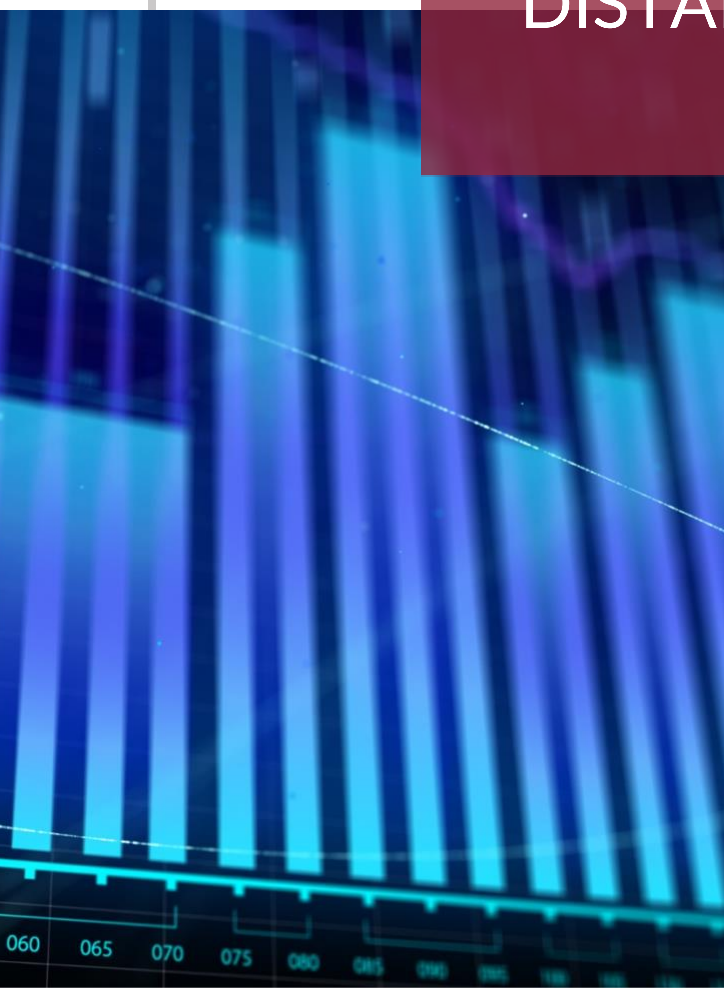
CIRANO
Knowledge into action

HANSEN-JAGANNATHAN DISTANCE WITH MANY ASSETS



MARINE CARRASCO
CHEIKH NOKHO

CS
2026s-13
WORKING PAPER

ISSN 2292-0838 (online version)

Hansen-Jagannathan distance with many assets

Marine Carrasco[*], *Cheikh Nokho*[†]

Abstract

This paper proposes methods to estimate and compare asset pricing models in settings with a large number of test assets. Models are specified through a linear stochastic discount factor (SDF). We propose two regularization schemes to extend the Hansen-Jagannathan distance to high-dimensional environments. In addition to stabilizing the inversion of the covariance matrix, the proposed regularizations admit an economic interpretation as relaxing the exact pricing restrictions, thereby accommodating market frictions.

We derive the asymptotic properties of the SDF parameter estimator under a double asymptotic framework in which both the cross-sectional and time dimensions grow. These results allow for inference on whether individual factors are priced. We further develop tests for comparing competing asset pricing models under misspecification, providing a formal procedure to identify the least misspecified model. The analysis covers both nested and non-nested specifications.

An empirical application compares 4 models using a dataset of 647 test portfolios.

Résumé

Cet article propose des méthodes pour estimer et comparer des modèles d'évaluation des actifs dans des contextes où le nombre d'actifs tests est élevé. Les modèles sont caractérisés par leur facteur d'actualisation stochastique (SDF) linéaire. Nous proposons deux méthodes de régularisation permettant d'étendre la distance de Hansen-Jagannathan aux environnements de grande dimension. Au-delà de la stabilisation de l'inversion de la matrice de covariance, les régularisations proposées admettent une interprétation économique comme un relâchement des restrictions exactes d'évaluation des actifs, permettant ainsi de prendre en compte les frictions de marché.

Nous établissons les propriétés asymptotiques de l'estimateur des paramètres du SDF dans un cadre de double asymptotique où les dimensions transversale et temporelle tendent toutes deux vers l'infini. Ces résultats permettent de mener des tests afin de déterminer si des facteurs individuels sont valorisés. Nous développons également des tests permettant de comparer différents modèles d'évaluation des actifs en présence d'erreurs de spécification, fournissant ainsi une procédure formelle pour identifier le modèle le moins mal spécifié. L'analyse couvre à la fois les spécifications emboîtées et non emboîtées.

Une application empirique consiste à comparer 4 modèles à l'aide de 647 portefeuilles tests.

[*]Université de Montréal, CIRANO Researcher and Fellow

[†] Université de Montréal

Keywords/Mots-clés: asset pricing models, misspecification, ridge, regularization, tests, Tikhonov, stochastic discount factor / modèles d'évaluation des actifs, spécification erronée, méthode Ridge, régularisation, tests, Tikhonov, facteur d'actualisation stochastique

JEL Codes/Codes JEL: C12, G12

Pour citer ce document / To quote this document
Carrasco, M. and Nokho, C. (2026). Hansen-Jagannathan distance with many assets (2026s-13, CIRANO). https://doi.org/10.54932/EVGB2502

1 Introduction

Asset pricing models are typically represented through a parametric stochastic discount factor (SDF), often specified as a linear function of a small number of risk factors. Given the large number of competing models proposed in the literature, it is essential to compare them using an objective and economically meaningful criterion.

A well-known measure of model misspecification is the Hansen-Jagannathan (HJ) distance, which measures the distance between a candidate pricing kernel and the closest valid one (see Hansen and Jagannathan (1997)). This distance is closely related to the Generalized Method of Moments (GMM) criterion, with the key difference that the weighting matrix is given by the inverse second moment matrix of the returns. Empirically, specification tests based on HJ distance are often rejected (Hodrick and Zhang (2001); Ludvigson (2013)), suggesting that most asset pricing models are misspecified. Consequently, rather than testing whether a model is exactly correct, a more relevant objective is to compare competing models and determine which one is less misspecified. In this context, the HJ distance provides a natural tool to rank models and assess whether particular factors are priced.

A central challenge in this comparison is that many models appear to perform similarly when evaluated on standard sets of test assets, such as the 25 size and book-to-market portfolios of Fama and French (1992). As emphasized by Daniel and Titman (2012) and Lewellen, Nagel, and Shanken (2010), these portfolios span a limited dimension of the return space and exhibit strong covariance structures, making it difficult to discriminate between models. This has led to the view that a larger and more diverse set of test assets is needed. For instance, Kan and Robotti (2009) augment the cross-section with industry portfolios but still find limited power, while Barras (2019) advocate using a very large number of micro portfolios. More generally, increasing the number of test assets improves the ability to detect misspecification, as the unconditional HJ distance approaches its conditional counterpart when the cross-section becomes sufficiently rich.

However, expanding the number of test assets raises some technical issues. The HJ distance relies on the inverse covariance matrix of returns, which becomes ill-conditioned when the number of assets is large relative to the time dimension. In practice, the sample covariance matrix is nearly singular or non-invertible when N is comparable to or exceeds T (Cochrane (2005)). As a result, the implied weighting matrix is unstable and can place excessive weight on directions associated with small eigenvalues, leading to economically meaningless portfolios and unreliable inference. This issue becomes particularly severe when individual stock returns or very large sets of portfolios are used as test assets.

This paper studies the evaluation and comparison of asset pricing models in such high-dimensional settings, explicitly allowing for model misspecification. Our approach is based on regularizing the HJ distance to stabilize the inversion of the covariance matrix. Specifically, we consider ridge and Tikhonov regularizations of the inverse covariance matrix (see Carrasco, Florens, and Renault (2007)). These regularizations yield a modified distance that can be interpreted as allowing for controlled pricing errors, which is consistent with the presence of transaction costs, liquidity frictions, or measurement error. Importantly, our objective is not to test whether the HJ distance is equal to zero, but rather to compare models through differences of (regularized) HJ distances, which remain well-defined even when the cross-section is large.

