complications. As a case in point, the running example on portfolio evaluation (recall Example 1) involves such issues, for which we provide a more detailed analysis in Section 3, below.

Second, the $\rho$ parameter governs how the test's type I error probability is "allocated" between the two steps. Choosing a smaller $\rho$ implies that the critical value is computed as a lower quantile (in step 2) of a larger bounding variable (from step 1), and vice versa. While this trade-off is intuitively clear, the literature does not offer an "optimal" choice for $\rho$. The two-step procedure is originally proposed by Berger and Boos (1994) and Silvapulle (1996) for a variety of inference problems and these authors suggest taking $\rho$ as a small number such as 0.001 or 0.005. In the context of unconditional moment inequality testing, Romano, Shaikh, and Wolf (2014) and Bai, Santos, and Shaikh (2022) recommend setting $\rho = \alpha/10$ based on simulation evidence; this recommendation in fact closely mirrors the suggestions of Berger and Boos (1994) and Silvapulle (1996) for typical significance levels such as $\alpha = 0.01$ or $0.05$. Our simulation results reported in Section 4.2 confirm that picking $\rho$ from the $[0.001, 0.005]$ range is adequate. For these reasons, we also recommend choosing $\rho$ from $[0.001, 0.005]$ and implementing the test for several values of $\rho$ as a way of checking robustness.

It is worth clarifying that, like Romano, Shaikh, and Wolf (2014), we treat the parameter $\rho$ as a fixed constant rather than as a tuning sequence that varies with the sample size. The theoretical validity of the proposed test will be established for any $\rho \in (0, \alpha)$ without any further restriction on this parameter. Yet an alternative theoretical approach is to consider $\rho$ as a tuning sequence $\rho_n = o(1)$. This *additional* tuning assumption on $\rho$ would allow one to invoke an extra layer of asymptotic approximation: Since the type I error probability $\rho_n$ from the first step vanishes asymptotically, the critical value may be "simply" computed in the second step as the $1-\alpha$ quantile of $\widehat{T}_n^*$, instead of the larger $1 - \alpha + \rho_n$ quantile. This alternative asymptotic embedding could be easily accommodated in our theory, but we refrain from pursuing it explicitly so as to avoid the said extra asymptotic approximation.

[5] See, for example, Andrews (1991), Newey (1997), and Chen (2007) for early contributions, and Chernozhukov, Lee, and Rosen (2013), Belloni, Chernozhukov, Chetverikov, and Kato (2015), Li and Liao (2020), and Li, Liao, and Quaedvlieg (2022) for more recent work that focuses more specifically on uniform functional inference.

### 2.3 Uniform asymptotic validity

In this subsection, we show that the proposed test controls size asymptotically under the null hypothesis and is consistent against certain local alternatives. Consider a (large) collection $\mathcal{P}_n$ of distributions that govern the observed data and let $\mathcal{P}_{0,n} \subseteq \mathcal{P}_n$ denote the subset of distributions that satisfy the null hypothesis (2.1). Following the terminology of Romano (2004), we provide conditions under which the proposed test is uniformly consistent in level in the sense that

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_{0,n}} P\left(\widehat{T}_n > \hat{c}_n\right) \leq \alpha,$$

where the asymptotic size $\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_{0,n}} P(\widehat{T}_n > \hat{c}_n)$ embodies uniformity with respect to the data generating process by definition.

The importance of controlling size uniformly has been emphasized by Imbens and Manski (2004) in the partial identification literature; also see Canay and Shaikh (2017) and Molinari (2020) for comprehensive reviews. Romano and Shaikh (2008) and Andrews and Guggenberger (2009) establish the first uniformity results for unconditional moment inequality tests. These results are further improved by Andrews and Soares (2010), Andrews and Barwick (2012), and Romano, Shaikh, and Wolf (2014), among others. The uniformity issue in the "many" unconditional moment inequality setting has been addressed in the recent work of Chernozhukov, Chetverikov, and Kato (2019) and Bai, Santos, and Shaikh (2022). The theory of Andrews and Soares (2010) and Chernozhukov, Chetverikov, and Kato (2019) accommodate certain types of time series dependence in the data. The uniformity issue in the context of testing unconditional moment inequalities is now relatively well understood.

Meanwhile, considerably less is known in the context of testing functional inequalities, especially when the data is allowed to be dependent. The existing literature on functional inequality test mainly pertains to the baseline setting with conditional moment inequalities. As noted in the introduction, the leading uniformity result is developed by Andrews and Shi (2013). The authors show that the instrumental variable approach yields a uniformly valid test under general primitive conditions for independent data; also see Andrews and Shi (2017) for a battery of examples. They also suggest that a similar result obtains for the time series setting under a high-level condition (Assumption EP), which requires the weak convergence of an empirical process indexed by a large collection of instruments (e.g., boxes and cubes on the conditioning space). Unfortunately, it is unclear from the discussion in Andrews and Shi (2013) how this condition may be verified or, more precisely, to which extent it implicitly restricts the data generating process.[6] As previously noted,

[6] Under the independent data setting, Andrews and Shi's Assumption EP can be easily verified using classical

the instrumental variable approach is also specific to the conditional moment inequality setting, and so, Andrews and Shi's uniformity result is largely tangential to the present setting with general functional inequalities.

To the best of our knowledge, our uniformity result presented below is the first in the literature to establish the uniform validity for general functional inequality tests while accommodating general forms of data dependence. Turning to the theoretical details, we start with introducing some regularity conditions.

**Assumption 1.** *Suppose that $k = o(n/\log n)$. Moreover, the following conditions hold for some finite constant $K > 0$ and all $j \in \mathcal{J}$:*

*(i)* $\sup_{P\in\mathcal{P}_n}\sup_{x\in\mathcal{X}} |\ell(h_j, x)| \leq K$*;*

*(ii)* $\sup_{P\in\mathcal{P}_n}\sup_{x_1,x_2\in\mathcal{X}} |\ell(h_j, x_1) - \ell(h_j, x_2)| / \|x_1 - x_2\| \leq K$*;*

*(iii)* $\sup_{P\in\mathcal{P}_n}\sup_{x\in\mathcal{X}} \|\dot{\ell}(h_j^*, x)\| \leq \xi_k$ *for some non-dereasing sequence* $\xi_k$*;*

*(iv)* $\sup_{P\in\mathcal{P}_n}\sup_{x_1,x_2\in\mathcal{X}} \|\dot{\ell}(h_j^*, x_1) - \dot{\ell}(h_j^*, x_2)\| / \|x_1 - x_2\| \leq \zeta_k^L$ *for some* $\zeta_k^L$ *satisfying* $\log(\zeta_k^L) = O(\log(k))$ *and* $\zeta_k^L = o\left(n/\log n\right)$*.*

Assumption 1 collects some regularity conditions on the evaluation criterion $\ell(h_j, x)$. The function is assumed to be bounded and Lipschitz over the compact set $\mathcal{X}$, with its growing dimensional derivative function $\dot{\ell}$ bounded by $\xi_k$ and Lipschitz continuous with coefficient $\zeta_k^L$, where $\xi_k$ and $\zeta_k^L$ are both allowed to be divergent. Since $\ell(h_j, x)$ depends on the $h_j$ function, which in turn is determined by $P$, these conditions are understood as restrictions on the class $\mathcal{P}_n$ of distributions. The conditions are easy to verify. For instance, in the baseline setting with $\ell(h_j, x) = h_j(x)$ (and hence, $\dot{\ell}(h_j^*, x) = \psi(x)$), they coincide with those commonly used for uniform series-based inference (see, e.g., Belloni, Chernozhukov, Chetverikov, and Kato (2015) and Li and Liao (2020)). More generally, the conditions accommodate smooth functional transformations such as derivatives.

Next, we describe the conditions for the functional estimators $(\hat{h}_j(\cdot))_{j\in\mathcal{J}}$ and the variance-covariance estimator $\widehat{\Omega}$. These conditions are high-level in nature and some of them are nontrivial to verify. We present our general results under these conditions to help clarify the econometric mechanism at play, and provide further details on how to verify them under primitive conditions

empirical process theory because the box-type instruments form a Vapnik–Cervonenkis class. The same logic extends trivially to a setting in which the data is finitely dependent. Although not claimed by Andrews and Shi (2013), it is also conceivable that this high-level condition might be verified under certain mixing-type conditions provided that the mixing coefficient decays sufficiently fast, though establishing this formally, if possible, is clearly nontrivial. Nevertheless, mixing-type conditions are well known to be a highly restrictive notion of weak dependence for developing limit theorems in time series econometrics. For example, Andrews (1984, 1985) shows that even simple autoregressive processes that are nearly i.i.d. can fail to be $\alpha$-mixing, let alone fulfilling the other more restrictive mixing conditions such as $\beta$, $\rho$, or $\phi$-mixing (i.e., absolutely regular, completely regular, or uniform mixing).

in Section 3. The conditions resemble those commonly adopted in uniform nonparametric series inference based on growing dimensional Gaussian coupling, but with a more stringent uniformity requirement across a class $\mathcal{P}_n$ of distributions.

**Assumption 2.** *The following conditions hold uniformly over $P \in \mathcal{P}_n$ for some finite constant $K > 0$: (i) there exists a sequence of $m$-dimensional standard Gaussian random vectors $\mathcal{N}_m$ such that*

$$n^{1/2}(\hat{\beta} - \beta^*) = \Omega^{1/2}\mathcal{N}_m + o_{\mathcal{P}}(\log(n)^{-1});$$

*(ii) uniformly over $x \in \mathcal{X}$,*

$$\frac{n^{1/2}(\ell(\hat{h}_j, x) - \ell(h_j, x))}{s_j(x)} = \frac{\dot{\ell}(\hat{h}_j, x)^\top n^{1/2}\mathrm{vec}(\hat{\beta}_j - \beta_j^*)}{s_j(x)} + o_{\mathcal{P}}(\log(n)^{-1});$$

*(iii)* $\max_{j\in\mathcal{J}} \sup_{x\in\mathcal{X}} \|\dot{\ell}(\hat{h}_j, x) - \dot{\ell}(h_j^*, x)\| + \|\widehat{\Omega} - \Omega\|_S = o_{\mathcal{P}}(k^{-1/2}\log(n)^{-1})$*; (iv)* $K^{-1} \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq K$*; (v)* $\inf_{x\in\mathcal{X}} \|\dot{\ell}(h_j^*, x)\| \geq K^{-1}$.

A few remarks on Assumption 2 are in order. Condition (i) states that the growing dimensional regression coefficient $\hat{\beta}$ from the series estimation admits a Gaussian coupling uniformly for all distributions from the $\mathcal{P}_n$ class. This is the main driver of our asymptotic analysis, particularly for establishing the size property of the proposed test. Verifying this condition mainly involves characterizing the influence function of the $\hat{\beta}$ estimator and invoking a strong Gaussian coupling for the related growing dimensional sample moment. In Section SA of the Supplemental Appendix, we provide a uniform (over $\mathcal{P}_n$) Gaussian approximation theory for general mixingale-type dependent data. The mixingale process is proposed by McLeish (1975) and further generalized by Andrews (1988), which forms a very general class of models, including MDS, linear processes, and various types of mixing and near-epoch dependent processes as special cases, and naturally allows for data heterogeneity. This dependence concept is substantially more general than commonly used mixing concepts (e.g., $\alpha$, $\beta$, $\rho$, $\phi$ mixing) and, in particular, is not subject to the critiques of Andrews (1984, 1985).[7] The other conditions in Assumption 2 are relatively easy to understand and verify. Condition (ii) states a first-order approximation for the functional estimation error $\ell(\hat{h}_j, \cdot) - \ell(h_j, \cdot)$, which mainly pertains to the smoothness of $\ell(\cdot)$. Condition (iii) ensures that standard errors can be feasibly estimated.[8] Conditions (iv) and (v) prevent degeneracy, which are quite standard in this type of analysis.

[7] Also see White and Gallant (1988), Davidson (1994), and Pötscher and Prucha (2013) for comprehensive reviews.

[8] The variance-covariance estimator $\widehat{\Omega}$ is generally a growing dimensional HAC estimator. The consistency of kernel-based HAC estimators, including the Newey–West esitmator, has been established by Li and Liao (2020) for general mixingale-type dependent data.

We are now ready to state the asymptotic size and power properties of the functional inequality test proposed in Algorithm 1. Below, for two sequences of positive numbers $a_n$ and $b_n$, we write $a_n \succ b_n$ if $a_n \geq c_n b_n$ for some strictly positive sequence $c_n \to \infty$.

**Theorem 1.** *Under Assumptions 1 and 2, the following statements hold for any significance level* $\alpha \in (0, 0.5)$*: (a) the test determined by the critical region* $\{\widehat{T}_n > \hat{c}_n\}$ *is uniformly consistent in level:*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_{0,n}} P\left(\widehat{T}_n > \hat{c}_n\right) \leq \alpha;$$

*(b) for any* $P \in \mathcal{P}_n \backslash \mathcal{P}_{0,n}$ *satisfying*

$$\max_{j\in\mathcal{J}} \sup_{x\in\mathcal{X}} \ell(h_j, x) \succ \xi_k \log(k)^{1/2} n^{-1/2}, \tag{2.12}$$

*we have* $P(\widehat{T}_n > \hat{c}_n) \to 1$ *as* $n \to \infty$.

Part (a) of Theorem 1 shows that the proposed test controls asymptotic size uniformly over the $\mathcal{P}_n$ class of distributions. Part (b) establishes the consistency of the test against local alternatives that approach the null at rates slightly slower than $\xi_k \log(k)^{1/2} n^{-1/2}$.[9]

Since the testing problem (2.1) includes moment inequality testing as a special case, it is useful to clarify how our theoretical results are related to those of Romano, Shaikh, and Wolf (2014) and Bai, Santos, and Shaikh (2022). Romano, Shaikh, and Wolf (2014) first propose the two-step approach for testing unconditional moment inequalities with independent data and, importantly, establish the uniform size control for their proposed test. Bai, Santos, and Shaikh (2022) extend this approach in an important direction by accommodating "many" moment inequalities, and show that the resulting test is more powerful than that of Chernozhukov, Chetverikov, and Kato (2019). Our analysis, when specialized in the moment inequality context, shows that the same insight can be further extended to accommodate conditioning and nonparametric functional inference, as well as general forms of data dependence, using the new probabilistic tools developed in the present paper (see Section 3 for additional details). The scope of our analysis is of course not restricted to the moment inequality context, but pertains to generic functional inequalities, for which the two-step strategy remains to yield a practical and uniformly valid inference procedure.