Our contributions are threefold. First, we introduce regularized versions of the HJ distance that remain well-behaved when the number of test assets is large, and we propose a data-driven procedure to select the regularization parameter. Second, we derive the asymptotic distribution of the SDF parameters obtained by minimizing the regularized distance under a double asymptotic framework where both N and T diverge. We allow for serial dependence and propose a consistent HAC estimator of the asymptotic variance. We show that these estimators are asymptotically normal and can be used to test whether specific factors are priced. Third, we develop inference procedures for comparing competing models based on differences in regularized HJ distances, treating separately the cases of nested and non-nested models. These results provide a formal basis for ranking models in environments where all models are potentially misspecified.

Monte Carlo simulations compare the power of the t-test based on Moore-Penrose inverse with that based on Tikhonov regularization and show that, while the power of the Moore-Penrose t-test is very low when N and T are of the same order, the regularized t-test maintains a high power. The other comparison tests are also shown to have good power. In an empirical application, we compare several prominent SDF specifications, including the Fama–French three- and five-factor models, the nonlinear model of Dittmar (2002), and the q5 model of Hou, Mo, Xue, and Zhang (2021). Our results indicate that the q5 model delivers significantly smaller pricing errors.

Our work relates to several strands of the literature at the intersection of asset pricing model evaluation and machine learning in finance. Several papers proposed methods to examine asset pricing misspecification (Hansen and Jagannathan (1997); Almeida and Garcia (2012); Almeida and Garcia (2017)). This paper is close to Kan and Robotti (2008) and Kan and Robotti (2009) who derived asymptotic distribution of the SDF parameter and model comparison methods using the HJ-distance under a misspecified setting using a finite number of test assets. Gagliardini and Ronchetti (2020) compare models by conditional Hansen-Jagannathan distances which involves a kernel smoothing step, see also Antoine, Proulx, and Renault (2020). Unlike their approach, we employ the unconditional version of the HJ-distance with many assets. However, when a very large number of test assets is used, the unconditional HJ distance should be close to the conditional one. Several papers also propose methods to either stabilize or improve the estimation of covariance matrices (Brodie, Daubechies, De Mol, and Loris (2009); Carrasco and Rossi (2016); Carrasco, Kone, and Noumon (2019); Ledoit and Wolf (2003); Ledoit and Wolf (2017); D′Hondt, De Winne, Ghysels, and Raymond (2020); Qiu and Otsu (2022); Lönn and Schotman (2024)). Ren and Shimotsu (2009) illustrate via simulations that shrinkage improves the finite sample properties of Hansen-Jagannathan distance. More recently, Chernov, Kelly, Malamud, and Schwab (2025) develop a test of portfolio efficiency in high-dimensional settings using ridge regularization and derive its asymptotic distribution accounting for regularization bias. Our approach is also related to Korsaye, Quaini, and Trojani (2025), who propose a strictly positive SDF that tolerates pricing errors; in contrast, we relax pricing restrictions without imposing positivity. Barillas and Shanken

(2018) study model comparison and show that test assets are irrelevant when only traded factors are used, but remain essential with non-traded factors, a setting we consider. While recent work focuses on high-dimensional factor models (Kozak, Nagel, and Santosh (2020), Didisheim, Ke, Kelly, and Malamud (2024)), we instead compare low-dimensional SDF specifications using a large cross-section of test assets. Finally, our asymptotic results relate to the literature on GMM under misspecification (Hall and Inoue (2003) and Hall and Pelletier (2011)).

The remainder of the paper is organized as follows. Section 2 introduces the framework under which we evaluate models and discusses the issues arising from the weighting matrix. Section 3 presents the regularization methods and derives the asymptotic properties of the SDF parameter estimators. Section 4 develops inference procedure for model comparison. Section 5 reports simulation results and Section 6 compares four empirical asset pricing models using a dataset of 647 test portfolios. Finally, Section 7 concludes. Appendices A to C give some extra details. The proofs are collected in the supplementary appendix.

2 Asset pricing model under misspecification

We believe most asset pricing models are misspecified. We want to compare models and determine which one is the less misspecified. To do this, we compare the differences between Hansen-Jagannathan distances.

2.1 Pricing errors and model specification using excess returns

Let r_t be the excess returns of N assets. Given the availability of K factors f_t, the estimation of asset pricing models can be summarized in finding the expression of the relevant stochastic discount factor y_t. The latter is admissible if it correctly prices the test assets r_t [1]:

$$E[r_t.y_t] = 0.$$

Define $Y_t = [f_t', r_t']'$. Its mean and covariance matrix are given by $\mu = E[Y_t] = [\mu_1', \mu_2']'$ and $V = V(Y_t) = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$. We define also the centered variables $\tilde{r}_t = r_t - \mu_2$ and $\tilde{f}_t = f_t - \mu_1$. In this paper, we focus on linear candidate SDF, $y_t(\theta) = 1 - \tilde{f}_t'\theta$. It is common to choose θ by minimizing the aggregate pricing errors $\alpha(\theta) = E[r_t.y_t(\theta)] = \mu_2 - V_{21}\theta$ via

$$\delta_W^2 = \alpha(\theta)' W \alpha(\theta), \tag{1}$$

where W is a positive-definite matrix.

[1]Note that the no-arbitrage constraint, $y_t \geq 0$, is not imposed. Gospodinov, Kan, and Robotti (2016) analyze the benefits and costs of using the HJ-distance with a no-arbitrage constraint and found that preventing the SDF from taking negative values may lower its pricing ability and may be problematic when comparing competing models. Therefore, we focus our attention to unconstrained HJ-distance.

The candidate SDF prices correctly the returns, when one can find θ such that $\delta_W^2(\theta) = 0$. Otherwise, the model is considered globally misspecified. The reason for demeaning the factors is to insure the invariance to affine transformations of the factors, as explained in Kan and Robotti (2008).

When models are misspecified, the SDF parameter, that minimizes (1), depends on the weighting matrix. Therefore, its choice is paramount.

One possibility is to use the GMM framework. In this case, W is the inverse of the asymptotic variance of $\hat{E}[r_t.y_t(\theta)] = \sum_t r_t y_t(\theta)/T$ and the objective function (1) equates the over-identification test of Hansen (1996). However, using this weighting matrix to compare asset pricing models may be misleading for several reasons. First, it has been shown that this diagnostic is model-dependent and tends to reward models with volatile SDF and pricing errors (Ludvigson (2013)). Second, it tends to produce huge leverage portfolios as W is near singular with many assets (Cochrane (1996)).

Other matrices can be used. For example, the inverse of $V_{22} - V_{21}V_{11}^{-1}V_{12}$, the covariance matrix of the residuals from the regression of r on f, is used in Shanken (1985) and Shanken and Zhou (2007) to estimate the risk premium λ. One can also use the identity matrix to circumvent the invertibility issue. Nonetheless, the models estimated will depend on the assets included. This setting is not preferable for researchers looking for results independent of particular dataset.

As established in Kan and Robotti (2008), the use of V_{22}^{-1} as weighting matrix enables the HJ-distance to be model-independent and suitable for asset pricing model comparison. We adopt this approach.

When $W = V_{22}^{-1}$, δ_W^2 is a modified Hansen and Jagannathan (1997) distance, where the mean of the SDF is constrained to 1. Let

$$\delta_{V_{22}}^2 = \delta^2 = (\mu_2 - V_{21}\theta)' V_{22}^{-1}(\mu_2 - V_{21}\theta). \tag{2}$$

We define θ_{HJ} as the solution to the minimization of (2).

$$\theta_{HJ} = \arg\min_{\theta} \delta^2 = (V_{12}V_{22}^{-1}V_{21})^{-1}V_{12}V_{22}^{-1}\mu_2.$$

θ_{HJ} is a pseudo-true value, which coincides with the true θ when the model is correctly specified. It can also be written as $V_{11}^{-1}(\beta' V_{22}^{-1}\beta)^{-1}\beta' V_{22}^{-1}\mu_2 = V_{11}^{-1}\lambda$ where $\beta = V_{21}V_{11}^{-1}$ is the exposure of the returns to the factors f_t and $\lambda = (\beta' V_{22}^{-1}\beta)^{-1}\beta' V_{22}^{-1}\mu_2$ represents the pseudo risk premium[2]. This particular form shows that the SDF parameter can also be estimated via the βs. Such representation is not new as a well-known equivalence between SDF representation, beta-representation and minimum-variance efficiency has already been established (see (Cochrane, 2005, p. 261), chapter 7 of Ferson (2019) or Goyal (2012)). In this setting, the asset pricing model is misspecified when $\alpha = \mu_2 - V_{21}V_{11}^{-1}\lambda = \mu_2 - \beta\lambda \neq 0$.

[2]The pseudo risk premium coincides with the risk premium when $\alpha = 0$.

We introduce the model in the following assumption.

Assumption 1. *(i)*

$$r_t = \alpha + \beta(\tilde{f}_t + \lambda) + \epsilon_t, \tag{3}$$

where β is a matrix $N \times K$, $\alpha \in \mathbb{R}^N$, $\lambda \in \mathbb{R}^K$, the $N \times 1$ error terms ϵ_t are assumed uncorrelated with the factors. In addition, the errors have mean 0 and variance $V(\epsilon_t \mid f_t) = \Sigma_\epsilon = [\sigma_{i,j}]_{i,j=1,\cdots,N}$ where $\sigma_{i,j} = E[\epsilon_{it}\epsilon_{jt}]$. Moreover, $tr(\Sigma_\epsilon)/N = O(1)$ and the maximum eigenvalue of Σ_ϵ is $o(N)$.

(ii) $\frac{1}{N}\sum_{i=1}^{N}\beta_i'\beta_i \to \Sigma_\beta$, *as $N \to \infty$, where Σ_β is positive-definite matrix.*

(iii) V_{11} *and* V_{22} *are non singular.*

Assumption 1(i) relaxes the approximate factor model of Chamberlain and Rothschild (1983). Indeed, Chamberlain and Rothschild (1983) assume that the eigenvalues of Σ_ϵ are uniformly bounded. The condition $tr(\Sigma_\epsilon)/N = O(1)$ is weaker and allows for the largest eigenvalue to diverge (albeit at a slower rate than N). Diverging largest eigenvalue appears in Gagliardini, Ossola, and Scaillet (2016) and Raponi, Robotti, and Zaffaroni (2020).

Assumption 1(ii) is the same as Assumption 2 of Raponi, Robotti, and Zaffaroni (2020). Positive-definite Σ_β excludes spurious factors and cross sectionally constant β_i. Also, this assumption implies that $\beta_k'\beta_k/N < \infty$, $k = 1, \cdots, K$. The factors included in the SDF need to be sufficiently correlated with the returns. However, there could be some weak factors hidden in the error term. This is consistent with what is discussed in Pesaran and Smith (2023): "The idiosyncratic errors, although uncorrelated with $\tilde{f}_t$, could be cross-sectionally correlated due to non-fundamental common factors uncorrelated with the SDF that arise from herding behavior or correlated beliefs for instance at times of crisis." Several authors have warned that using weak factors could yield incorrect econometric inference, see Kleibergen and Zhan (2020) and Manresa, Panaranda, and Sentana (2023) and references therein. Pesaran and Smith (2021) argue that only strong factors should be used to predict returns. So, we stick to the situations where the factors $\tilde{f}_t$ are all strong even though here the aim is not to estimate β but to compare different asset pricing models.

Assumption 1(iii) implies that, in population, none of the factors and none of the assets are redundant. However, in finite sample, $\widehat{V}_{22}$ may be close to singular when N approaches T and will be necessarily singular when $N > T$.

Note that the error term ϵ_t may be serially correlated and the intercept α_i may vary with the asset i.

2.2 Overview of the approach and main results

The covariance matrix of the returns, V_{22}, is often nearly singular when the number of assets N is large, due to the strong cross-sectional dependence induced by common factors. This nearsingularity deteriorates the finite-sample properties of both the SDF estimator and the associated misspecification tests. More formally, under a factor structure, the largest eigenvalue of V_{22} is typically of order N, while the smallest eigenvalues remain bounded.

For this reason, it is convenient to work with the normalized covariance matrix $\Sigma = V_{22}/N$, and to express the SDF parameter as

$$\theta_{HJ} = V_{11}^{-1}(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}\mu_2$$

While the matrix Σ has bounded largest eigenvalue, its smallest eigenvalues accumulate toward zero as $N \to \infty$. As a result, its condition number $\kappa = \lambda_1/\lambda_N$ diverges, making inversion unstable.

To illustrate this instability, consider the equation

$$\Sigma\varphi = \beta \tag{4}$$

where $\beta \in \mathbb{R}^N$ is given. Let $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N > 0$ denote the eigenvalues of Σ with corresponding orthonormal eigenvectors ϕ_j. The solution of (4) is

$$\varphi = \Sigma^{-1}\beta = \sum_{j=1}^{N} \frac{1}{\lambda_j}\left(\phi_j'\beta\right)\phi_j.$$

Now perturb β by $\tilde{\beta} = \beta + c\phi_N$, where c is scalar. The solution becomes $\tilde{\varphi} = \Sigma^{-1}\tilde{\beta} = \varphi + \dfrac{c}{\lambda_N}\phi_N$. Hence,

$$\|\tilde{\varphi} - \varphi\| / \left\|\tilde{\beta} - \beta\right\| = 1/\lambda_N = \kappa/\lambda_1$$

which diverges when N goes to infinity. Small perturbations in β are therefore greatly amplified along directions associated with small eigenvalues. This instability arises even when Σ is known, and becomes more severe when it is estimated. It explains why standard HJ-based inference becomes unreliable when the number of assets is large.

To restore stability, we introduce Tikhonov regularization. The regularized solution to (4) is

$$\varphi^{\gamma} = \left(\Sigma^2 + \gamma I\right)^{-1}\Sigma\beta = \sum_{j=1}^{N} \frac{\lambda_j}{\lambda_j^2 + \gamma}\left(\phi_j'\beta\right)\phi_j$$

where $\gamma > 0$ is a regularization parameter. Unlike the Moore-Penrose inverse (obtained as γ goes to zero), this mapping is continuous. Indeed for a perturbation $\Delta\beta$,

$$\|\tilde{\varphi}^{\gamma} - \varphi\| = \left\|\sum_{j=1}^{N} \frac{\lambda_j}{\lambda_j^2 + \gamma}\left(\phi_j'\Delta\beta\right)\phi_j\right\| \leq \sup_j \frac{\lambda_j}{\lambda_j^2 + \gamma}\|\Delta\beta\| \leq \frac{1}{2\sqrt{\gamma}}\|\Delta\beta\|.$$

Continuity is necessary to establish the convergence of our estimators. Another interesting property of Tikhonov is that the eigenvectors associated with the 0 eigenvalue vanish from the sum.

Building on results from the inverse problem literature (Kress (2014), Carrasco, Florens, and Renault (2007)), we show that $\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\hat{\beta}/N$ and $\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\hat{\mu}_2/N$ are consistent, where $\hat{\Sigma}_\gamma^{-1}$ is the regularized inverse of Σ and γ goes to zero at an appropriate rate. It follows that the regularized estimator $\hat{\theta}_{HJ}^\gamma$ is consistent for the pseudo-true value θ_{HJ} and $\sqrt{T}\left(\hat{\theta}_{HJ}^\gamma - \theta_{HJ}\right)$ is asymptotically normal. This result is used to test whether some factor is priced by testing the nullity of the corresponding SDF coefficient. In practice, implementation requires the estimation of variance terms, which we achieve using regularized HAC estimators. We establish their consistency in this high-dimensional framework.

We next discuss the behavior of the regularized Hansen–Jagannathan distance $\hat{\delta}_\gamma^2$. It has been shown (see Chernov, Kelly, Malamud, and Schwab (2025)) that $T\hat{\delta}_\gamma^2$ tends to diverge as N increases, requiring centering and rescaling for valid inference. However, our objective is not to test whether a model is correctly specified ($\delta^2 = 0$), but rather to compare competing misspecified models.

A key insight of this paper is that differences of regularized HJ distances remain well behaved even when each distance may diverge individually. Let a and b be two asset pricing models with regularized HJ distances $\hat{\delta}_{a,\gamma}^2$, $\hat{\delta}_{b,\gamma}^2$.

When the models are nested or when they are not nested but share the same SDF, we show that $T\left(\hat{\delta}_{a,\gamma}^2 - \hat{\delta}_{b,\gamma}^2\right)$ can be written as a quadratic function of the parameters $\hat{\theta}_{HJ}^\gamma$. Since $\hat{\theta}_{HJ}^\gamma$ is asymptotically normal, this statistic converges to a weighted sum of chi-square distributions with a finite number of terms (bounded by the number of factors).

When the models are not nested and have different SDFs, the convergence rate is slower. In this case, $\sqrt{T}\left(\hat{\delta}_\gamma^2 - \delta^2\right)$ is asymptotically normal and no bias term arises because we operate under global misspecification ($\delta^2 \neq 0$). As a result, $\sqrt{T}\left(\hat{\delta}_{a,\gamma}^2 - \hat{\delta}_{b,\gamma}^2\right)$ is also asymptotically normal.