[9] Recall from Assumption 1 that $\xi_k$ is a uniform bound for the vector norm of $\dot{\ell}(h_j^*, \cdot)$, which in turn quantifies the smoothness of the mapping $\beta_j \mapsto \ell(\psi(\cdot)^\top \beta_j, \cdot)$. In a typical setting the mapping $h \mapsto \ell(h, \cdot)$ is Lipschitz and $\xi_k$ admits a more explicit form as $\sup_{x\in\mathcal{X}} \|\psi(x)\|$, the sup-norm of the approximating series functions. For orthonormal polynomials, we thus have $\xi_k = O(k)$, and for splines and wavelets, $\xi_k = O(k^{1/2})$. Theorem 1(b) then implies that the proposed test is consistent against local alternatives going to zero slower than $k \log(k)^{1/2} n^{-1/2}$ (resp. $k^{1/2} \log(k)^{1/2} n^{-1/2}$) if orthonormal polynomials (resp. splines or wavelets) are used in the series estimation of the unknown functions $h_j(\cdot)$.

It is also worth noting that Romano, Shaikh, and Wolf (2014) and Bai, Santos, and Shaikh (2022) mainly advocate resampling methods, particularly various versions of the bootstrap, for feasible inference, whereas we rely directly on the asymptotic Gaussian approximation. Andrews and Soares (2010) note that, while the bootstrap method is preferred under the independent data setting, it is unclear whether it may outperform the asymptotic Gaussian approximation in the time series setting. Justifying (a version of) the bootstrap under the general setting of the present paper is an interesting but challenging task, which we leave for future research.

The theory developed in this section is admittedly abstract because of the high-level econometric structures employed in our presentation. This serves our goal of developing a general inference theory, but it might not be immediately clear to an applied researcher how to actually implement the test in practice. We turn next to a concrete demonstration.

# 3 Conditional Evaluation of Portfolio Performance

We apply the generic testing procedure to formally evaluate the relative state-dependent performance of portfolios measured by their conditional Sharpe ratio functions. Portfolio evaluation is of tremendous importance for both academic finance and investment practice. The setting is also econometrically interesting because the nonparametric inference naturally involves nontrivial complications such as shape restrictions, sequential estimation, and nonparametrically generated variables, which may arise in other types of applications as well. This section is organized as follows. We formally introduce the portfolio evaluation problem in Section 3.1 and construct the conditional Sharpe ratio inequality test in Section 3.2. The procedure involves shape-restricted series estimation, for which a data-driven tuning scheme is recommended in Section 3.3. Formal justifications and theoretical details are discussed in Section 3.4, which may be skipped by readers who are mainly interested in the application.

## 3.1 The portfolio evaluation problem

The econometric setting expands Example 1. Let $R_{t+1} = (R_{1,t+1}, \ldots, R_{N,t+1})^\top$ denote the vector of excess returns for $N$ risky assets during the period $[t, t+1]$. The conditional mean and variance-covariance matrix of $R_{t+1}$ given the time-$t$ information set $\mathcal{F}_t$ is denoted by $\mu_t$ and $\Sigma_t$, respectively. At time $t$, a risky portfolio $j$ is characterized by an $N$-dimensional vector of portfolio weights $w_{j,t}$ such that $\mathbf{1}_N^\top w_{j,t} = 1$, where $\mathbf{1}_N$ denotes the $N$-dimensional vector of ones. The associated portfolio return is $r_{j,t+1} \equiv w_{j,t}^\top R_{t+1}$, with its conditional mean and variance given by $\mu_{j,t} \equiv w_{j,t}^\top \mu_t$ and $\sigma_{j,t}^2 \equiv w_{j,t}^\top \Sigma_t w_{j,t}$, respectively. Under the classical mean-variance framework for portfolio choice, it is well-known that the optimal risky portfolio is the one that maximizes the Sharpe ratio, solving

$$\max_{w_t} \frac{w_t^\top \mu_t}{\sqrt{w_t^\top \Sigma_t w_t}}, \quad \text{s.t.} \quad \mathbf{1}_N^\top w_t = 1. \tag{3.13}$$

Since $\mu_t$ and $\Sigma_t$ are unknown, practical portfolios cannot realistically attain this theoretical ideal. A more relevant empirical question is to evaluate the relative performance of the portfolios by comparing their Sharpe ratios estimated using historical records.

We stress that the Sharpe ratio is intrinsically a *conditional* quantity, for it is derived from the conditional moments $\mu_t$ and $\Sigma_t$ of the excess returns. A portfolio's Sharpe ratio is thus state-dependent. One portfolio may have high Sharpe ratio in a volatile/crisis environment but performs poorly during a tranquil period, another portfolio may do the opposite, whereas an "all-weather" portfolio forever pursued by investors is hoped to deliver adequate performance under "all" circumstances.

The conditional perspective directly suggests that the definition of Sharpe ratio depends on the evaluator's conditioning information set $\mathcal{F}_t$. To construct a tractable conditional evaluation procedure we evidently need some structure on the information set (Giacomini and White, 2006). Following common practice, we suppose that the information set is spanned by a conditioning variable $X_t$, so that the conditional mean and variance of the portfolio return may be represented as conditional moment functions $\mu_j(x) = \mathbb{E}[r_{j,t+1}|X_t = x]$ and $\sigma_j^2(x) = Var(r_{j,t+1}|X_t = x)$, respectively. The choice of conditioning variable determines the working definition of the Sharpe ratio, and as such, pins down the evaluation criterion. To clarify, our econometric framework does not dictate which conditioning variable should be chosen. The evaluator is tasked to make this choice according to their own view on which variable best reveals the "economic state" that is relevant for discriminating the portfolios under evaluation. This is precisely the channel through which an evaluator's expertise can lead to a more informative empirical investigation within our econometric framework. That being said, we suggest using the VIX as a default choice because it is arguably the most popular real-time barometer for financial market conditions.

We aim to test the null hypothesis that a benchmark portfolio (indexed by $j = 0$) weakly dominates all competing portfolios indexed by $j \in \mathcal{J}$ across all conditioning states. If so, the benchmark may be deemed as an all-weather portfolio. Recall that this problem is a special case of (2.1) corresponding to $h_j(\cdot) = \mathrm{SR}_j(\cdot) - \mathrm{SR}_0(\cdot)$ and $\ell(h_j, x) = h_j(x)$. The choice of benchmark is often self-evident in practice provided that the evaluator has an articulated goal.[10] Nevertheless, in certain applications it might be desirable to treat all portfolios under study symmetrically without

[10]Consider two examples (i) if the researcher is interested in examining whether some actively managed portfolios can ever outperform a passive index fund (especially after fee), it is natural to use the passive portfolio as a

designating any portfolio as *the* benchmark. In that case, we recommend "inverting" the test by rotating the benchmark role across all competing portfolios and constructing a *confidence set for the most superior (CSMS)*. Formally, we denote the set of superior portfolios by

$$\mathcal{M}^* \equiv \left\{ j \in \mathcal{J} : \mathrm{SR}_j(\cdot) \geq \mathrm{SR}_{j'}(\cdot) \text{ for all } j' \in \mathcal{J} \right\}.$$[11]

Note that $j \in \mathcal{M}^*$ if and only if portfolio $j$ weakly dominates all the other competitors in $\mathcal{J}$, which can be tested by taking $j$ as the benchmark. The confidence set of the most superior portfolio is then formed as

$$\widehat{\mathcal{M}} \equiv \{ j \in \mathcal{J} : \text{the null hypothesis } H_0 : j \in \mathcal{M}^* \text{ is not rejected} \}. \tag{3.14}$$

By the duality between tests and confidence sets, it is easy to see that for each $j \in \mathcal{M}^*$, $\widehat{\mathcal{M}}$ contains $j$ with probability at least $1 - \alpha$ asymptotically.[12]

### 3.2 Conditional Sharpe ratio inequality test

In order to apply the general testing procedure proposed in Section 2 to the present portfolio evaluation setting, we need to properly design the $\hat{h}_j(\cdot)$ and $\widehat{\Omega}$ estimators and verify the high-level conditions in Theorem 1. Recall that the null hypothesis under study,

$$H_0 : \mathrm{SR}_0(x) \geq \mathrm{SR}_j(x) \text{ for all } j \in \mathcal{J} \text{ and } x \in \mathcal{X}, \tag{3.15}$$

is a special case of (2.1) with $h_j(x) = \mathrm{SR}_j(x) - \mathrm{SR}_0(x)$ and $\ell(h_j, x) = h_j(x)$. The main task is thus to construct and analyze nonparametric series estimators for the conditional Sharpe ratio functions $\mathrm{SR}_j(\cdot)$. In this subsection, we describe how to construct the requisite estimators and implement the conditional evaluation test. The accompanying theoretical discussion is postponed to Section 3.4.

To nonparametrically estimate the conditional Sharpe ratio function, the most straightforward approach is to first estimate the conditional mean function $\mu_j(\cdot)$ by running a nonparametric series regression for $r_{j,t+1}$ on $X_t$, with the resulting estimator given by

$$\hat{\mu}_j(\cdot) \equiv \psi(\cdot)^\top \widehat{Q}^{-1} \left( n^{-1} \sum_{t=1}^{n} \psi(X_t) r_{j,t+1} \right) \quad \text{where} \quad \widehat{Q} \equiv n^{-1} \sum_{t=1}^{n} \psi(X_t) \psi(X_t)^\top. \tag{3.16}$$

benchmark; (ii) if an investor tries to decide whether their prevailing trading strategy should be replaced by some new ones, the strategy currently in use should clearly serve as the benchmark.

[11] The set $\mathcal{M}^*$ may be empty because the functional inequality only induces a partial, instead of complete, order.

[12] We note that the confidence set $\widehat{\mathcal{M}}$ is designed to cover each element in $\mathcal{M}^*$ instead of the set itself. This distinction is relevant only when $\mathcal{M}^*$ contains more than one portfolio, which necessarily implies that all portfolios in $\mathcal{M}^*$ have identical Sharpe ratios across all conditioning states. To the extent that two distinct portfolios are unlikely to share exactly the same conditional Sharpe ratio *function*, the aforementioned distinction appears empirically unimportant.

Then, the conditional variance function $\sigma_j^2(\cdot)$ may be estimated via a nonparametric series regression for the squared residual $\left(r_{j,t+1} - \hat{\mu}_j(X_t)\right)^2$ on $X_t$.[13] Plugging the estimators for $\mu_j(\cdot)$ and $\sigma_j^2(\cdot)$ into the definition $\mathrm{SR}_j(x) = \mu_j(x)/\sigma_j(x)$ yields an asymptotically valid estimator for the conditional Sharpe ratio function $\mathrm{SR}_j(\cdot)$.

A drawback of this seemingly natural approach, however, is that the said estimator of the conditional variance function is not guaranteed to be positive in finite samples, which is evidently quite undesirable for our analysis on the conditional Sharpe ratio (because the conditional variance appears in the denominator). We thus propose a sign-restricted estimator for $\sigma_j^2(\cdot)$ as a regularized alternative. This type of shape restriction is relatively straightforward to obtain in the series estimation framework (Chetverikov and Wilhelm, 2017; Chetverikov, Santos, and Shaikh, 2018; Chiang, Kato, Sasaki, and Ura, 2021).[14] Specifically, we implement a constrained series regression for the squared residual $\left(r_{j,t+1} - \hat{\mu}_j(X_t)\right)^2$ by solving

$$\hat{\gamma}_j \equiv \operatorname*{argmin}_{\gamma \in \mathbb{R}^k} n^{-1} \sum_{t=1}^{n} \left( \left(r_{j,t+1} - \hat{\mu}_j(X_t)\right)^2 - \psi(X_t)^\top \gamma \right)^2, \quad \text{s.t.} \min_{1 \le t \le n} \psi(X_t)^\top \gamma \ge \underline{\sigma}_{j,n}^2, \tag{3.17}$$

where the tuning parameter $\underline{\sigma}_{j,n}^2 > 0$ imposes a lower bound for the estimated $\sigma_j^2(\cdot)$ over the observed values of $X_t$.[15] The estimator for $\sigma_j^2(\cdot)$ is then given by

$$\hat{\sigma}_j^2(\cdot) \equiv \psi(\cdot)^\top \hat{\gamma}_j.$$

The sign restriction will be handled as a finite-sample regularization in our formal econometric analysis detailed in Section 3.4 below. The tuning sequence $\underline{\sigma}_{j,n}^2$ is assumed to shrink to zero, so that the sign restrictions become unbinding in large samples. Although setting $\underline{\sigma}_{j,n}^2$ as any positive $o(1)$ sequence would technically fulfill this requirement, the accompanying arbitrariness is clearly undesirable. To rectify this issue, we recommend a data-driven choice of this tuning parameter in Section 3.3, which is based on a preliminary consistent estimator of the lower bound of the

[13] Fan and Yao (1998) propose a kernel-type nonparametric estimator for the conditional variance function using the squared residuals and show that the resulting estimator is adaptive with respect to the unknown conditional mean function. The authors also show theoretically that this residual-based approach outperforms the "direct" approach under which the conditional variance estimator is formed as the difference between estimates of the conditional second moment and the squared conditional mean. Inspired by Fan and Yao's insight, we adopt the residual-based estimation approach, though we focus on the series method; this is more convenient for our uniform functional inference, which is not considered in Fan and Yao (1998). The series-based estimator also enjoys the said adaptiveness property, as shown in Lemma S11 in the Supplemental Appendix.

[14] The other type of shape restrictions such as monotonicity and convexity can also be readily incorporated, which may be relevant elsewhere.

[15] The minimization in (3.17) is a quadratic program with linear constraints, which can be easily and efficiently solved using standard computational packages.

conditional variance function. This is our recommended default choice. It is worth noting that, in practice, the user may also exploit prior knowledge and/or additional information to decide what may be a reasonable choice for this lower bound in their specific application.[16]

To clarify, the sign-restricted estimator $\hat{\sigma}_j^2(\cdot)$ is designed to satisfy $\hat{\sigma}_j^2(x) \geq \underline{\sigma}_{j,n}^2$ only for $x$ in the realized support $\{X_t : 1 \leq t \leq n\}$ rather than for all $x \in \mathcal{X}$. This "seemingly incomplete" restriction is in fact sufficient for our purpose of nonparametrically estimating the conditional Sharpe ratio function. The idea is to consider the normalized return

$$r_{j,t+1}^* \equiv \frac{r_{j,t+1}}{\sigma_j(X_t)},$$

and observe that the conditional mean function of $r_{j,t+1}^*$ is exactly the conditional Sharpe ratio of $r_{j,t+1}$. While $r_{j,t+1}^*$ is not directly observable, it can be proxied by the generated variable

$$\hat{r}_{j,t+1}^* \equiv \frac{r_{j,t+1}}{\hat{\sigma}_j(X_t)},$$

which only relies on $\hat{\sigma}_j(\cdot)$ evaluated at the $X_t$'s. The Sharpe ratio function $\mathrm{SR}_j(\cdot)$ may then be estimated by nonparametrically regressing $\hat{r}_{j,t+1}^*$ on $X_t$, with the resulting estimator given by

$$\widehat{\mathrm{SR}}_j(\cdot) \equiv \psi(\cdot)^\top \hat{b}_j, \quad \text{where} \quad \hat{b}_j \equiv \widehat{Q}^{-1}\left(n^{-1}\sum_{t=1}^n \psi(X_t)\hat{r}_{j,t+1}^*\right). \tag{3.18}$$

We then estimate the $h_j(\cdot)$ function by $\hat{h}_j(\cdot) = \widehat{\mathrm{SR}}_j(\cdot) - \widehat{\mathrm{SR}}_0(\cdot)$, which may be represented in the series form as

$$\hat{h}_j(\cdot) = \psi(\cdot)^\top \hat{\beta}_j, \quad \text{where} \quad \hat{\beta}_j = \hat{b}_j - \hat{b}_0 = \widehat{Q}^{-1}\left(n^{-1}\sum_{t=1}^n \psi(X_t)\left(\hat{r}_{j,t+1}^* - \hat{r}_{0,t+1}^*\right)\right). \tag{3.19}$$

The remaining task is to construct the $\widehat{\Omega}$ estimator for the joint variance-covariance matrix of the $\hat{\beta}_j$'s, for which we need to account for the effect of the generated dependent variables $\hat{r}_{j,t+1}^* - \hat{r}_{0,t+1}^*$ on $\hat{h}_j(\cdot)$. We define

$$\begin{aligned}\hat{v}_{j,t+1} \;\equiv\;& \psi(X_t)\left(\hat{r}_{j,t+1}^* - \widehat{\mathrm{SR}}_j(X_t)\right) \\ & - \left(n^{-1}\sum_{t=1}^n \frac{r_{j,t+1}}{2\hat{\sigma}_j^3(X_t)}\psi(X_t)\psi(X_t)^\top \widehat{Q}^{-1}\right)\psi(X_t)\left(\left(r_{j,t+1} - \hat{\mu}_j(X_t)\right)^2 - \hat{\sigma}_j^2(X_t)\right),\end{aligned} \tag{3.20}$$

[16] For example, if high-frequency data on portfolio returns are available, a reasonable approach is to first estimate the realized variance (RV) of the portfolio and then set $\underline{\sigma}_{j,n}^2$ to be a "shrunk" version of the historical low of the RV series.

where the first component corresponds to the influence function in (3.18) and the second component captures the influence of the proxy error in the generated dependent variable. It is also convenient to define two vectors $\hat{b}$ and $\hat{v}_{t+1}$ by stacking $(\hat{b}_j)_{j\in\{0\}\cup\mathcal{J}}$ and $(\hat{v}_{j,t+1})_{j\in\{0\}\cup\mathcal{J}}$, respectively. We can show that $\hat{b}$ admits an asymptotic linear representation and its influence function can be approximated by $(I_{J+1}\otimes\widehat{Q}^{-1})\hat{v}_{t+1}$. It then follows that the variance-covariance matrix of the $\hat{b}$ vector may be estimated by

$$\widehat{\Omega}_b \equiv (I_{J+1}\otimes\widehat{Q}^{-1})\widehat{\Omega}_v(I_{J+1}\otimes\widehat{Q}^{-1}),$$

where $\widehat{\Omega}_v$ is taken as a (long-run) variance-covariance estimator for the $\hat{v}_{t+1}$ series. We further use $\tilde{L}_j$ to denote the selection matrix such that $\hat{b}_j = \tilde{L}_j\hat{b}$ and set

$$\tilde{L} \equiv \begin{pmatrix} \tilde{L}_1 - \tilde{L}_0 \\ \vdots \\ \tilde{L}_J - \tilde{L}_0 \end{pmatrix}.$$

With this notation, we can write $\hat{\beta} = \tilde{L}\hat{b}$, which suggests constructing $\widehat{\Omega}$ as

$$\widehat{\Omega} = \tilde{L}\widehat{\Omega}_b\tilde{L}^\top. \tag{3.21}$$

We are now ready to implement the generic testing procedure proposed in Section 2.2. Since $\ell(\psi(\cdot)^\top\beta, x) = \psi(x)^\top\beta$ in the present setting, its derivative takes a simple form $\dot{\ell}(\hat{h}_j, x) = \psi(x)$, and so, the $\hat{s}_j(\cdot)$ estimator for the standard error function defined in (2.5) may be more explicitly written as

$$\hat{s}_j(x) \equiv \sqrt{\psi(x)^\top L_j\widehat{\Omega}L_j^\top\psi(x)}.$$

The $\widehat{T}_n$ test statistic can then be computed according to (2.6), which can also be equivalently written as

$$\widehat{T}_n \equiv \max\left\{\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\frac{n^{1/2}\big(\widehat{\mathrm{SR}}_j(x) - \widehat{\mathrm{SR}}_0(x)\big)}{\hat{s}_j(x)}, 0\right\}.$$

The test statistic quantifies the maximal extent to which any competitor $j\in\mathcal{J}$ outperforms the benchmark. We reject the null hypothesis that the benchmark portfolio weakly dominates all the others if $\widehat{T}_n$ exceeds the critical value described in Algorithm 1.

### 3.3 A data-driven tuning sequence for sign-restricted estimation

We recommend a data-driven choice of $\underline{\sigma}_{j,n}^2$, the tuning parameter for the sign-restricted estimation of the conditional variance function $\sigma_j^2(\cdot)$ described in (3.17). In theory, we require it to be a slack lower bound for the $\sigma_j^2(\cdot)$ function, so that the constraint $\sigma_j^2(\cdot) \geq \underline{\sigma}_{j,n}^2$ eventually becomes unbinding in large samples; then, the sign-restricted estimator $\hat{\sigma}_j^2(\cdot)$ shares the same asymptotic properties with its unrestricted counterpart. To achieve this goal, we consider a preliminary estimator for the $\sigma_j^2(\cdot)$ function, defined as

$$\tilde{\sigma}_j^2(\cdot) \equiv \frac{\psi(\cdot)^\top \widehat{Q}^{-1}\widehat{\Lambda}_j\widehat{Q}^{-1}\psi(\cdot)}{\psi(\cdot)^\top \widehat{Q}^{-1}\psi(\cdot)} \qquad \text{where} \qquad \widehat{\Lambda}_j \equiv n^{-1}\sum_{t=1}^{n}\left(r_{j,t+1} - \hat{\mu}_j(X_t)\right)^2 \psi(X_t)\,\psi(X_t)^\top. \tag{3.22}$$

In the Supplemental Appendix, we show that this estimator is uniformly consistent; see Proposition S2. It is worth noting that $\tilde{\sigma}_j^2(\cdot)$ is positively valued in finite samples by construction but unfortunately has a relatively slow rate of convergence. This explains why we do not employ it directly in the estimation of conditional Sharpe ratio function, but only use it as a pilot estimator for tuning $\underline{\sigma}_{j,n}^2$.

Our recommended choice of $\underline{\sigma}_{j,n}^2$ is then given by

$$\underline{\sigma}_{j,n}^2 = \frac{\min_{1\leq t\leq n}\tilde{\sigma}_j^2(X_t)}{\max\{1, \log n\}}. \tag{3.23}$$

Here, $\min_{1\leq t\leq n}\tilde{\sigma}_j^2(X_t)$ consistently estimates the lower bound of $\sigma_j^2(\cdot)$, which is then shrunk towards zero by the slowly growing $\log(n)$ factor. Under the maintained assumption that the conditional variance function is bounded away from zero, this construction ensures that $\underline{\sigma}_{j,n}^2 < \min_{1\leq t\leq n}\sigma_j^2(X_t)$ with probability approaching 1 as $n \to \infty$, and so, fulfills the said theoretical requirement on the $\underline{\sigma}_{j,n}^2$ tuning sequence.

## 3.4 Theoretical justifications

We now present the formal justification for the conditional Sharpe ratio inequality test proposed in Section 3.2 above. By specializing the general theoretical framework developed in Section 2.2 within the current specific context, we can incarnate the high-level conditions in a more concrete form, which in turn directly suggests how they may be verified under general primitive conditions. For readability, we focus on the main theoretical points here, relegating various technical details to the Supplemental Appendix. Among them, arguably the most important is our new results on the uniform Gaussian coupling for mixingale-type dependent data, which is the main driver for the uniform size property of the proposed test.

More notation is needed. Denote $\mathcal{J}_0 = \{0\} \cup \mathcal{J}$ for simplicity. Let $\epsilon_{t+1} \equiv (\epsilon_{j,t+1})_{j\in\mathcal{J}_0}$, where

$$\epsilon_{j,t+1} \equiv (\varepsilon_{j,t+1}, \eta_{j,t+1})^\top, \quad \varepsilon_{j,t+1} \equiv r_{j,t+1}^* - \mathrm{SR}_j(X_t), \quad \eta_{j,t+1} \equiv \sigma_j^2(X_t)(\varepsilon_{j,t+1}^2 - 1),$$

and further set $\Psi_\epsilon$ and $\Psi_u$ as the variance-covariance matrices of $n^{-1/2}\sum_{t=1}^{n}\epsilon_{t+1}\otimes\psi(X_t)$ and $n^{-1/2}\sum_{t=1}^{n}\sigma\left(X_t\right)\varepsilon_{j,t+1}\otimes\psi(X_t)$, respectively. Denote

$$
\begin{aligned}
\tilde{\Gamma}_j &\equiv n^{-1}\sum_{t=1}^{n}\sigma\left(X_t\right)\varepsilon_{j,t+1}\psi(X_t)\psi(X_t)^\top, \\
\tilde{\Lambda}_j &\equiv n^{-1}\sum_{t=1}^{n}\sigma^2\left(X_t\right)\varepsilon_{j,t+1}^2\psi(X_t)\psi(X_t)^\top, \\
\tilde{\Upsilon}_j &\equiv n^{-1}\sum_{t=1}^{n}\frac{r_{j,t+1}}{2\sigma_j^3\left(X_t\right)}\psi(X_t)\psi(X_t)^\top Q^{-1},
\end{aligned}
$$

where $Q\equiv n^{-1}\sum_{t=1}^{n}\mathbb{E}[\psi(X_t)\psi(X_t)^\top]$. Finally, we complement the definition in (3.20) with its population version by setting $v_{t+1}\equiv\left(v_{j,t+1}\right)_{j\in\mathcal{J}_0}$, where

$$
\begin{aligned}
v_{j,t+1} \equiv\ & \psi(X_t)(r_{j,t+1}^*-\mathrm{SR}_j\left(X_t\right)) \\
& -n^{-1}\sum_{t=1}^{n}\mathbb{E}\left[\frac{r_{j,t+1}}{2\sigma_j^3\left(X_t\right)}\psi(X_t)\psi(X_t)^\top\right]Q^{-1}\psi(X_t)\left(\left(r_{j,t+1}-\mu_j\left(X_t\right)\right)^2-\sigma_j^2\left(X_t\right)\right).
\end{aligned}
$$

The variance-covariance matrix of $n^{-1/2}\sum_{t=1}^{n}v_{t+1}$ is denoted by $\Omega_v$.

In what follows, we introduce some conditions that are tailored for the conditional Sharpe ratio inequality test, explain how to verify them under primitive conditions, and then show that they imply the high-level conditions used in Theorem 1.

**Assumption 3.** *The following conditions hold uniformly over $P\in\mathcal{P}_n$: (i) for each $j\in\mathcal{J}_0$, $\mu_j(\cdot)$ and $\sigma_j^2(\cdot)$ are continuously differentiable with bounded derivatives and $\sigma_j^2(\cdot)$ is bounded above and away from zero; (ii) the eigenvalues of $Q$, $\Psi_\epsilon$, and $\Psi_u$ are bounded above and below by some fixed positive constants; (iii) for some constant $K>0$ and any $\varphi(\cdot)\in\cup_{j\in\mathcal{J}_0}\{\mu_j(\cdot),\sigma_j^2(\cdot),\mathrm{SR}_j(\cdot)\}$, there exists a sequence $b_\varphi^*$ of $k$-dimensional nonrandom vectors such that*

$$
\sup_{x\in\mathcal{X}} n^{1/2}\left|\varphi(x)-\psi(x)^\top b_\varphi^*\right|\leq K\log(n)^{-2};
$$

*(iv) there exists a sequence of $m$-dimensional standard Gaussian random vectors $\mathcal{N}_m$ such that*

$$
n^{-1/2}\sum_{t=1}^{n}v_{t+1}=\Omega_v^{1/2}\mathcal{N}_m+o_{\mathcal{P}}(\log(n)^{-1});
$$

*(v) $\|\widehat{Q}-Q\|_S+\|\widehat{\Omega}_v-\Omega_v\|_S=o_{\mathcal{P}}(k^{-1/2}\log(n)^{-1})$; (vi) $\|\tilde{\Lambda}_j-\mathbb{E}[\tilde{\Lambda}_j]\|_S+\|\tilde{\Gamma}_j\|_S=o_{\mathcal{P}}(k^{-1/2}\log(n)^{-1})$ for $j\in\mathcal{J}_0$; (vii) $\inf_{x\in\mathcal{X}}\|\psi(x)\|>K^{-1}$, $k^2n^{-1/2}+k^{-1}=o(\log(n)^{-1})$, $\log(\xi_{1,k})=O(\log(k))$, and $\xi_{0,k}kn^{-1/2}+\xi_{1,k}n^{-1}=o\left(\log(n)^{-1}\right)$ where $\xi_{j,k}\equiv\sup_{x\in\mathcal{X}}\|\partial^j\psi(x)/\partial x^j\|$ for $j=0,1$.*

Conditions (i)–(iii) and (vii) are fairly standard for series estimation (Andrews, 1991; Newey, 1997). Conditions (iv)–(vi) are high-level in nature, in that they directly restrict the asymptotic behavior of various (scaled) sample moments uniformly over $\mathcal{P}_n$. Specifically, condition (iv) requires that the sequence $n^{-1/2}\sum_{t=1}^{n} v_{t+1}$ of growing dimensional random vectors admits a Gaussian coupling. This condition can be verified for general mixingale processes by directly applying the new uniform Gaussian coupling theory that we develop in Section SA of the Supplemental Appendix. Conditions (v) and (vi) are less demanding, in that they "only" pertain to the convergence rates of $\widehat{Q}$, $\widehat{\Omega}_v$, $\tilde{\Upsilon}_j$, $\tilde{\Lambda}_j$, and $\tilde{\Gamma}_j$, which can be verified under primitive conditions following the same law-of-large-numbers type of argument in Chen and Christensen (2015) and Li and Liao (2020).