Overall, regularization plays two roles. First, it stabilizes the estimation of the SDF parameters by addressing the ill-posed nature of the inverse problem. Second, it enables meaningful comparison of asset pricing models in high-dimensional settings by ensuring that differences in HJ distances remain well behaved.

3 Regularized SDF parameter estimator

As stated earlier, inference using the modified HJ-distance may not be robust to a large number of correlated securities that makes the weighting matrix near singular or singular. Relying on the literature on inverse problems in an infinite dimensional space (see Kress (2014) and Carrasco, Florens, and Renault (2007)), we introduce two regularization methods to stabilize the weighting matrix and improve the estimation of asset pricing models.

3.1 Regularization as penalization

The following lemma sheds light on the properties of the matrix $\Sigma = V_{22}/n$.

Lemma 1. *Under Assumptions 1, assuming* (ϵ_t, f_t) *is stationary and* $e'e/N < \infty$*, we have the following results as* $N \to \infty$*,*

1. $tr(\Sigma) = O(1)$.
2. $E[r_t' r_t/N] = O(1)$.

Lemma 1 shows that the normalized covariance matrix Σ has bounded trace when N increases. As a consequence, Σ behaves like a compact operator in the limit, in the sense that its eigenvalues are bounded but accumulate at zero. This implies that the inverse of Σ, when it exists, is not a continuous operator and the associated inverse problem is ill-posed. In particular, small perturbations in the data can lead to large variations in Σ^{-1}, making estimation unstable. Regularization therefore becomes necessary to stabilize the inversion and obtain well-defined estimators. For this reason, it is more natural to regularize Σ rather than V_{22}, as this allows us to directly exploit results from the inverse problem literature.

Let $\hat{\Sigma}$ be the sample counterpart of Σ. Let $\gamma > 0$ be a regularization parameter. We consider two techniques which consist in replacing the singular or nearly singular matrix $\hat{\Sigma}$ by a well-conditioned matrix before inverting the matrix. These two regularization schemes give the following inverses:

Ridge regularization

$$\hat{\Sigma}_\gamma^{-1} = (\hat{\Sigma} + \gamma I_N)^{-1}.$$

Tikhonov regularization

$$\hat{\Sigma}_\gamma^{-1} = (\hat{\Sigma}^2 + \gamma I_N)^{-1}\hat{\Sigma}.$$

Appendix A gives some further details on these regularizations. The regularization parameter γ plays a crucial role in stabilizing the inverse of the matrix Σ. For small γ, the regularized inverse will be close to the actual inverse while being much more stable. In practice, the tuning parameter γ is chosen to go to zero with the sample size. Its choice is discussed later.

We focus on ridge and Tikhonov regularizations for their favorable properties and clear interpretation. While alternative approaches exist—such as nonlinear shrinkage (Ledoit and Wolf (2017)), principal components, Landweber–Fridman regularization (Carrasco, Kone, and Noumon (2019)), thresholding (Fan, Liao, and Mincheva (2013)), graphical lasso (Callot, Caner, Onder, and Ulasan (2021)), or L_2-boosting (Lönn and Schotman (2024))—many rely on sparsity assumptions on Σ or Σ^{-1}, which are unlikely to hold in asset returns. In contrast, ridge and Tikhonov regularizations are well suited to settings with strong cross-sectional dependence, as in our application.

Below, we provide an interpretation of the regularized HJ-distance in terms of a penalization on the norm of the pricing error.

Let L^2 be the set of random variables X such that $E\left(X^2\right) < \infty$. Suppose the candidate SDF y_t belongs to L^2. First, recall that, when N is fixed, Kan and Robotti (2008) showed that δ^2 (not regularized) represents the distance between the proposed SDF y_t and the set of admissible SDFs of mean 1 in L^2.

$$\delta^2 = \min_{m_t \in L^2, E[m_t]=1} E\left(m_t - y_t\right)^2 \quad \text{subject to } E[m_t r_t] = 0. \tag{5}$$

How does this result change when one introduces regularization? Let

$$\begin{aligned} \delta_R^2 &= \alpha'(\Sigma + \gamma I_N)^{-1}\alpha/N, \\ \delta_K^2 &= \alpha'(\Sigma^2 + \gamma I_N)^{-1}\Sigma\alpha/N \end{aligned}$$

be respectively the HJ-distance obtained by Ridge and Tikhonov regularizations. Below, we are going to show that these regularized HJ-distances measure how far y is to the closest valid SDF of mean 1 which prices returns with an error controlled by γ. We use the notation $\|v\|_N^2 = v'v/N$.

Proposition 1. *The following results hold:*

1. For ridge,

$$\delta_R^2 = \inf_{m \in L^2, E[m]=1} E[(m-y)^2] + \frac{1}{\gamma} \parallel E[mr] \parallel_N^2, \tag{6}$$

2. For Tikhonov,

$$\delta_K^2 = \mathop{inf}_{m \in L^2, E[m]=1} E[(m-y)^2] + \frac{1}{\gamma} \parallel E[mr] \parallel_{N,\Sigma}^2, \tag{7}$$

where $\parallel x \parallel_{N,\Sigma}^2 = \dfrac{x'\Sigma x}{N}$ *for any* $x \in \mathbb{R}^N$.

The previous proposition shows that regularization is equivalent to relaxing the constraint of problem (5). Low values of γ put the emphasis on the asset pricing equation, while high values allow for possible errors in the pricing of assets. Note that the optimal m from (6) also solves

$$\inf_{m \in L^2, E[m]=1} E[(m-y)^2] \text{ s.t. } \parallel E[mr] \parallel_N^2 \leq \tau$$

for an appropriately chosen τ related to γ. Similarly, the optimal m from (7) solves

$$\mathop{inf}_{m \in L^2, E[m]=1} E[(m-y)^2] \text{ s.t. } \parallel E[mr] \parallel_{N,\Sigma}^2 \leq \tau$$

for an appropriately chosen τ related to γ.

In Appendix B, we also provide an interpretation of regularization as a penalization on the Lagrange multipliers. Note that when $N \to \infty$, duality breaks down if no regularization is applied, see Borwein (1993).

Now we examine how Tikhonov regularization affects the constraint. Since directions associated with small eigenvalues receive disproportionately large weights in the HJ distance, regularization effectively rebalances these contributions.

Let $P^{'}\Lambda P$ be the spectral decomposition of Σ. Then the penalization term can be written as

$$\frac{1}{\gamma} \parallel E[mr] \parallel^2_{N,\Sigma} = \frac{1}{\gamma N} E[mr]^{'} P^{'} \Lambda P E[mr] = \sum_{j=1}^{N} \omega_j E[m(Pr)_j]^2,$$

where $\omega_j = \dfrac{\lambda_j}{\gamma N}$ and $(Pr)_j$ denotes the jth principal component of returns. This representation shows that regularization entails the repackaging of the assets into N portfolios $(Pr)_j$ where directions associated with small eigenvalues receive less weight.

Korsaye, Quaini, and Trojani (2025) provide a related interpretation in terms of transaction costs. In their framework, pricing errors are penalized for "dubious" assets, leading to a modified SDF that tolerates deviations from exact pricing. In our setting, ridge and Tikhonov regularizations induce a quadratic penalization, which can be interpreted as transaction costs. Unlike their approach, however, we do not impose positivity of the SDF but allow for large N and large T asymptotics.

3.2 Asymptotic distribution of the regularized SDF parameter of misspecified models

Let $R = [r_1, \cdots, r_T]^{'}$ and $F = [f_1, \cdots, f_T]^{'}$ be respectively the $T \times N$ and $T \times K$ matrices of returns and factors. The OLS estimates of β is given by

$$\hat{\beta} = (\bar{R}^{'}\bar{F})(\bar{F}^{'}\bar{F})^{-1} = \hat{V}_{21}\hat{V}_{11}^{-1}$$

where $\bar{R} = R - 1_T\hat{\mu}_2^{'}$ and $\bar{F} = F - 1_T\hat{\mu}_1^{'}$. $\bar{R} = \begin{bmatrix} \bar{r}_1^{'} \\ \vdots \\ \bar{r}_T^{'} \end{bmatrix}$ and $\bar{F} = \begin{bmatrix} \bar{f}_1^{'} \\ \vdots \\ \bar{f}_T^{'} \end{bmatrix}$ with $\bar{r}_t = r_t - \hat{\mu}_2$ and $\bar{f}_t = f_t - \hat{\mu}_1$. $\hat{\mu}_1 = \dfrac{1}{T}\sum_{t=1}^{T} f_t$ and $\hat{\mu}_2 = \dfrac{1}{T}\sum_{t=1}^{T} r_t$ are respectively the estimators of μ_1 and μ_2.

For any regularization scheme, the estimator of the pseudo-true value θ_{HJ} is given by

$$\hat{\theta}_{HJ}^{\gamma} = \arg\min_{\theta} (\hat{\mu}_2 - \hat{V}_{21}\theta)^{'}\hat{\Sigma}_{\gamma}^{-1}(\hat{\mu}_2 - \hat{V}_{21}\theta). \tag{8}$$

$$\hat{\theta}_{HJ}^{\gamma} = \hat{V}_{11}^{-1}(\hat{\beta}^{'}\hat{\Sigma}_{\gamma}^{-1}\hat{\beta})^{-1}\hat{\beta}^{'}\hat{\Sigma}_{\gamma}^{-1}\hat{\mu}_2$$

and the regularized HJ-distance is

$$
\begin{aligned}
\hat{\delta}^2_\gamma &= \frac{\hat{\mu}_2'\hat{\Sigma}_\gamma^{-1}\hat{\mu}_2}{N} - \frac{\hat{\mu}_2'\hat{\Sigma}_\gamma^{-1}\hat{V}_{21}}{N}\left(\frac{\hat{V}_{12}\hat{\Sigma}_\gamma^{-1}\hat{V}_{21}}{N}\right)^{-1}\frac{\hat{V}_{12}\hat{\Sigma}_\gamma^{-1}\hat{\mu}_2}{N} \\
&= \frac{\hat{\mu}_2'\hat{\Sigma}_\gamma^{-1}\hat{\mu}_2}{N} - \frac{\hat{\mu}_2'\hat{\Sigma}_\gamma^{-1}\hat{\beta}}{N}\left(\frac{\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\hat{\beta}}{N}\right)^{-1}\frac{\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\hat{\mu}_2}{N}
\end{aligned}
$$

where $\hat{\Sigma}_\gamma^{-1}$ is the regularized inverse of $\hat{\Sigma}$ obtained either by Ridge or Tikhonov regularization.

Using the definition of the asset pricing model, the average of the excess return can be rewritten as

$$\hat{\mu}_2 = \hat{\beta}(\hat{\mu}_1 - \mu_1 + \lambda) + (\beta - \hat{\beta})(\hat{\mu}_1 - \mu_1 + \lambda) + \alpha + \bar{\epsilon},$$

where $\bar{\epsilon} = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t$. Remark that $\mu_2 = \alpha + \beta\lambda$ and $\beta' V_{22}^{-1}\alpha = 0$ by the first order condition of (2). Then, $\hat{\theta}^\gamma_{HJ} - \theta_{HJ}$ can be decomposed as

$$
\begin{aligned}
\hat{\theta}^\gamma_{HJ} - \theta_{HJ} = {} & (\hat{V}_{11}^{-1} - V_{11}^{-1})\lambda + \hat{V}_{11}^{-1}(\hat{\mu}_1 - \mu_1) \\
& + \hat{V}_{11}^{-1}(\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\hat{\beta})^{-1}\Big[\hat{\beta}'\hat{\Sigma}_\gamma^{-1}(\beta - \hat{\beta})(\lambda + \hat{\mu}_1 - \mu_1) \\
& + \hat{\beta}'\hat{\Sigma}_\gamma^{-1}\alpha + \hat{\beta}'\hat{\Sigma}_\gamma^{-1}\bar{\epsilon}\Big].
\end{aligned}
\tag{9}
$$

Appendix A shows that Ridge and Tikhonov regularizations are closely related. For this reason, we focus on Tikhonov regularization. From now on, $\hat{\theta}^\gamma_{HJ}$ and $\hat{\delta}^2_\gamma$ correspond to the estimators obtained by Tikhonov regularization.

Assumption 2. *(i) The process $x_t = (\epsilon_{it}, f_{kt})_{t=1,2,\cdots,T}$ is stationary and strong mixing with mixing coefficients $\alpha_x(l)$ verifying*

$$\sum_{l=1}^{\infty} \alpha_x(l)^{\frac{\rho}{2+\rho}} < \infty,$$

for some $\rho > 0$. $\alpha_x(l) = \sup_{i,k \geq 1}\sup_{A,B}\left[|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}^0_{-\infty}, B \in \mathcal{F}^\infty_l\right]$, where $l \geq 1$ and $\mathcal{F}^v_u = \sigma(x_t : u \leq t \leq v)$ is the σ-field generated by the data from time u to v.

(ii) $E[\epsilon_{it}^{4+2\rho}] < c$, for $i = 1, 2, \cdots, N$, where c is a constant and ρ is given in (i).

(iii) $E[\|f_t\|^{4+2\rho}] < \infty$.

Assumption 2(i) specifies the rate of decay for the mixing coefficient in terms of a parameter ρ. When the data are independent, $\alpha_x = 0$ and this condition is automatically satisfied for all

$\rho > 0$. If x_t is exponentially strong mixing, then A2(i) is also valid for any $\rho > 0$. Assumption 2(ii) implies that $E[\epsilon_t'\epsilon_t/N] = O(1)$ as $N \to \infty$.

Definition 1. *The matrix Σ can be viewed as an operator $\Sigma : \mathbb{R}^N \to \mathbb{R}^N$ where $\mathbb{R}^N$ is endowed with norm $\|v\|_N$ and inner product $\langle v_1, v_2 \rangle_N = \dfrac{v_1' v_2}{N}$. We define the range of Σ^ω (for some scalar $\omega > 0$), denoted $\mathcal{R}(\Sigma^\omega)$, as the set $\{\psi \in \mathbb{R}^N : \psi = \Sigma^\omega \phi$ for some $\phi \in \mathbb{R}^N$ such that $\|\phi\|_N < \infty\}$.*

As the eigenvalues λ_j (associated with the eigenvectors ϕ_j) of Σ decline to zero as N goes to infinity, the elements of $\mathcal{R}(\Sigma^\omega)$ are such that the inner products $\langle \psi, \phi_j \rangle_N$ decline to zero faster than λ_j^ω. The sequence $(\mathcal{R}(\Sigma^\omega))_{\omega > 0}$ is a decreasing family of subspaces of $\mathbb{R}^N$ as ω increases. The parameter ω permits to quantify the regularization bias and determine conditions on γ.

We introduce the following assumption on the β_k and α.

Assumption 3. *(i) $\beta_k, \alpha \in \mathcal{R}(\Sigma^2)$, for $k = 1, 2, ..., K$.*

(ii) As $N \to \infty$, $C_\beta = \dfrac{1}{N}\beta' \Sigma^{-1} \beta = \left\langle \Sigma^{-\frac{1}{2}}\beta, \Sigma^{-\frac{1}{2}}\beta \right\rangle_N \to C$, where C is positive-definite matrix.

Assumption 3(i) states that β_k and α belong to the range of Σ^ω with $\omega = 2$ so that objects $\Sigma^{-2}\beta_k$ and $\Sigma^{-2}\alpha$ are well defined even when N goes to infinity. This condition rules out concentration of β_k and α on directions associated with very small eigenvalues and helps control the regularization bias. It ensures that quadratic forms such as $\beta'\Sigma^{-1}\beta$, $e'\Sigma^{-1}e$ and hence δ^2 remain bounded when $N \to \infty$. In Appendix A, we show that $\hat{\beta}$ belongs to the range of $\hat{\Sigma}^{1/2}$.

The following assumption is needed to derive the distribution of the regularized SDF parameter when N and T go to ∞.

Assumption 4. *For $\rho > 0$ defined in Assumption 2 (i), we assume:*

(i) $\lim\limits_{N \to \infty} E[\|r_t\|_N^{4+2\rho}] < \infty$.

(ii) $0 < \lim\limits_{N,T \to \infty} Var\left(\dfrac{1}{\sqrt{T}} \sum\limits_{t=1}^{T} \langle r_t, \Sigma^{-1}\alpha \rangle_N\right) < \infty$.

(iii) $0 < \lim\limits_{N,T \to \infty} Var\left(\dfrac{1}{\sqrt{T}} \sum\limits_{t=1}^{T} \langle \Sigma^{-1}\beta, r_t \rangle_N\right) < \infty$.

(iv) $0 < \lim\limits_{N,T \to \infty} Var\left(\dfrac{1}{\sqrt{T}} \sum\limits_{t=1}^{T} \langle \epsilon_t, \Sigma^{-1}\alpha \rangle_N\right) < \infty$.