**Assumption 4.** *For some $K > 0$ and each $j \in \mathcal{J}_0$, $\underline{\sigma}_{j,n}^2 \leq (1+K)^{-1}\min_{1\leq t\leq n}\sigma_j^2(X_t)$ with probability approaching 1 uniformly for $P \in \mathcal{P}_n$.*

Assumption 4 formalizes the requirement that the tuning parameter $\underline{\sigma}_{j,n}^2$ needs to form a lower bound for the conditional variance function. Under the maintained Assumption 3, this condition is automatically satisfied for the recommended data-driven choice described in Section 3.3; see Section SD in the Supplemental Appendix for theoretical details.

**Proposition 1.** *Under the setting described in Section 3.2, Assumptions 3 and 4 imply Assumptions 1 and 2.*

In summary, Assumptions 3 and 4 can be verified by combining well-known results from the classic series estimation literature and some new technical tools developed in the Supplemental Appendix for uniform strong approximation. By Proposition 1, these assumptions imply the high-level conditions used in Theorem 1, and so, ensure the theoretical validity of the proposed conditional Sharpe ratio inequality test. We turn next to apply this test on both simulated and real data.

# 4 Simulations

In this section, we investigate the finite-sample performance of the proposed functional inequality test. Section 4.1 presents the simulation design. Results are reported in Section 4.2.

## 4.1 The Monte Carlo setting

Our Monte Carlo design is based on the conditional portfolio evaluation setting discussed in Section 3. We simulate the return of the $j$th portfolio as follows:

$$r_{j,t+1} = \mu_j(X_t) + \sigma_j(X_t)\varepsilon_{j,t+1}, \qquad X_t = \gamma X_{t-1} + (1-\gamma^2)^{1/2}\nu_t, \tag{4.1}$$

where $\mu_j(\cdot)$ and $\sigma_j(\cdot)$ denote the conditional mean and the conditional standard deviation functions of $r_{j,t+1}$ given the conditioning variable $X_t$, respectively. Moreover, $\nu_t$ and $\varepsilon_{j,t}$ are *i.i.d.* standard normal random variables that are mutually independent. The $\gamma$ parameter controls the serial dependence of the dynamic system, which we set as $\gamma = 0.4$ or $0.8$. Following the notation in Section 2.3, we index the benchmark portfolio and the competing portfolios using $j = 0$ and $j \in \mathcal{J}$, respectively. Below, $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and the probability density function of the standard normal distribution, respectively.

To ensure that the functional forms of $\mu_j(\cdot)$ and $\sigma_j(\cdot)$ employed in the Monte Carlo design are empirically realistic, we calibrate their specifications to some empirical estimates obtained from one of the datasets, Size/BM-100, used in our empirical analysis in Section 5. Specifically, the conditional standard deviation function $\sigma_j(x)$ is given by

$$\sigma_j(x) = -3\phi(-3\Phi^2(x) - 0.4\Phi(x) + 1.8) + 1.6089, \tag{4.2}$$

which mimics the empirical estimate of a global minimum variance portfolio. Recalling $\mu_j(\cdot) = \sigma_j(\cdot)\mathrm{SR}_j(\cdot)$, it remains to specify the conditional Sharpe ratio functions. Since our test mainly concerns the difference between Sharpe ratio functions rather than the functions themselves, we normalize the benchmark Sharpe ratio function $\mathrm{SR}_0(\cdot)$ to a constant 0.0335. We then specify the functional Sharpe ratio differentials as

$$\mathrm{SR}_j(x) - \mathrm{SR}_0(x) = a_j + 0.1(\phi(2.6\Phi(x)^2 + 2.5\Phi(x) - 0.2) - \phi(0)). \tag{4.3}$$

This specific functional form is again calibrated to an empirical estimate.

Under this specification, we have $a_j = \sup_x(\mathrm{SR}_j(x) - \mathrm{SR}_0(x))$ by design, and hence, the null hypothesis holds if and only if $a_j \leq 0$ for all $j \in \mathcal{J}$. We may also introduce slackness under the null hypothesis by setting $a_j < 0$. With this in mind, we examine a range of configurations for this data generating process (DGP) by varying the number of functional inequalities (i.e., $J$) and the values of the $a_j$ constants specifically as follows: for $a \in [0, 0.25]$, consider

$$\begin{aligned} &\textbf{DGP 1}. \quad J = 4, a_1 = 0, a_2 = a_3 = -1, \text{and } a_4 = a; \\ &\textbf{DGP 2}. \quad J = 8, a_1 = a_2 = 0, a_3 = \cdots = a_6 = -1, \text{and } a_7 = a_8 = a. \end{aligned} \tag{4.4}$$

Under both configurations, one quarter of the functional inequalities is partially binding (i.e., $a_j = 0$), one half is globally slack (i.e., $a_j < 0$), and the remaining quarter with $a_j = a$ may violate the null hypothesis to the degree quantified by the constant $a$. The null hypothesis obtains when $a = 0$. Increasing the value of $a$ allows us to trace out power curves of the test.

We consider three sample sizes: $n = 1000$, 2500, or 5000. As a point of reference, the sample size in our empirical application is 5788 (corresponding to roughly 20 years of daily observations).

Simulation results for the $n = 5000$ case are thus directly informative for our subsequent empirical study, while the other two cases shed additional light on the test's behavior for smaller samples.

The series estimation involves a few additional implementation details. To construct the approximating basis functions, we first rescale $X_t$ onto the $[-1, 1]$ interval via the transformation $\widetilde{X}_t = 2\Phi((X_t - \mu_{X,n})/\sigma_{X,n}) - 1$, where $\mu_{X,n}$ and $\sigma_{X,n}$ denote the sample mean and standard deviation of $X_t$, respectively, and then set the basis function as $\varphi_l(\widetilde{X}_t)$, $l = 0, \ldots, k-1$, where $\varphi_l(\cdot)$ denotes the $l$th-order Legendre polynomial (orthonormalized with respect to the uniform distribution on $[-1, 1]$). Following Li, Liao, and Quaedvlieg (2022), the number of series terms is determined as $k = \max\{4, \lfloor 1.2n^{1/5} \rfloor\}$ so that the series regression fits at least a third-order polynomial and grows slowly with the sample size. This scheme results in $k = 4$, 5, and 6 for $n = 1000$, 2500, and 5000, respectively. Since $v_{t+1}$ forms an MDS in the present setting, we estimate the variance-covariance matrix $\Omega_v$ using its sample analogue. The MDS structure arises naturally in the current context of analyzing asset returns. It is important to note that such time series may be serially highly dependent through higher-order conditional moments, which prevents the application of asymptotic theory developed for random samples. For completeness, we also consider in the Supplemental Appendix a more complicated simulation design under which $v_{t+1}$ is no longer an MDS and $\Omega_v$ is estimated using the Newey–West estimator. The latter setting yields very similar findings, and so, are omitted from our main discussion here.

We implement the test at significance level $\alpha = 5\%$. As discussed in Section 2.2, we take $\rho \in \{0.001, 0.005\}$ following the recommendations from Berger and Boos (1994), Silvapulle (1996), and Romano, Shaikh, and Wolf (2014). For comparison, we also consider cases with $\rho = 0.025$ or 0. The $\rho = 0$ setting corresponds to the least favorable approach. Finite-sample rejection frequencies of the tests are computed based on 10000 Monte Carlo replications.

## 4.2 Simulation results

We first examine the size property of the proposed test for conditional Sharpe ratio inequalities. The null hypothesis is imposed by setting $a = 0$; recall (4.4). Table 1 reports the Monte Carlo rejection rates of the test for various specifications of $\gamma$, $\rho$, $J$, and $n$. Consistent with the asymptotic theory, we find that the test controls size well under all scenarios. We also note that, as is typical for this kind of inference, the rejection rates are notably below the 5% nominal level. The observed "conservativeness" mainly stems from the fact that the functional inequalities in our Monte Carlo experiment are slack, either partially or globally. The undersizing is particularly evident for the least favorable approach. This is not a small-sample phenomenon, as it is present even when $n$ becomes larger.

Table 1: Rejection Rates Under the Null Hypothesis

| | $\gamma = 0.4$ | | | | $\gamma = 0.8$ | | | |
|---|---|---|---|---|---|---|---|---|
| | L.F. | $\rho = 0.001$ | $\rho = 0.005$ | $\rho = 0.025$ | L.F. | $\rho = 0.001$ | $\rho = 0.005$ | $\rho = 0.025$ |
| | | | | $n = 1000$ | | | | |
| $J = 4$ | 0.019 | 0.028 | 0.027 | 0.017 | 0.019 | 0.028 | 0.027 | 0.016 |
| $J = 8$ | 0.018 | 0.028 | 0.026 | 0.015 | 0.020 | 0.034 | 0.031 | 0.018 |
| | | | | $n = 2500$ | | | | |
| $J = 4$ | 0.015 | 0.029 | 0.026 | 0.014 | 0.016 | 0.029 | 0.028 | 0.015 |
| $J = 8$ | 0.015 | 0.029 | 0.027 | 0.014 | 0.017 | 0.034 | 0.031 | 0.017 |
| | | | | $n = 5000$ | | | | |
| $J = 4$ | 0.014 | 0.027 | 0.025 | 0.014 | 0.013 | 0.027 | 0.024 | 0.013 |
| $J = 8$ | 0.016 | 0.030 | 0.027 | 0.016 | 0.014 | 0.028 | 0.025 | 0.013 |

*Note:* The table reports the rejection rates of the conditional Sharpe ratio inequality test at the 5% significance level under the null hypothesis. The test is implemented using $\rho \in \{0.001, 0.005, 0.025\}$. L.F. denotes the test constructed using the least favorable approach. The rejection rates are computed based on 10000 Monte Carlo replications.

Meanwhile, we observe that the proposed two-step test is able to mitigate the undersizing issue as desired. The rejection rates of the test with $\rho = 0.001$ are notably closer to the nominal level than those of the least favorable test. Moreover, the performance of the test is insensitive with respect to the choice of $\rho$, evidenced by the very similar results from the $\rho = 0.005$ implementation. The robustness with respect to the choice of $\rho$ is desirable in practice and confirms that the rule-of-thumb choices recommended by Berger and Boos (1994), Silvapulle (1996), Romano, Shaikh, and Wolf (2014), among others, are adequate in the present setting as well. That being said, we note that $\rho$ should not be taken "too large." Indeed, when $\rho = 0.025$, the two-step test becomes even more conservative than the least favorable test.

We turn next to the power analysis. In Figure 1, we plot the power curves as functions of the $a$ parameter. The results for the $\gamma = 0.4$ and $\gamma = 0.8$ cases are very similar, and so, we only discuss the former here for brevity and relegate the latter to the Supplemental Appendix. As expected, the rejection rates of the tests increase as the alternative hypothesis deviates farther away from the null (as $a$ becomes larger), and eventually reach probability one. Other things being equal, the rejection rates are also higher for larger samples. Importantly, the proposed two-step test with the recommended choice of $\rho$ are notably more powerful than the least favorable test, uniformly across all specifications. The power curves for the $\rho = 0.001$ and $\rho = 0.005$ tests are nearly identical, mirroring the previously noted robustness with respect to the choice of $\rho$. Taking $\rho$ too large would hurt the test's power.

In summary, the simulation results show that the proposed test has excellent size control. Relative to the least favorable approach, it is notably less conservative under the null and more powerful under the alternative, uniformly across all the configurations considered in our Monte Carlo analysis. These findings are consistent with the theory. The rule-of-thumb choice of $\rho$ from the $[0.001, 0.005]$ range is adequate for our inferential setting and generates robust numerical findings. Like the cited prior work, we also recommend such choice for practical applications.

# 5 An Empirical Application on Conditional Portfolio Evaluation

As an empirical illustration, we apply the proposed econometric method to evaluate the out-of-sample performance of four trading strategies using twelve datasets from the U.S. equity market. Section 5.1 describes the empirical setting. Section 5.2 reports the results.