Proposition 2. *Suppose Assumptions 1-4 are satisfied.*

As T, N go to infinity and γ goes to zero, if γ is chosen such that $\gamma T \to \infty$ and $\gamma^2 T \to 0$, we have the following results for Tikhonov regularization

1. $\hat{\theta}^{\gamma}_{HJ} \xrightarrow{P} \theta_{HJ}$
2. $\sqrt{T}(\hat{\theta}^{\gamma}_{HJ} - \theta_{HJ}) \xrightarrow{d} \mathcal{N}(0_K, V_{11}^{-1} S V_{11}^{-1})$

where $S = \underset{N,T\to\infty}{lim}\ var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_t\right]$ *and* h_t *is defined as*

$$h_t = \widetilde{f}_t y_t + \lambda + C_\beta^{-1}\frac{\beta'\Sigma^{-1}}{N}\left(\epsilon_t y_t - \tilde{r}_t\tilde{u}_t\right) + C_\beta^{-1}V_{11}^{-1}\widetilde{f}_t\frac{\epsilon_t'\Sigma^{-1}\alpha}{N} \text{ with } \tilde{u}_t = \frac{\tilde{r}_t'\Sigma^{-1}\alpha}{N}.$$

Proposition 2 holds for both correctly specified (case where $\alpha = 0$) and misspecified models.

Note that there is no condition connecting N and T so that N can be smaller or larger than T. Even though N does not explicitly appear in the proposition, N is present in the weighting matrix $\Sigma = V_{22}/N$, so that γ depends on N.

Let us discuss the impact of regularization on the properties of the estimator. It is well-known that when the number of moments increase, the variance of the standard GMM estimator decreases but its bias increases (Newey and Smith (2004)) and regularization permits to mitigate the bias at the cost of a higher variance (Carrasco (2012)). Here, we are not in the standard setting because of the misspecification and the fact that the weighting matrix is not the optimal one. A formal investigation of the bias and variance as a function of the number of test assets would require a higher-order expansion and is beyond the scope of this paper. The primary role of the regularization is to stabilize the inverse of Σ when the number of test assets is large, hence reducing the variance of the estimator. However, the regularization introduces a regularization bias which vanishes asymptotically when $\gamma\sqrt{T} \to 0$ (see term (33) in the proof). Proposition 2 shows that $\hat{\theta}^{\gamma}_{HJ}$ is a consistent estimator of the pseudo-true value provided the regularization parameter γ converges to zero at an appropriate rate as $T \to \infty$. This rate is satisfied if $\gamma \sim \frac{1}{T^{\kappa}}$, with $\kappa \in (1/2, 1)$. In practice, we select γ using a data-driven procedure.

The results of Proposition 2 provide the basis for comparing competing asset pricing models. They can also be used to test whether a given factor is priced via a $t-$test on individual components of $\hat{\theta}^{\gamma}_{HJ}$. Indeed, if r_t is not correlated with a factor, say f_j, then the jth row of V_{12} equals zero and the j^{th} element of θ_{HJ} is zero. So even if the model is misspecified in the sense that $\alpha \neq 0$, the jth component of $\hat{\theta}^{\gamma}_{HJ}$ consistently converge to 0 when the jth factor is not priced. To implement the test, we estimate the standard error by replacing Σ^{-1} in h_t by the regularized inverse $\hat{\Sigma}_{\gamma}^{-1}$. So regularization is also useful to get a reliable standard error estimate. If h_t is autocorrelated, we need to rely on a HAC estimator.

3.3 HAC estimator

The long-run variance S can be estimated using the HAC estimator of Andrews (1991) as explained below.

Let $\widehat{h}_t$ be an estimator of h_t:

$$\widehat{h}_t = \bar{f}_t \hat{y}_t + \hat{V}_{11}\hat{\theta} + \widehat{C}_\beta^{-1} \frac{\hat{\beta}' \hat{\Sigma}_\gamma^{-1}}{N} \left(\hat{\epsilon}_t \hat{y}_t - \overline{r}_t \widehat{u}_t \right) + \widehat{C}_\beta^{-1} \widehat{V}_{11}^{-1} \overline{f}_t \frac{\widehat{\epsilon}_t' \hat{\Sigma}_\gamma^{-1} \widehat{\alpha}}{N}$$

with $\hat{y}_t = 1 - \bar{f}_t \hat{\theta}$, $\hat{\epsilon}_t = \overline{r}_t - \hat{\beta} \bar{f}_t$, $\widehat{\alpha} = \hat{\mu}_2 - \widehat{V}_{21}\hat{\theta}$, $\widehat{u}_t = \dfrac{\overline{r}_t' \hat{\Sigma}_\gamma^{-1} \widehat{\alpha}}{N}$, and $\hat{\Sigma}_\gamma^{-1}$ is obtained by Tikhonov regularization.

For a kernel ω and a bandwidth m_T, the HAC estimator of S is defined as

$$\hat{S}_T = \sum_{|j|<T} \omega \left(\frac{j}{m_T} \right) \widehat{\Gamma}_j$$

where $\widehat{\Gamma}_j = \dfrac{1}{T} \displaystyle\sum_{t=j+1}^{T} \widehat{h}_t \widehat{h}_{t-j}'$ for $j \geq 0$ and $\widehat{\Gamma}_j = \widehat{\Gamma}_{-j}'$ for $j < 0$. The kernel ω is assumed to below to the class $\mathcal{W}$:

$$\begin{aligned} \mathcal{W} \;\; = \;\; & \{ \omega(.) : \mathbb{R} \rightarrow [-1, 1], \omega(0) = 1, \omega(x) = \omega(-x), \forall x \in \mathbb{R}, \\ & \int |\omega(x)| \, dx < \infty, \omega(.) \text{ is continuous at 0 and at all but a finite number of points} \Big\}. \end{aligned}$$

This class corresponds to the class $\mathcal{K}_1$ of Andrews (1991) with the extra condition $\int |\omega(x)| \, dx < \infty$.

Note that h_t depends on two types of parameters, the finite dimensional parameters denoted ξ_1 and the $N \times 1$ dimensional parameters denoted ξ_2. More specifically, $\xi_1 = \left(\theta', \mu_1', vecV_{11}', vecC_\beta' \right)'$, $\xi_2 = \left((\Sigma^{-1}\beta)', \left(\Sigma^{-1}\alpha \right)', \mu_2', \beta' \right)'$. To establish the consistency of $\hat{S}_T$, we need some extra conditions.

Assumption 5. *(i) $\omega \in \mathcal{W}$.*

(ii) $\{h_t\}$ is $\alpha - mixing$ with

$$\sum_{l=1}^{\infty} l^2 \alpha(l)^{\frac{\rho}{2+\rho}} < \infty$$

where ρ is the same as in Assumption 2.

(iii) $E\left[\|h_t\|^{4+2\rho} \right] < \infty$ and $E\left[\sup_{\xi_1 \in \Xi} \left\| \dfrac{\partial h_t}{\partial \xi_1} \right\|^2 \right] < \infty$, where Ξ is a convex neighborhood of the true value of ξ_1.

Proposition 3. *Suppose Assumptions 1-5 hold, T, N go to infinity, γ is chosen such that $\gamma T \rightarrow \infty$ and $\gamma^2 T \rightarrow 0$, and $m_T \rightarrow \infty$ as $T \rightarrow \infty$. Then, we have*

$$\left\| \widehat{S}_T - S \right\| = O_p \left(m_T \left(\frac{1}{\sqrt{\gamma T}} + \sqrt{\gamma} \right) \right).$$

Assumption 5(ii) is slightly stronger than Assumption 2(i). Our conditions are similar to Assumptions A and B in Andrews (1991). The proof of the consistency of $\widehat{S}_T$ is more involved than that of Andrews because of the presence of parameters which dimension N increases. To the best of our knowledge, Proposition 3 is the first result establishing the convergence of HAC estimator based on Tikhonov regularization in high dimension. Some results on the HAC estimator in high dimension have been derived by Li and Liao (2020) for series estimators and Babii, Ghysels, and Striaukas (2024) for Lasso estimators. Proposition 3 provides an upper bound for the convergence rate of $\widehat{S}_T - S$. We observe that this rate depends on the regularization parameter γ and is slower than that obtained by Andrews (1991) which was $m_T/\sqrt{T} \to 0$. This is due to the large dimensional parameters ξ_2 which estimators converge slower than $\sqrt{T}$.