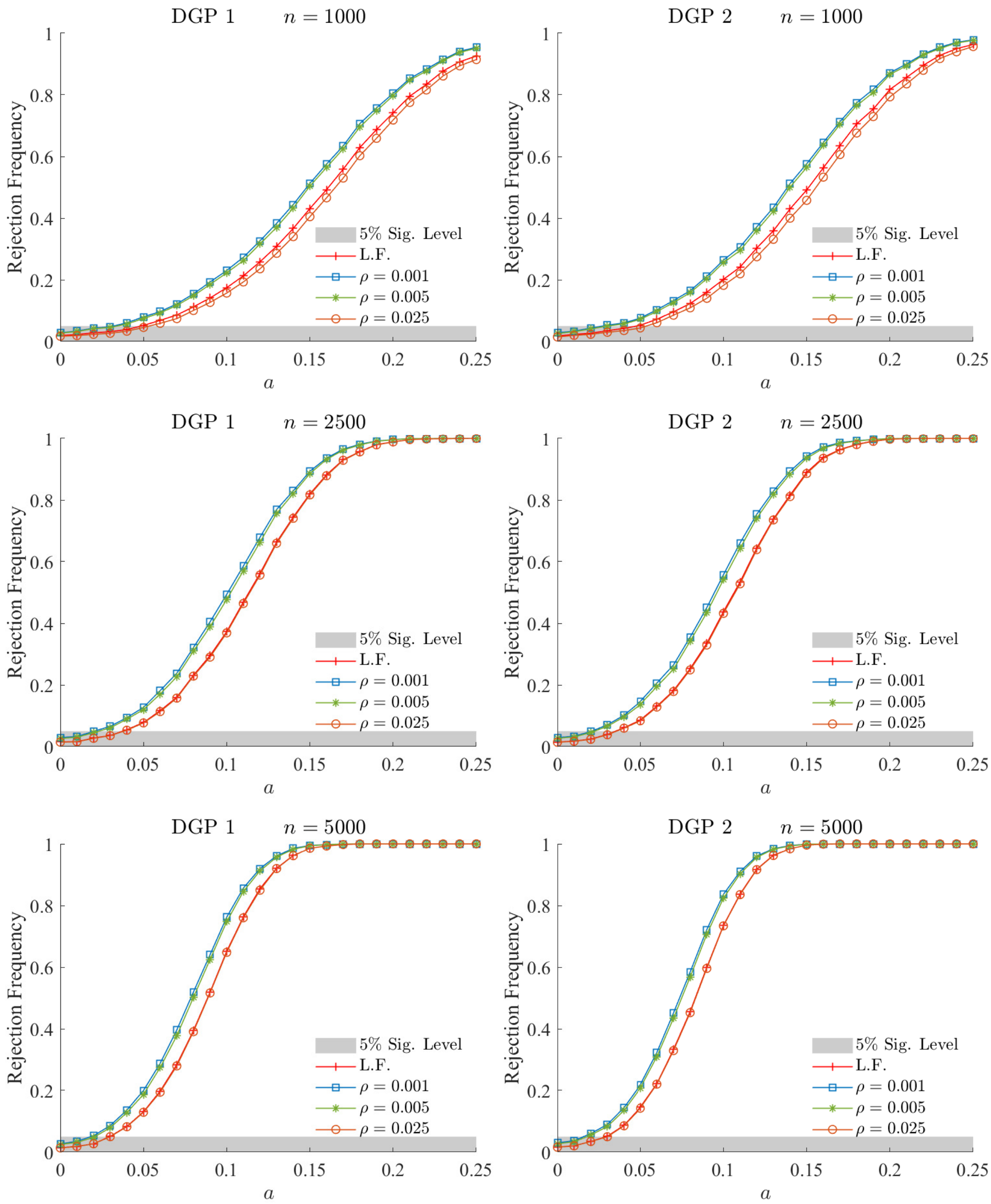
## 5.1 The setting

The literature on portfolio construction is vast with countless proposals (Guerard Jr, 2009). An enormous amount of research has been devoted to comparing their performance using conventional unconditional evaluation methods (Ferson, 2010). Equipped with the new conditional Sharpe ratio inequality test, we aim to shed light on this important evaluation problem from a conditional perspective.

Our empirical design is motivated by the following important studies. In an influential paper, DeMiguel, Garlappi, and Uppal (2009) provide striking evidence that many well-known quantitative trading strategies cannot outperform the simple equally weighted "$1/N$" portfolio. Jagannathan and Ma (2003) find that imposing short-sale constraints can greatly improve the performance of Markowitz-type portfolios and show that this has the effect of applying shrinkage on the estimated covariance matrix. In a recent paper, Ledoit and Wolf (2017) derive the optimal nonlinear shrinkage for estimating the covariance matrix in the portfolio construction context and document its superior performance.

In light of the said evidence, we shall focus on a few leading prototype strategies without re-investigating portfolio methods that are similar or have been documented to be inferior. Specifically, we compare the relative state-dependent performance of four trading strategies, including

Figure 1: Simulation Results: Power Curves




*Note:* The figure plots the power curves of the conditional Sharpe ratio inequality test at the 5% significance level (highlighted by the shaded area). In each panel, we plot the power curves for the tests constructed with the least favorable (L.F.) bound, and data-driven bounds corresponding to $\rho = 0.001, 0.005$, and $0.025$, respectively. The level of autocorrelation in $X_t$ is fixed at $\gamma = 0.4$. The rejection rates are computed based on 10000 Monte Carlo replications.

the $1/N$ portfolio, a global minimum variance (GMV) portfolio implemented using Ledoit and Wolf's (2017) nonlinear shrinkage method, a GMV portfolio with no-short-sale constraint, and a GMV portfolio with both nonlinear shrinkage and no-short-sale constraint. The four portfolios are denoted as $1/N$, GMV-NLS, GMV-NSS, and NLS-NSS, respectively. Restricting attention to the GMV class of quantitative strategies sharpens our focus on understanding the benefit of incorporating short-sale constraint and/or nonlinear shrinkage (Jagannathan and Ma, 2003; Ledoit and Wolf, 2017) and the inclusion of the $1/N$ portfolio allows us to revisit DeMiguel, Garlappi, and Uppal's (2009) message from a conditional perspective. The detailed descriptions of these well-known strategies are relegated to the Supplemental Appendix to save space.

Since the relative performance of trading strategies may depend on the composition of the investable universe, we conduct tests using twelve distinct datasets listed in Table 2. These datasets contain the daily returns of various factor portfolios provided in Kenneth French's data library.[17] Following DeMiguel, Garlappi, and Uppal (2009), among others, we treat the factor portfolios as investable assets in the portfolio construction problem. Our empirical analysis thus mainly concerns the allocation of capital across investment styles, rather than direct stock picking. This selection of datasets is fairly representative, as it covers the factors most commonly used in academic finance and investment practice, including industry, size, value, profitability, investment, momentum, and reversal. It is worth noting that the dimensionality (i.e., $N$) of the investable universe differs substantially across datasets. The smallest consists of only ten industry portfolios, whereas the largest dataset with $N = 375$ contains three $10 \times 10$ two-way sorted portfolios between size and book-to-market/profit/investment, and three $5 \times 5$ two-way sorted portfolios between size and prior returns (i.e., momentum, long-reversal, and short-reversal). Our sample spans the period from 1998 to 2020, which covers the boom and bust of the dot-com bubble, the Great Recession, the European sovereign debt crisis, and the beginning period of the COVID-19 pandemic.

We rebalance each portfolio at the end of each calendar month, with parameters trained using daily return observations from the one-year rolling window preceding the rebalancing date. The out-of-sample daily portfolio returns are then used to carry out the conditional evaluation using the proposed conditional Sharpe ratio inequality test. The portfolios under consideration turn out to have quite different turnover rates, and hence, it is important to take their transaction costs into consideration. We follow a standard approach (see, e.g., DeMiguel, Garlappi, and Uppal (2009), Robert, Robert, and Jeffrey (2012), Yen (2016), and Ao, Li, and Zheng (2019)) by computing the

[17] See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Table 2: List of Datasets

| # | Dataset | $N$ | Abbreviation |
|---|---|---|---|
| 1 | Ten industry portfolios | 10 | Industry-10 |
| 2 | Twenty five size and book-to-market portfolios | 25 | Size/BM-25 |
| 3 | Twenty five size and operating profitability portfolios | 25 | Size/Profit-25 |
| 4 | Twenty five size and investment portfolios | 25 | Size/Investment-25 |
| 5 | Twenty five size and momentum portfolios | 25 | Size/Momentum-25 |
| 6 | Twenty five size and long-term reversal portfolios | 25 | Size/Long reversal-25 |
| 7 | Twenty five size and short-term reversal portfolios | 25 | Size/Short reversal-25 |
| 8 | One hundred size and book-to-market portfolios | 100 | Size/BM-100 |
| 9 | One hundred size and operating profitability portfolios | 100 | Size/Profit-100 |
| 10 | One hundred size and investment portfolios | 100 | Size/Investment-100 |
| 11 | Size/BM-100 & Size/Profit-100 & Size/Investment-100 | 300 | Portfolio-300 |
| 12 | Portfolio-300 & Size/Momentum-25 & Size/Long reversal-25 & Size/Short reversal-25 | 375 | Portfolio-375 |

*Note:* This table provides details on the datasets we analyze, including the description, the number of investable assets $N$ in each dataset, and the corresponding abbreviation. All datasets contain daily excess returns relative to the one-month Treasury bill rate, obtained from Kenneth R. French's data library.

monthly transaction cost as

$$\text{monthly transaction cost} = c \times \text{monthly turnover},$$

where $c$ denotes the unit transaction cost. Pinning down a "representative" level of $c$ is a formidable task because it involves many practical idiosyncrasies such as implicit and explicit brokerage commissions, technological expenses for efficient execution, bid-ask spread, market impact, and stock loan fees for short-selling, which may vary across assets, sample periods, market conditions, and/or investors. Being agnostic, we consider two cost levels with $c = 10$ or 50 basis points (bps), respectively. The former appears to be a fairly forgiving lower bound, whereas the latter has been adopted in prior work such as DeMiguel, Garlappi, and Uppal (2009). We refer to these two configurations for ease of discussion as the low-cost and high-cost scenarios, respectively. By convention, we account the monthly transaction cost as an operational expense in the preceding trading month, and distribute it evenly (in a geometric manner) across the individual trading days. In so doing, we avoid attributing the monthly transaction cost solely to the rebalance day, which would otherwise lead to an uninteresting peculiarity in the daily time series of portfolio returns. Our empirical analysis will be based on the post-fee daily portfolio returns.

We use the VIX (in log) as the conditioning variable $X_t$, which is a reasonable choice as the VIX is arguably the most popular real-time barometer for financial market conditions. The data is available at the daily frequency for the 1998-2020 sample period considered here. Using data from the daily frequency instead of the lower monthly or quarterly frequency greatly improves the precision of volatility estimation, which in turn benefits the estimation of Sharpe ratios. We rescale the log VIX onto the $[-1, 1]$ interval using the same transformation as in the simulation. The significance level is fixed at $\alpha = 5\%$. We implement the test for $\rho \in \{0.001, 0.005\}$ as recommended. Since the results are similar, we focus on the $\rho = 0.001$ case here and relegate the $\rho = 0.005$ case to the Supplemental Appendix. As in the simulation study, the number of series terms is determined as $k = \max\{4, \lfloor 1.2n^{1/5} \rfloor\}$, resulting in $k = 6$. In the Supplemental Appendix, we further report results for $k = 5$ or 7. Our main empirical findings are robust to these changes.

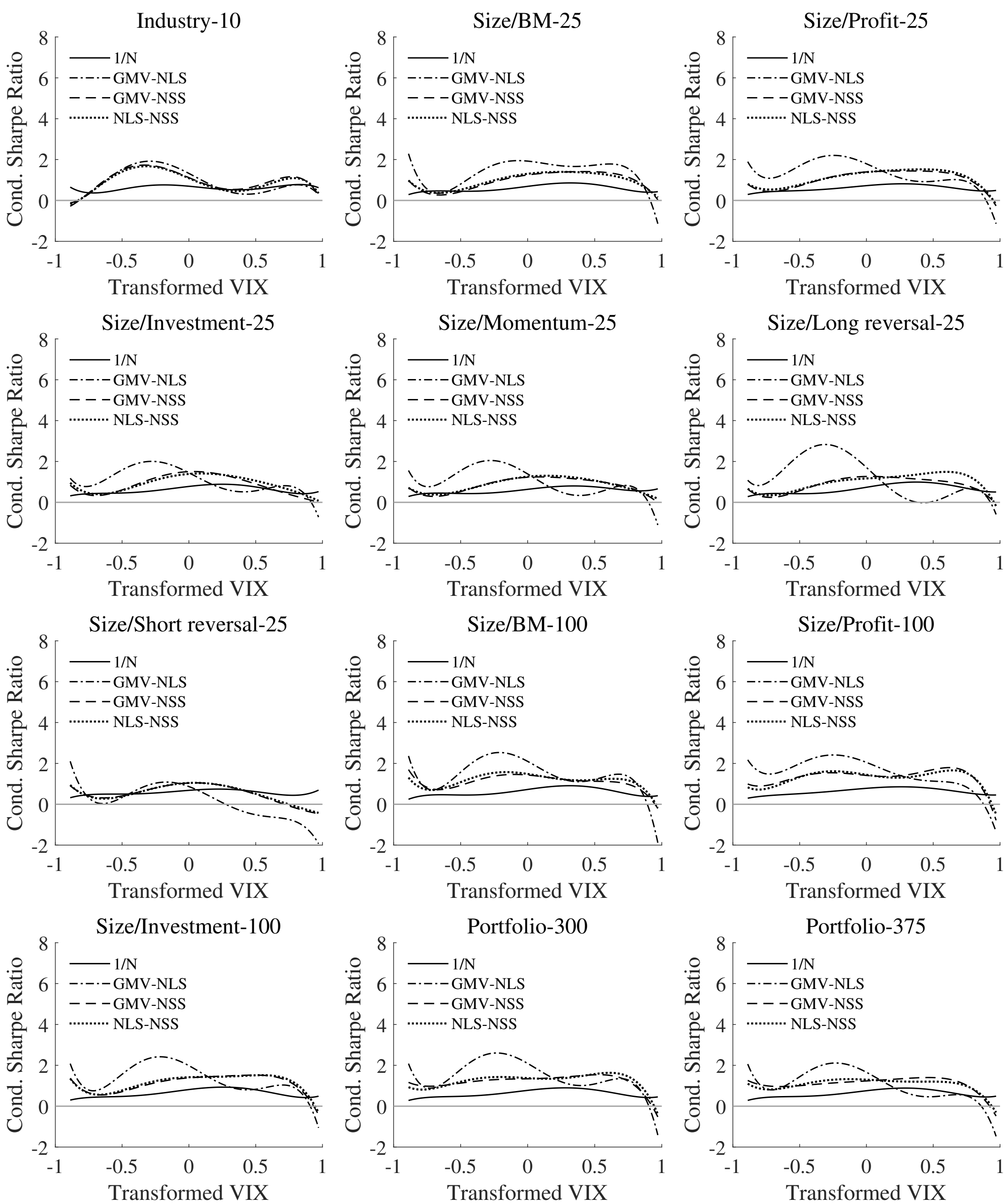
## 5.2 Empirical results

We treat each of the twelve datasets as a stand-alone investment environment and evaluate the relative conditional performance of the four competing trading strategies for each dataset separately. The null hypothesis (3.15) states that the benchmark portfolio weakly dominates all the other competing portfolios uniformly across the conditioning space, characterized here by the VIX. For symmetry, we rotate the benchmark role across the four portfolios and, as discussed in Section 2.1, form the CSMS as the collection of non-rejected benchmarks according to (3.14).