3.4 Choice of the regularization parameter

We adopt a data-driven procedure to select the regularization parameter γ. From Proposition 2, consistency requires that $\gamma T \to \infty$ and $\gamma^2 T \to 0$ which implies that γ must decrease at a rate $T^{-\kappa}$ with $\kappa \in (1/2, 1)$. We impose this rate by setting $\gamma = c\lambda_{\max} T^{-0.8}$ where $\lambda_{\max}$ is the largest eigenvalue of $\left(\widehat{V}_{22}/N\right)^2$ and c is a scalar selected from a grid. This parametrization separates scale (captured by $\lambda_{\max}$) from shrinkage intensity (captured by c). Scaling by $\lambda_{\max}$ ensures that the regularization is invariant to the magnitude of the covariance matrix, while restricting c to a grid allows for a simple one-dimensional search. To select c, we use a two-fold cross-validation procedure designed to approximate the out-of-sample Hansen–Jagannathan distance.

For a given sample size T, we split the data into two equal subsamples. For each candidate c, we estimate θ using the first subsample, evaluate the pricing errors on the second subsample, and compute the out-of-sample criterion:

$$HJ_{oos} = (\bar{\mu}_2^o - \widehat{V}_{21}^o \hat{\theta}_c)' \left(diag\left(\widehat{V}_{22}^o\right)\right)^{-1} (\bar{\mu}_2^o - \widehat{V}_{21}^o \hat{\theta}_c), \tag{10}$$

where quantity with ^{o} are estimated from the withheld sample. We then select the value of c that minimizes the average of the HJ_{oos} obtained on the two different subsamples.

In principle, the Hansen–Jagannathan distance involves weighting by $\left(\widehat{V}_{22}^o\right)^{-1}$. However, in high dimension, $\left(\widehat{V}_{22}^o\right)^{-1}$ is poorly estimated and its inverse is highly unstable. Using a regularized inverse with the same γ inside the cross-validation criterion would mechanically favor large values of γ, since stronger regularization reduces the criterion. To avoid this issue, we replace $\widehat{V}_{22}^{o-1}$ by $\left(diag\left(\widehat{V}_{22}^o\right)\right)^{-1}$, which provides a stable and scale-correct weighting scheme.

4 Tests of equality of HJ-distances of two asset pricing models

The distribution of the difference in HJ distances depends on whether models are nested, share the same SDF, or are distinct. We treat these cases separately because they lead to different convergence rates and limiting distributions.

We compare two competing models (Models a and b) using their regularized HJ-distances. Their SDFs are defined as $y_{at}(\eta) = 1 - (x_{at} - E[x_{at}])'\theta_a$ and $y_{bt}(\lambda) = 1 - (x_{bt} - E[x_{bt}])'\theta_b$. $x_{at} = [f'_{1t}, f'_{2t}]'$ and $x_{bt} = [f'_{1t}, f'_{3t}]'$ are two sets of factors, that are used in Model a and Model b, respectively. f_{it} is of dimension $K_i \times 1$, $i = 1, 2, 3$. $\theta_a = [\theta'_{a1}, \theta'_{a2}]'$and $\theta_b = [\theta'_{b1}, \theta'_{b2}]'$.

The two corresponding pricing models are respectively

$$r_t = \alpha_a + \beta_a(x_{at} - E[x_{at}] + \lambda_a) + \epsilon_{at}, \tag{11}$$

and

$$r_t = \alpha_b + \beta_b(x_{bt} - E[x_{bt}] + \lambda_b) + \epsilon_{bt},$$

with $\theta_a = V_{a,11}^{-1}\lambda_a$ and $\theta_b = V_{b,11}^{-1}\lambda_b$. α_m represents the vector of pricing errors of model $m = a, b$. We note $\delta_m^2, m = a, b$ the HJ-distances of the two models.

$$\delta_m^2 = \mu_2' V_{22}^{-1}\mu_2 - \mu_2' V_{22}^{-1} V_{m,21}(V_{m,12}V_{22}^{-1}V_{m,21})^{-1}V_{m,12}V_{22}^{-1}\mu_2 \tag{12}$$

Model m is estimated using solely factors in x_m. Let $\theta_m = (\theta'_{m1}, \theta'_{m2})'$ for $m = a, b$ where θ_{m1} corresponds to the coefficient of the common factors f_{1t} and θ_{m2} to that of the other factors.

When $K_1 = 0$, the two models do not share factors. When $K_2 = 0$ or $K_3 = 0$, one of the models nests the other one. Finally, when $K_1 > 0$, $K_2 > 0$, and $K_3 > 0$, the two models are non-nested with overlapping factors.

4.1 Comparison of nested models

In this section, we assume without loss of generality that $K_2 = 0$. When the models are nested, the equality of HJ-distances is equivalent to the equality of the SDFs as pointed out by Kan and Robotti (2009). Let $V_{a,12} = E(x_{at}r'_t) = [E(f_{1t}r'_t)\ \ E(f_{2t}r'_t)]$ and $V_{b,12} = E(x_{bt}r'_t) = [E(f_{1t}r'_t)\ \ E(f_{3t}r'_t)]$. Let us define $C_m = \left(V_{m,12}V_{22}^{-1}V_{m,21}\right)^{-1}$, $m = a, b$ and partition it as below

$$C_m = \begin{bmatrix} C_{m,11} & C_{m,12} \\ C_{m,21} & C_{m,22} \end{bmatrix}. \tag{13}$$

We assume $C_{b,22}^{-1}$ is a full rank matrix. It follows from Lemma 11 in appendix that the difference of HJ-distances ($\delta_a^2 - \delta_b^2$) between the two models is equal to

$$\delta_a^2 - \delta_b^2 = \theta_{b2}' C_{b,22}^{-1} \theta_{b2}. \tag{14}$$

The following proposition can be viewed as a generalization of Kan and Robotti (2009) Proposition 2 where N and T are allowed to go to ∞ using regularization.

Proposition 4. *Suppose Assumptions 1-4 are satisfied and $C_{b,22}$ is non-singular. We have the following results:*

1. $\delta_a^2 = \delta_b^2$ if and only if $\theta_{b2} = 0_{K_3}$

2. Under the hypothesis $\theta_{b2} = 0_{K_3}$, as T, N go to infinity and if γ is chosen such that $\gamma T \to \infty$ and $\gamma^2 T \to 0$,

$$T(\hat{\delta}_{a,\gamma}^2 - \hat{\delta}_{b,\gamma}^2) \xrightarrow{d} \sum_{j=1}^{K_3} \xi_j \chi_j^2(1)$$

where $\chi_j^2(1)$ are independent chi-square random variables with 1 degree of freedom and ξ_j are the eigenvalues of $V\left(\widehat{\theta}_{b2}^{\gamma}\right)^{1/2} C_{b,22}^{-1} V\left(\widehat{\theta}_{b2}^{\gamma}\right)^{1/2}$ *and $V\left(\widehat{\theta}_{b2}^{\gamma}\right)$ is the asymptotic variance of $\widehat{\theta}_{b2}^{\gamma}$.*

Proposition 4 implies that we can perform two kinds of tests to compare Model a with factor f_1 and Model b with factors f_1 and f_3.

On the one hand, we can focus on the SDF parameter θ_{b2} (coefficient of f_3) and test $H_0 : \theta_{b2} = 0_{K_3}$ using a t- or Wald statistics based on the misspecification-robust asymptotic distribution given by Proposition 2 in a framework where returns are governed by (11).

On the other hand, we can compute the HJ-distance difference of the two models $T(\hat{\delta}_{a,\gamma}^2 - \hat{\delta}_{b,\gamma}^2)$ using the same level of penalization or (14) and use the asymptotic distribution given in point 2 of Proposition 4 to compare them. Note that the coefficients ξ_j are all nonnegative, hence the test presented in Proposition 4 is a one-sided test.

When comparing models, it is essential to use the same penalization value γ across models. A small value of penalization provides maximum power without compromising size, while a larger value diminishes it. This comes from the fact that large values of γ shrinks the regularized HJ-distances toward 0.

4.2 Comparison of non-nested models

In this section, we assume $K_1 > 0$, $K_2 > 0$, and $K_3 > 0$. The two models are non-nested with overlapping factors. In this case, equality of HJ-distance can be achieved in two cases. The first case corresponds to the setting where the SDFs coincide. The second is when $y_a \neq y_b$ but $\delta_a^2 = \delta_b^2$. Both cases need to be treated separately.