To set the stage, we first plot the conditional Sharpe ratio functions of the four portfolios estimated using each of the twelve datasets. Figure 2 and Figure 3 show the plots for the low-cost and high-cost scenarios, respectively. Since nonparametric estimates are invariably noisy near the boundary, we plot the estimated curves in restriction to the 2.5% and 97.5% empirical quantiles of the conditioning variable. The conditioning domain $\mathcal{X}$ is set accordingly for the conduct of the functional inequality test. The trimming is mainly employed here to prevent noisy estimates from obscuring the visualization and it has mild effect on the formal testing result. Testing results without trimming are in fact identical and reported in the Supplemental Appendix for completeness; see Table S2.

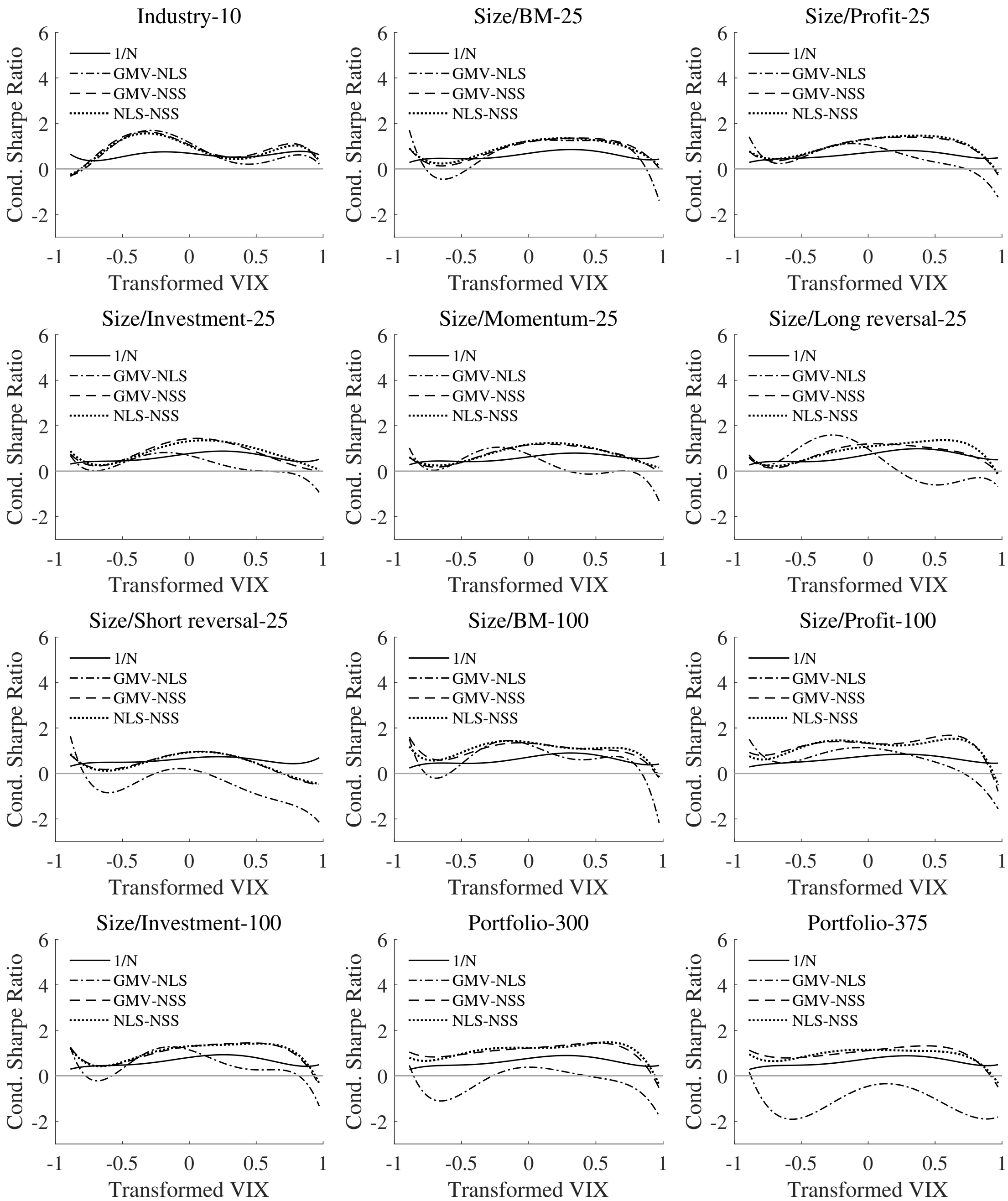
The plots show that the conditional Sharpe ratio of the $1/N$ portfolio varies only moderately across the conditioning space, evidenced by the nearly flat functional estimates. The other two long-only portfolios, GMV-NSS and NLS-NSS, also have relatively stable conditional Sharpe ratios across different conditioning states. However, the estimated Sharpe ratio of GMV-NLS exhibits a considerably greater level of state dependence. Further comparing the estimated curves between the low-cost and high-cost schemes, we see that increasing the transaction cost tends to worsen the

Figure 2: Conditional Sharpe Ratios for All Strategies: Low Cost Scenario




*Note:* This figure plots the estimated conditional Sharpe ratio functions for the four portfolios when the unit transaction cost is $c = 10$ bps.

Figure 3: Conditional Sharpe Ratios for All Strategies: High Cost Scenario




*Note:* This figure plots the estimated conditional Sharpe ratio functions for the four portfolios when the unit transaction cost is $c = 50$ bps.

performance of the three GMV portfolios relative to the $1/N$ portfolio, which is not surprising as the latter passive strategy requires less rebalancing. This effect is relatively mild for the two long-only GMV portfolios, which may be explained by the reduced amount of transactions resulting from the short-sale constraints. In contrast, GMV-NLS does not involve any direct constraint on portfolio weights, requires more rebalancing, and suffers the most from a higher level of transaction cost.

Turning next to the formal inference, we report results from the conditional Sharpe ratio inequality test in Table 3. The top and bottom panels correspond to the low-cost and high-cost scenarios, respectively. The rows correspond to different datasets and the columns indicate the benchmark trading strategy. If a portfolio—when used as the benchmark—is not rejected for a given dataset, we indicate this non-rejection with a check mark ($\checkmark$) in the table. Hence, for each dataset/row, the collection of check-marked trading strategies form the CSMS. A larger number of check marks on a column may thus be regarded as a higher "score" for the corresponding trading strategy. Below, we summarize the findings from this score board.

Arguably the most important finding is that NLS-NSS, the long-only global minimum variance portfolio implemented using the nonlinear shrinkage covariance matrix, is the best trading strategy in the sense that it "passes" the conditional evaluation test for the largest number of datasets (eleven out of twelve). The uniform superiority of this portfolio is rejected by the functional inequality test only for the Size/Short reversal-25 dataset. Importantly, this finding obtains under both low-cost and high-cost scenarios, suggesting that the portfolio achieves superior performance without "too much" trading.

To better appreciate the superior performance of the NLS-NSS portfolio, it is instructive to connect more precisely these testing results with the functional estimates plotted in Figures 2 and 3. We compare NLS-NSS with each of the other three portfolios in turn. First note that the estimated conditional Sharpe ratio functions of NLS-NSS are almost always above those of the $1/N$ portfolio, often by an economically nontrivial amount. The only notable exception is the Size/Short reversal-25 dataset, for which NLS-NSS is rejected as a superior benchmark because it significantly underperforms the $1/N$ portfolio for certain high-VIX states. Secondly, Figure 3 shows that when the transaction cost is high, NLS-NSS typically dominates GMV-NLS, sometimes quite substantially. The comparison is less clear under the low-cost scenario depicted in Figure 2. When the cost is low, GMV-NLS appears to have higher conditional Sharpe ratios than NLS-NSS over low-VIX states. However, since the nonparametric estimates for the GMV-NLS portfolio is relatively noisy, this "raw" evidence turns out to be insufficient for rejecting NLS-NSS as a superior benchmark. Finally, the comparison between the NLS-NSS and GMV-NSS portfolios is more subtle

Table 3: Conditional Sharpe Ratio Inequality Tests for Alternative Portfolios

| | *Benchmark Trading Strategy* | | | |
|---|---|---|---|---|
| | 1/N | GMV-NLS | GMV-NSS | NLS-NSS |
| | *Low Transaction Cost (10 bps)* | | | |
| Industry-10 | ✓ | ✓ | ✓ | ✓ |
| Size/BM-25 | | ✓ | ✓ | ✓ |
| Size/Profit-25 | | ✓ | ✓ | ✓ |
| Size/Investment-25 | | ✓ | ✓ | ✓ |
| Size/Momentum-25 | | ✓ | ✓ | ✓ |
| Size/Long reversal-25 | ✓ | ✓ | ✓ | ✓ |
| Size/Short reversal-25 | ✓ | | | |
| Size/BM-100 | | | ✓ | ✓ |
| Size/Profit-100 | ✓ | ✓ | ✓ | ✓ |
| Size/Investment-100 | | ✓ | | ✓ |
| Portfolio-300 | | ✓ | | ✓ |
| Portfolio-375 | | ✓ | | ✓ |
| | *High Transaction Cost (50 bps)* | | | |
| Industry-10 | ✓ | ✓ | ✓ | ✓ |
| Size/BM-25 | | ✓ | ✓ | ✓ |
| Size/Profit-25 | | ✓ | ✓ | ✓ |
| Size/Investment-25 | | | ✓ | ✓ |
| Size/Momentum-25 | | | ✓ | ✓ |
| Size/Long reversal-25 | ✓ | ✓ | ✓ | ✓ |
| Size/Short reversal-25 | ✓ | | | |
| Size/BM-100 | | | ✓ | ✓ |
| Size/Profit-100 | ✓ | ✓ | | ✓ |
| Size/Investment-100 | ✓ | ✓ | | ✓ |
| Portfolio-300 | ✓ | | | ✓ |
| Portfolio-375 | ✓ | | | ✓ |

*Note:* This table reports the results of the conditional Sharpe ratio inequality test. The conditioning variable is the log(VIX). Columns correspond to the benchmark strategies and rows specify the datasets. The ✓ symbol indicates that the null hypothesis of the benchmark's uniform superiority is not rejected at 5% significance level.

in that their estimated conditional Sharpe ratio functions appear similar. But their difference is also estimated accurately, permitting a statistically informative discrimination. A closer examination reveals that the GMV-NSS portfolio is rejected more often than NLS-NSS mainly due to the former's relative underperformance over high-VIX states and when the dimensionality $N$ of the investable universe is relatively large. This highlights the circumstance under which Ledoit and Wolf's (2017) nonlinear shrinkage method may help refine the classic GMV-NSS portfolio.

The performance of the GMV-NLS portfolio worth some additional discussion. As noted above, this portfolio suffers the most from a high level of transaction cost. Since it permits short selling, which tends to be more costly to execute, the high-cost scenario is not an ignorable possibility. In such an environment, applying the nonlinear shrinkage method alone does not generate satisfactory portfolio performance, especially when $N$ is large, as we can clearly see from the plots on the bottom row of Figure 3. This may not be surprising, because the nonlinear shrinkage method is designed to mitigate the estimation risk of the unknown variance-covariance matrix, without attempting to explicitly tame a potentially high transaction cost. On the other hand, for investors who are able to maintain the transaction cost at a low level, the plots in Figure 2 suggest that the conditional Sharpe ratio of GMV-NLS may be much higher than those of the other (long-only) portfolios under tranquil (i.e., low VIX) market conditions. Further corroborating these observations, the testing results shown on the top panel of Table 3 indicate that the GMV-NLS portfolio is in fact a close runner-up on the score board under the low-cost scenario, as it passes the functional inequality test for ten out of twelve datasets.

Finally, we note that the relative standing of the $1/N$ portfolio also depends nontrivially on the level of transaction cost. The top panel of Table 3 shows that under the low-cost scenario the $1/N$ portfolio is rejected as a superior benchmark for eight datasets, whereas the other portfolios are rejected notably less often. Correspondingly, the estimated conditional Sharpe ratios of this passive portfolio are often the lowest over large conditioning regions as shown in Figure 2. The evidence suggests that the $1/N$ portfolio is generally inferior to the minimum variance portfolios in a low-cost environment. The performance of the $1/N$ portfolio is more impressive when the transaction cost is high. From the bottom panel of Table 3, we see that it passes the test for more than half of the datasets, and measured as such, its overall performance is comparable with that of GMV-NLS and GMV-NSS; it is also interesting to note that the comparative advantage of the $1/N$ portfolio stems from investment problems with larger $N$. These findings may be related to the evidence documented by DeMiguel, Garlappi, and Uppal (2009) that many well-known and seemingly more sophisticated strategies cannot consistently outperform the "naive" $1/N$ portfolio. The authors consider the same 50-bps transaction cost but draw inference using unconditional evaluation methods. While their empirical evidence indicates the on-average adequacy of the $1/N$ portfolio, our finding further suggests a uniform sense of adequacy of this simple strategy for several datasets, as the null of its uniform superiority cannot be rejected at the given 5% level.

To sum up, the empirical exercise illustrates the practical applicability of the proposed test. The estimated conditional Sharpe ratio functions often exhibit nontrivial state dependence and the proposed functional inequality test provides a formal way of conducting state-dependent evaluation. The main takeaway of our empirical finding is that both short-sale constraint and nonlinear shrinkage are useful for improving portfolio performance, as the resulting NLS-NSS portfolio generates high conditional Sharpe ratios across most conditioning states and almost always passes the conditional Sharpe ratio inequality test. The short-sale constraint has an effect of reducing transactions and the associated cost, and nonlinear shrinkage offers further improvement over volatile states. Our nonparametric estimates and the formal inference provide clear evidence that the NLS-NSS portfolio generally outperform the simple $1/N$ portfolio, although the latter sometimes appear adequate (up to the available statistical precision) in trading environments with high transaction cost.

# 6 Conclusion

The contribution of this paper is to develop a uniformly valid test for general functional inequalities while accommodating general forms of time series dependence. The econometric framework naturally extends the important problem of testing (conditional) moment inequalities and affords a broader range of empirical applications pertaining to the evaluation of decisions. As a by-product of our analysis, we offer a new alternative—in the spirit of Romano, Shaikh, and Wolf (2014)—for testing conditional moment inequalities, complementing the existing approaches of Andrews and Shi (2013) and Chernozhukov, Lee, and Rosen (2013), and shed new light on the uniformity issue especially for times series applications. The econometric theory relies on a new uniform Gaussian coupling for general mixingale processes, capturing both martingale-type and mixing-type dependence as special cases, and readily accommodates most applications in economics and finance. This result is of independent interest and likely useful for the other inferential settings as well. We study in detail a specific functional inequality test for comparing the conditional Sharpe ratio functions of asset portfolios and provide empirical evidence for the uniform superiority of a shrinkage-based long-only global minimum variance portfolio using twelve datasets with distinct features.

## Appendix: Proofs

We prove Theorem 1 and Proposition 1, relying on several technical lemmas given in the Supplemental Appendix. Throughout the proofs, $K$ denotes a generic finite positive constant that may change from line to line.

Proof of Theorem 1. We first introduce some notation. Since $\mathcal{X}$ is a compact subset of $\mathbb{R}^{d_x}$, we can find a sequence of finite subsets $\mathcal{X}_n$ of $\mathcal{X}$ such that $\min_{x' \in \mathcal{X}_n} \sup_{x \in \mathcal{X}} \|x' - x\| \leq n^{-3}$ and the cardinality of $\mathcal{X}_n$ satisfies $|\mathcal{X}_n| \leq K n^{3d_x}$. We then set

$$\begin{aligned} \widetilde{T}_n^* &\equiv \max\left\{ \max_{j \in \mathcal{J}} \sup_{x \in \mathcal{X}} \frac{\widehat{\mathbb{G}}_{j,n}^*(x) + n^{1/2} u_j(x)}{\hat{s}_j(x)}, 0 \right\}, \\ T_n^0 &\equiv \max\left\{ \max_{j \in \mathcal{J}} \sup_{x \in \mathcal{X}_n} \frac{\mathbb{G}_{j,n}^*(x) + n^{1/2} u_j(x)}{s_j(x)}, 0 \right\}, \end{aligned}$$

where $u_j(x) \equiv \min\{\ell(h_j, x), 0\}$. For any $q \in (0,1)$, let $cv^*(q)$ denote the $1-q$ quantile of $\widetilde{T}_n^*$ conditional on data and $cv^0(q)$ denote the $1-q$ quantile of $T_n^0$.

We now prove the assertion in part (a). By Lemma S3 in the Supplemental Appendix, we have

$$\inf_{P \in \mathcal{P}_n} P\left( \min_{j \in \mathcal{J}} \inf_{x \in \mathcal{X}} \left( \hat{u}_j(x) - u_j(x) \right) \geq 0 \right) \geq 1 - \rho - o(1),$$

which implies that

$$\widetilde{T}_n^* \leq \widehat{T}_n^* \quad \text{and} \quad cv^*(\alpha - \rho) \leq \hat{c}_n \tag{A.1}$$

hold for any $\alpha \in (0, 1/2)$ and $\rho \in (0, \alpha)$ with probability at least $1 - \rho - o(1)$ uniformly over $P \in \mathcal{P}_n$. Therefore, for some sequence $v_n = o(1)$,

$$\begin{aligned} \sup_{P \in \mathcal{P}_{0,n}} P\left( \widehat{T}_n > \hat{c}_n \right) &\leq \sup_{P \in \mathcal{P}_{0,n}} P\left( \widehat{T}_n > cv^*(\alpha - \rho) \right) + \rho + o(1) \\ &\leq \sup_{P \in \mathcal{P}_{0,n}} P\left( \widehat{T}_n > cv^0(\alpha - \rho + v_n) \right) + \rho + o(1) \\ &= 1 - \inf_{P \in \mathcal{P}_{0,n}} P\left( \widehat{T}_n \leq cv^0(\alpha - \rho + v_n) \right) + \rho + o(1), \end{aligned} \tag{A.2}$$

where the second inequality is by Lemma S4 in the Supplemental Appendix. Moreover, by Lemma S5 in the Supplemental Appendix,

$$\begin{aligned} \inf_{P \in \mathcal{P}_{0,n}} P\left( \widehat{T}_n \leq cv^0(\alpha - \rho + v_n) \right) &\geq \inf_{P \in \mathcal{P}_{0,n}} P\left( T_n^0 \leq cv^0(\alpha - \rho + v_n) \right) - o(1) \\ &\geq 1 - \alpha + \rho - o(1) \end{aligned}$$

which together with (A.2) implies that

$$\sup_{P \in \mathcal{P}_{0,n}} P\left( \widehat{T}_n > \hat{c}_n \right) \leq \alpha + o(1). \tag{A.3}$$

This proves the assertion in part (a) of the theorem.

We next turn to the proof of part (b). For ease of reference, we collect a few preliminary estimates from the proof of Lemma S1 in the Supplemental Appendix:

$$\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left|\frac{\hat{s}_j(x)}{s_j(x)}-1\right| = o_{\mathcal{P}}(k^{-1/2}\log(n)^{-1}) \quad \text{and} \quad K^{-1}\leq\frac{s_j(x)}{\|\dot{\ell}(h_j^*,x)\|}\leq K, \tag{A.4}$$

and

$$K^{-1}-o_{\mathcal{P}}(1)\leq\lambda_{\min}(\widehat{\Omega})\leq\lambda_{\max}(\widehat{\Omega})\leq K+o_{\mathcal{P}}(1). \tag{A.5}$$

By Lemma S1 in the Supplemental Appendix,

$$\begin{aligned}
\widehat{T}_n &= \max\left\{\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left(\frac{n^{1/2}(\ell(\hat{h}_j,x)-\ell(h_j,x))}{\hat{s}_j(x)}+\frac{n^{1/2}\ell(h_j,x)}{\hat{s}_j(x)}\right),0\right\}\\
&\geq \max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left(\frac{\mathbb{G}_{j,n}(x)}{s_j(x)}+\frac{n^{1/2}\ell(h_j,x)}{\hat{s}_j(x)}\right)-o_{\mathcal{P}}(\log(n)^{-1})\\
&\geq \max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\frac{n^{1/2}\ell(h_j,x)}{s_j(x)}(1-o_{\mathcal{P}}(\log(n)^{-1})-\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left|\frac{\mathbb{G}_{j,n}(x)}{s_j(x)}\right|-o_{\mathcal{P}}(\log(n)^{-1})\\
&\geq \max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\frac{n^{1/2}\ell(h_j,x)}{s_j(x)}(1-o_{\mathcal{P}}(\log(n)^{-1})-O_{\mathcal{P}}(\log(k)^{1/2}) \qquad (A.6)
\end{aligned}$$

where the second inequality is by (A.4) and the third inequality is by

$$\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left|\frac{\mathbb{G}_{j,n}(x)}{s_j(x)}\right| = O_{\mathcal{P}}(\log(k)^{1/2}), \tag{A.7}$$

which holds by Lemma S10 in the Supplemental Appendix and Markov's inequality.

Since $\hat{u}_j(x)\leq 0$ for any $j\in\mathcal{J}$ and any $x\in\mathcal{X}$, we have

$$\begin{aligned}
\widehat{T}_n^* &\equiv \max\left\{\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left(\frac{\widehat{\mathbb{G}}_{j,n}^*(x)+n^{1/2}\hat{u}_j(x)}{\hat{s}_j(x)}\right),0\right\}\\
&\leq \max\left\{\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\frac{\widehat{\mathbb{G}}_{j,n}^*(x)}{\hat{s}_j(x)},0\right\}\leq\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left|\frac{\widehat{\mathbb{G}}_{j,n}^*(x)}{\hat{s}_j(x)}\right|\\
&\leq \max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left|\frac{\widehat{\mathbb{G}}_{j,n}^*(x)}{\hat{s}_j(x)}-\frac{\mathbb{G}_{j,n}^*(x)}{s_j(x)}\right|+O_{\mathcal{P}}(\log(k)^{1/2}) \qquad (A.8)
\end{aligned}$$

where the last inequality follows from (A.7). By the triangle inequality, (A.4) and (A.7),

$$\begin{aligned}
\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left|\frac{\widehat{\mathbb{G}}_{j,n}^*(x)}{\hat{s}_j(x)}-\frac{\mathbb{G}_{j,n}^*(x)}{s_j(x)}\right| &\leq \max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left|\frac{\widehat{\mathbb{G}}_{j,n}^*(x)-\mathbb{G}_{j,n}^*(x)}{\hat{s}_j(x)}\right|\\
&\quad+\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left|\frac{\mathbb{G}_{j,n}^*(x)}{\hat{s}_j(x)}-\frac{\mathbb{G}_{j,n}^*(x)}{s_j(x)}\right|\\
&= \max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left|\frac{\widehat{\mathbb{G}}_{j,n}^*(x)-\mathbb{G}_{j,n}^*(x)}{\hat{s}_j(x)}\right|+o_{\mathcal{P}}(\log(n)^{-1/2}). \qquad (A.9)
\end{aligned}$$

By the Cauchy–Schwarz inequality,

$$\begin{aligned}\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left|\frac{\widehat{\mathbb{G}}^*_{j,n}(x)-\mathbb{G}^*_{j,n}(x)}{\hat{s}_j(x)}\right| &\leq \max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\frac{\lambda^{1/2}_{\max}(\widehat{\Omega})\|\dot{\ell}(\hat{h}_j,x)-\dot{\ell}(h^*_j,x)\|}{\hat{s}_j(x)}\|\mathcal{N}^*_m\| \\ &\quad+\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\frac{\|\dot{\ell}(h^*_j,x)\|\|\widehat{\Omega}-\Omega\|_S}{\lambda^{1/2}_{\min}(\Omega)\hat{s}_j(x)}\|\mathcal{N}^*_m\|. \end{aligned}\tag{A.10}$$

Since $\|\mathcal{N}^*_m\| = O_{\mathcal{P}}(m^{1/2})$ and $m = O(k)$, by Assumptions 2(iii, v), (A.4), and (A.5), we have

$$\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\frac{\lambda^{1/2}_{\max}(\widehat{\Omega})\|\dot{\ell}(\hat{h}_j,x)-\dot{\ell}(h^*_j,x)\|}{\hat{s}_j(x)}\|\mathcal{N}^*_m\| = o_{\mathcal{P}}(\log(n)^{-1/2}). \tag{A.11}$$

Similarly, we can show that

$$\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\frac{\|\dot{\ell}(h^*_j,x)\|\|\widehat{\Omega}-\Omega\|_S}{\lambda^{1/2}_{\min}(\Omega)\hat{s}_j(x)}\|\mathcal{N}^*_m\| = o_{\mathcal{P}}(\log(n)^{-1/2}),$$

which together with (A.9), (A.10), and (A.11) implies that

$$\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\left|\frac{\widehat{\mathbb{G}}^*_{j,n}(x)}{\hat{s}_j(x)}-\frac{\mathbb{G}^*_{j,n}(x)}{s_j(x)}\right| = o_{\mathcal{P}}(\log(n)^{-1/2}). \tag{A.12}$$

Combining (A.8) and (A.12), we obtain $\widehat{T}^*_n = O_{\mathcal{P}}(\log(k)^{1/2})$ and hence $\hat{c}_n = O_{\mathcal{P}}(\log(k)^{1/2})$. Therefore by (A.6),

$$\begin{aligned} P\left(\widehat{T}_n > \hat{c}_n\right) &\geq P\left(\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\frac{n^{1/2}\ell(h_j,x)}{s_j(x)} > O_{\mathcal{P}}(\log(k)^{1/2})\right) - o(1) \\ &\geq P\left(\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}\ell(h_j,x) > O_{\mathcal{P}}(\xi_k\log(k)^{1/2}n^{-1/2})\right) - o(1) \end{aligned}\tag{A.13}$$

where the second inequality is by (A.4) and Assumption 1(iii). The claim of the theorem now follows from (2.12) and (A.13). *Q.E.D.*

Proof of Proposition 1. We first verify Assumption 1. Recall that under the present setting we have $\ell(h_j,x) = h_j(x) = \mathrm{SR}_j(x) - \mathrm{SR}_0(x)$ and $\dot{\ell}(h^*_j,x) = \psi(x)$. By Assumption 3(i) and the compactness of $\mathcal{X}$,

$$\max_{j\in\mathcal{J}}\sup_{x\in\mathcal{X}}|\ell(h_j,x)| \leq K$$

uniformly over $P \in \mathcal{P}_n$, which verifies Assumption 1(i). By Assumption 3(i), we have for any $x_1, x_2 \in \mathcal{X}$ and any $j \in \mathcal{J}$,

$$\begin{aligned} |\mathrm{SR}_j(x_1) - \mathrm{SR}_j(x_2)| &\leq \left|\frac{\mu_j(x_1)-\mu_j(x_2)}{\sigma_j(x_1)}\right| + \left|\frac{\mu_j(x_2)(\sigma^2_j(x_2)-\sigma^2_j(x_1))}{\sigma_j(x_1)\sigma_j(x_2)(\sigma_j(x_2)+\sigma_j(x_1))}\right| \\ &\leq K\left(|\mu_j(x_1)-\mu_j(x_2)| + \left|\sigma^2_j(x_1)-\sigma^2_j(x_2)\right|\right) \leq K\left\|x_1-x_2\right\|, \end{aligned}$$

which shows Assumption 1(ii). Since $\dot{\ell}(h_j^*, x) = \psi(x)$, we can let $\xi_k \equiv \xi_{0,k}$ and $\zeta_k^L \equiv \xi_{1,k}$. Then Assumptions 1(iii, iv) follows from Assumption 3(vii).

Next, we verify Assumption 2. Recall that for each $j \in \mathcal{J}$,

$$\ell(\hat{h}_j, x) = \widehat{\mathrm{SR}}_j(x) - \widehat{\mathrm{SR}}_0(x) = \psi(x)^\top(\hat{b}_j - \hat{b}_0) = \psi(x)^\top(\tilde{L}_j - \tilde{L}_0)\hat{b}, \tag{A.14}$$

and so, $\hat{\beta}_j = (\tilde{L}_j - \tilde{L}_0)\hat{b}$. Therefore, by Lemma S15 in the Supplemental Appendix,

$$n^{1/2}(\hat{\beta} - \beta^*) = \tilde{L}(I_{J+1} \otimes Q^{-1})n^{-1/2}\sum_{t=1}^{n} v_{t+1} + o_{\mathcal{P}}(\log(n)^{-1}), \tag{A.15}$$

where $\beta^* \equiv \tilde{L}b^*$ and $b^* \equiv (b^*_{\mathrm{SR}_j})_{j\in\mathcal{J}_0}$. This representation together with Assumptions 3(ii, iv) implies that Assumption 2(i) holds with

$$\Omega = \tilde{L}(I_{J+1} \otimes Q^{-1})\Omega_v(I_{J+1} \otimes Q^{-1})\tilde{L}^\top. \tag{A.16}$$

By the definition of $s_j(x)$ in (2.8) and the expression of $\Omega$ above, we find

$$s_j(x) = \sqrt{\psi(x)^\top\tilde{L}(I_{J+1} \otimes Q^{-1})\Omega_v(I_{J+1} \otimes Q^{-1})\tilde{L}^\top\psi(x)}. \tag{A.17}$$

Let $\mathrm{diag}((A_j)_{j\in\mathcal{J}})$ denote the block diagonal matrix with matrix $A_j$ on its $j$th diagonal entry. Then by the definition of $\Psi_\epsilon$ and $\Omega_v$, we can write

$$\Omega_v = \mathrm{diag}((I_k, -\mathbb{E}[\tilde{\Upsilon}_j])_{j\in\mathcal{J}})\,\Psi_\epsilon\,\mathrm{diag}((I_k, -\mathbb{E}[\tilde{\Upsilon}_j])_{j\in\mathcal{J}})^\top. \tag{A.18}$$

Note that

$$\mathrm{diag}((I_k, -\mathbb{E}[\tilde{\Upsilon}_j])_{j\in\mathcal{J}})\mathrm{diag}((I_k, -\mathbb{E}[\tilde{\Upsilon}_j])_{j\in\mathcal{J}})^\top = \mathrm{diag}((I_k + \mathbb{E}[\tilde{\Upsilon}_j]\mathbb{E}[\tilde{\Upsilon}_j^\top])_{j\in\mathcal{J}}). \tag{A.19}$$

By Assumptions 3(i, ii)

$$\lambda_{\max}(\mathbb{E}[\tilde{\Upsilon}_j]\mathbb{E}[\tilde{\Upsilon}_j^\top]) \leq \lambda_{\min}^{-1}(Q)\lambda_{\max}\left(n^{-1}\sum_{t=1}^{n}\mathbb{E}\left[\frac{r_{j,t+1}^2}{4\sigma_j^6(X_t)}\psi(X_t)\psi(X_t)^\top\right]\right) \leq K, \tag{A.20}$$

which implies that

$$1 \leq \min_{j\in\mathcal{J}}\lambda_{\min}(I_k + \mathbb{E}[\tilde{\Upsilon}_j]\mathbb{E}[\tilde{\Upsilon}_j^\top]) \leq \max_{j\in\mathcal{J}}\lambda_{\max}(I_k + \mathbb{E}[\tilde{\Upsilon}_j]\mathbb{E}[\tilde{\Upsilon}_j^\top]) \leq K. \tag{A.21}$$

Using Assumption 3(ii), (A.18), (A.19), and (A.21), we have

$$K^{-1} \leq \lambda_{\min}(\Omega_v) \leq \lambda_{\max}(\Omega_v) \leq K. \tag{A.22}$$

By Assumption 3(ii) and (A.22),

$$K^{-1} \leq \frac{s_j(x)}{\|\psi(x)\|} \leq K \tag{A.23}$$

for any $x \in \mathcal{X}$ and $j \in \mathcal{J}$. Since $\|\psi(x)\|$ is bounded away from zero for any $x$ by Assumption 3(vii), Assumption 3(iii) and (A.23) imply that

$$\frac{n^{1/2}(\mathrm{SR}_j(x) - \mathrm{SR}_j^*(x))}{s_j(x)} = O(\log(n)^{-2}), \tag{A.24}$$

where $\mathrm{SR}_j^*(x) \equiv \psi(x)^\top b_{\mathrm{SR}_j}^*$. Using (A.14), (A.24), and the definition of $\hat{\beta}_j$ and $\beta_j^* \equiv (\tilde{L}_j - \tilde{L}_0)b^*$, we obtain

$$\frac{n^{1/2}(\ell(\hat{h}_j, x) - \ell(h_j, x))}{s_j(x)} = \frac{n^{1/2}\psi(x)^\top(\hat{\beta}_j - \beta_j^*)}{s_j(x)} + o(\log(n)^{-1})$$

uniformly over $x \in \mathcal{X}$ and $P \in \mathcal{P}_n$, which verifies Assumption 2(ii).

The estimator of $\Omega$ is constructed as

$$\widehat{\Omega} = \tilde{L}(I_{J+1} \otimes \widehat{Q}^{-1})^\top \widehat{\Omega}_v (I_{J+1} \otimes \widehat{Q}^{-1})\tilde{L}^\top.$$

Using Assumptions 3(ii, v) and (A.22), we can show that

$$\|\widehat{\Omega} - \Omega\|_S = o_{\mathcal{P}}(k^{-1/2}\log(n)^{-1}). \tag{A.25}$$

Since $\dot{\ell}(\psi(\cdot)^\top\beta_j, x) = \psi(\cdot)$, $\dot{\ell}(\hat{h}_j, x) - \dot{\ell}(h_j^*, x) = 0$ trivially, which together with (A.25) shows Assumption 2(iii). Assumption 2(iv) follows directly from Assumption 3(ii), (A.22), and the definition of $\Omega$ in (A.16). Assumption 2(v) is assumed in Assumption 3(vii). This finishes the verification of Assumption 2. *Q.E.D.*

# References


Andrews, D. W. K. (1984): "Non-Strong Mixing Autoregressive Processes," *Journal of Applied Probability*, pp. 930–934.

——— (1985): "A Nearly Independent, but Non-Strong Mixing, Triangular Array," *Journal of Applied Probability*, pp. 729–731.

——— (1988): "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables," *Econometric Theory*, 4(3), 458–467.

——— (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59(2), 307–345.

Andrews, D. W. K., and P. J. Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80(6), 2805–2826.

Andrews, D. W. K., and P. Guggenberger (2009): "Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25(3), 669–709.

Andrews, D. W. K., and X. Shi (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81(2), 609–666.

——— (2017): "Inference Based on Many Conditional Moment Inequalities," *Journal of Econometrics*, 196(2), 275–287.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Ao, M., Y. Li, and X. Zheng (2019): "Approaching Mean-Variance Efficiency for Large Portfolios," *Review of Financial Studies*, 32(7), 2890–2919.

Armstrong, T. B. (2015): "Asymptotically Exact Inference in Conditional Moment Inequality Models," *Journal of Econometrics*, 186(1), 51–65.

Bai, Y., A. Santos, and A. M. Shaikh (2022): "A Two-Step Method for Testing Many Moment Inequalities," *Journal of Business & Economic Statistics*, 40(3), 1070–1080.

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015): "Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," *Journal of Econometrics*, 186(2), 345 – 366, High Dimensional Problems in Econometrics.

Berger, R. L., and D. D. Boos (1994): "P Values Maximized Over a Confidence Set for the Nuisance Parameter," *Journal of the American Statistical Association*, 89(427), 1012–1016.

Canay, I. A., and A. M. Shaikh (2017): "Practical and Theoretical Advances in Inference for Partially Identified Models," *Advances in Economics and Econometrics*, 2, 271–306.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, chap. 76. Elsevier, 1 edn.

Chen, X., and T. M. Christensen (2015): "Optimal Uniform Convergence Rates and Asymptotic Normality for Series Estimators Under Weak Dependence and Weak Conditions," *Journal of Econometrics*, 188(2), 447–465.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013): "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-dimensional Random Vectors," *Annals of Statistics*, 41(6), 2786–2819.

——— (2014): "Gaussian Approximation of Suprema of Empirical Processes," *Annals of Statistics*, 42(4), 1564–1597.

——— (2017): "Central Limit Theorems and Bootstrap in High Dimensions," *Annals of Probability*, 45(4), 2309–2352.

——— (2019): "Inference on Causal and Structural Parameters using Many Moment Inequalities," *Review of Economic Studies*, 86(5), 1867–1900.

Chernozhukov, V., D. Chetverikov, K. Kato, and Y. Koike (2022): "Improved Central Limit Theorem and Bootstrap Approximations in High Dimensions," *Annals of Statistics, Forthcoming.*

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

Chetverikov, D. (2018): "Adaptive Tests of Conditional Moment Inequalities," *Econometric Theory*, 34(1), 186–227.

Chetverikov, D., A. Santos, and A. M. Shaikh (2018): "The Econometrics of Shape Restrictions," *Annual Review of Economics*, 10(1), 31–63.

Chetverikov, D., and D. Wilhelm (2017): "Nonparametric Instrumental Variable Estimation under Monotonicity," *Econometrica*, 85(4), 1303–1320.

Chiang, H. D., K. Kato, Y. Sasaki, and T. Ura (2021): "Linear Programming Approach to Nonparametric Inference under Shape Restrictions: with An Application to Regression Kink Designs," *arXiv preprint arXiv:2102.06586*.

Cox, G., and X. Shi (2019): "Simple Adaptive Size-exact Testing for Full-vector and Subvector Inference in Moment Inequality Models," *arXiv preprint arXiv:1907.06317*.

Davidson, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.

DeMiguel, V., L. Garlappi, and R. Uppal (2009): "Optimal versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?," *Review of Financial Studies*, 22(5), 1915–1953.

Fan, J., and Q. Yao (1998): "Efficient Estimation of Conditional Variance Functions in Stochastic Regression," *Biometrika*, 85(3), 645–660.

Fang, Z., and J. Seo (2021): "A Projection Framework for Testing Shape Restrictions That Form Convex Cones," *Econometrica*, 89(5), 2439–2458.

Ferson, W. E. (2010): "Investment Performance Evaluation," *Annual Review of Financial Economics*, 2(1), 207–234.

Giacomini, R., and H. White (2006): "Tests of Conditional Predictive Ability," *Econometrica*, 74(6), 1545–1578.

Guerard Jr, J. B. (2009): *Handbook of Portfolio Construction: Contemporary Applications of Markowitz Techniques*. Springer Science & Business Media.

Hansen, P. R. (2005): "A Test for Superior Predictive Ability," *Journal of Business & Economic Statistics*, 23(4), 365–380.

Imbens, G. W., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

Jagannathan, R., and T. Ma (2003): "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," *Journal of Finance*, 58(4), 1651–1683.

Ledoit, O., and M. Wolf (2017): "Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks," *Review of Financial Studies*, 30(12), 4349–4388.

Lee, S., K. Song, and Y.-J. Whang (2013): "Testing Functional Inequalities," *Journal of Econometrics*, 172(1), 14–32.

Li, J., and Z. Liao (2020): "Uniform Nonparametric Inference for Time Series," *Journal of Econometrics*, 219(1), 28–51.

Li, J., Z. Liao, and R. Quaedvlieg (2022): "Conditional superior predictive ability," *Review of Economic Studies*, 89(2), 843–875.

McLeish, D. L. (1975): "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability*, 3(5), 829–839.

Molinari, F. (2020): "Microeconometrics with Partial Identification," *Handbook of Econometrics*, 7, 355–486.

Newey, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79(1), 147 – 168.

Pötscher, B. M., and I. R. Prucha (2013): *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Springer Science & Business Media.

Robert, E., F. Robert, and R. Jeffrey (2012): "Measuring and Modeling Execution Cost and Risk," *Journal of Portfolio Management*, 38(2), 14–28.

Romano, J. P. (2004): "On Non-parametric Testing, the Uniform Behaviour of the t-test, and Related Problems," *Scandinavian Journal of Statistics*, 31(4), 567–584.

Romano, J. P., and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138(9), 2786–2807.

Romano, J. P., A. M. Shaikh, and M. Wolf (2014): "A Practical Two-step Method for Testing Moment Inequalities," *Econometrica*, 82(5), 1979–2002.

Silvapulle, M. J. (1996): "A Test in the Presence of Nuisance Parameters," *Journal of the American Statistical Association*, 91(436), 1690–1693.

White, H. (2000): "A Reality Check for Data Snooping," *Econometrica*, 68(5), 1097–1126.

White, H., and R. Gallant (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Basil Blackwell.

Wu, W. B. (2005): "Nonlinear System Theory: Another Look at Dependence," *Proceedings of the National Academy of Sciences*, 102(40), 14150–14154.

YEN, Y.-M. (2016): "Sparse Weighted-Norm Minimum Variance Portfolios," *Review of Finance*, 20(3), 1259–1287.

ZHANG, D., AND W. B. WU (2017): "Gaussian Approximation for High Dimensional Time Series," *Annals of Statistics*, 45(5), 1895–1919.

ZHANG, X., AND G. CHENG (2018): "Gaussian Approximation for High Dimensional Vector Under Physical Dependence," *Bernoulli*, 24(4A), 2640 – 2675.