4.2.1 Test of SDFs equality

In this section, we test the equality of the SDFs, $y_a = y_b$. Given the models are non-nested, the equality of SDFs can be achieved only if the both SDF depend on f_1 only. Consider $C_a = \left(V_{a,12}V_{22}^{-1}V_{a,21}\right)^{-1}$, partition it as in (13), and assume $C_{a,22}$ is a full rank matrix. Lemma 11 shows that the difference between the HJ-distances is

$$\delta_a^2 - \delta_b^2 = -\theta_{a2}^{'}C_{a,22}^{-1}\theta_{a2} + \theta_{b2}^{'}C_{b,22}^{-1}\theta_{b2}.$$

The following proposition outlines the asymptotic distribution of the test statistics.

Proposition 5. *Suppose Assumptions 1-4 are satisfied. Moreover, assume that $C_{a,22}$ and $C_{b,22}$ are non-singular. We have the following result:*

1. $y_a = y_b$ if and only if $\theta_{a2} = 0_{K_2}$ and $\theta_{b2} = 0_{K_3}$.

2. Under the hypothesis $y_a = y_b$, as T, N go to infinity and γ goes to zero such that $\gamma T \to \infty$ and $\gamma^2 T \to 0$,

$$T(\hat{\delta}_{a,\gamma}^2 - \hat{\delta}_{b,\gamma}^2) \xrightarrow{d} \sum_{i=1}^{K_2+K_3} \xi_i \chi_i^2(1) \tag{15}$$

where ξ_i are the eigenvalues of $\begin{bmatrix} -C_{a,22}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & C_{b,22}^{-1} \end{bmatrix} V \left(\begin{bmatrix} \hat{\theta}_{a2} \\ \hat{\theta}_{b2} \end{bmatrix}\right)$, and $\chi_i^2(1)$ are independent $\chi^2(1)$ random variables.

Note that the ξ_i in Equation (15) may be positive or negative, hence this test is a two-sided test. Proposition 5.1. shows that to compare asset pricing models with overlapping factors, one can also test the simultaneous nullity of the coefficients of the non common factors (θ_{a2} and θ_{b2}). Then, Proposition 2 can be used to construct a classic Wald test. This option does not directly test the nullity of the difference in HJ-distances, but the equality of the SDFs of the two models.

4.2.2 Comparison of non-nested models with distinct SDFs

In this section, we wish to compare two non-nested models with distinct SDFs ($y_a \neq y_b$) when both models are misspecified. Hansen, Heaton, and Luttmer (1995) and Kan and Robotti (2008) have already given the distribution of the HJ-distance (for gross returns) and the modified HJ-distance (for excess returns) respectively when models are misspecified and N is fixed.

Kan and Robotti (2008) show that the modified HJ-distance has the following distribution

$$\sqrt{T}(\hat{\delta}^2 - \delta^2) \xrightarrow{d} \mathcal{N}(0, v_1)$$

where $v_1 = var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} q_t^m\right)$ and $q_t^m = y_t^2 - (y_t - \nu^{'}(r_t - \mu_2))^2 + 2\nu^{'}\mu_2 - \delta^2$ with ν is the Lagrange multiplier ($\nu = V_{22}^{-1}E[r_t y_t]$) of problem (5). It is worth noticing that the distribution of the distance is not affected by the uncertainty brought forth by the estimation of the Lagrange multiplier ν.

Below, we give the distribution of the penalized HJ-distance when models are misspecified and N goes to infinity. To do so, we exploit the following expression of the penalized HJ-distance (δ_p^2) given in Equations (19) and (21) in Appendix B:

$$\delta^2 = \min_{\theta} \underset{\nu_1 \in \mathbb{R}^N, \nu_2 \in \mathbb{R}}{max} E\left[2y\nu_1'\frac{r}{N} - \frac{\nu_1' r r' \nu_1}{N^2} - \nu_2^2 - 2\frac{\nu_1' r \nu_2}{N} + \psi(\nu_1)\right], \tag{16}$$

where $\psi : \mathbb{R}^N \to \mathbb{R}$ is a concave function representing the penalty, namely $\psi(\nu_1) = -\gamma \|\nu_1\|_N^2$ for Ridge and $\psi(\nu_1) = -\gamma \|\nu_1\|_{N,\Sigma^{-1}}^2$ for Tikhonov.

Let $\nu_{1,\gamma}$ and $\nu_{2,\gamma}$ be the optimal values of the Lagrange multipliers ν_1 and ν_2, and θ_γ the solution of (16) in θ. We can rewrite δ^2 as $E\left[2y\nu_{1,\gamma}'\frac{r}{N} - \frac{\nu_{1,\gamma}' r r' \nu_{1,\gamma}}{N^2} - \nu_{2,\gamma}^2 - 2\frac{\nu_{1,\gamma}' r \nu_{2,\gamma}}{N} + \psi(\nu_{1,\gamma})\right]$. Now, by the first-order condition, we can replace $\nu_{2,\gamma}$ by $-\nu_{1,\gamma}' E(r)/N$. This yields

$$\delta^2 = E\left[2\nu_{1,\gamma}'\frac{yr}{N} - \frac{\nu_{1,\gamma}'\tilde{r}_t\tilde{r}_t'\nu_{1,\gamma}}{N^2} + \psi(\nu_{1,\gamma})\right]$$

or equivalently using $E(y_t) = 1$,

$$\begin{aligned} \delta^2 &= E\left(q_t(\eta_\gamma)\right) \text{ with} \\ q_t(\nu_{1,\gamma}) &= 2\nu_{1,\gamma}'\frac{y_t\tilde{r}_t}{N} - \frac{\nu_{1,\gamma}'\tilde{r}_t\tilde{r}_t'\nu_{1,\gamma}}{N^2} + \frac{2\nu_{1,\gamma}'\mu_2}{N} + \psi(\nu_{1,\gamma}) \end{aligned}$$

and $\eta_\gamma = \left(\theta_\gamma', \nu_{1,\gamma}\right)'$.

We can also show that, for both types of regularizations,

$$\widehat{\delta}_\gamma^2 = \frac{1}{T}\sum_{t=1}^{T}\left[2\widehat{\nu}_{1,\gamma}'\frac{y_t\left(\widehat{\theta}_{HJ}\right)r_t}{N} - \frac{\widehat{\nu}_{1,\gamma}'\tilde{r}_t\tilde{r}_t'\widehat{\nu}_{1,\gamma}}{N^2} + \psi(\widehat{\nu}_{1,\gamma})\right] = \frac{1}{N}\widehat{\alpha}'\hat{\Sigma}_\gamma^{-1}\widehat{\alpha}$$

where $\widehat{\nu}_{1,\gamma} = \hat{\Sigma}_\gamma^{-1}\widehat{\alpha}$ and $\widehat{\alpha} = \sum_{t=1}^{T} y_t\left(\widehat{\theta}_{HJ}\right)r_t/T$.

Assumption 6. *For $0 < \gamma < \infty$, $q_t(\eta)$ is differentiable on an open set N of $\nu_{1\gamma}$ and*
$E\left[\underset{\nu_1 \in N}{sup} \parallel \nabla q_t(\eta) \parallel < \infty\right]$.

This assumption ensures the interchangeability between integration and differentiation for any $0 < \gamma < \infty$.

Proposition 6. *Let $\hat{\delta}_\gamma^2$ be the regularized Hansen-Jagannathan distance with Ridge or Tikhonov regularization. Suppose Assumptions 1-4 and 6 are satisfied and $\delta \neq 0$. As T, N go to infinity and $\gamma T \to \infty$, and $\gamma^2 T \to 0$,*

$$\sqrt{T}\left(\hat{\delta}_\gamma^2 - \delta^2\right) \xrightarrow{d} \mathcal{N}(0, v_2),$$

where

$$v_2 = \lim_{N,T\to\infty} \sum_{j=-\infty}^{\infty} E\left(l_t l_{t-j}\right),$$

and $l_t = 2y_t\nu_1' \frac{\tilde{r}_t}{N} - \frac{\nu_1'\tilde{r}_t\tilde{r}_t'\nu_1}{N^2} + \frac{2\nu_{1,\gamma}'\mu_2}{N} + \psi\left(\nu_1\right) - \delta^2 = 2y_t\tilde{u}_t - \tilde{u}_t^2 + \delta^2 + \psi\left(\nu_1\right),\ \tilde{u}_t = \tilde{r}_t'\Sigma^{-1}\alpha/N,$ *and* $\nu_1 = \Sigma^{-1}\alpha$.

Proposition 6 gives the distribution of the penalized HJ-distance when the model is misspecified. It cannot be used when the model is correctly specified. Indeed, when $\alpha = 0$, $\nu_1 = 0$ and hence $l_t = 0$ which yields a degenerate asymptotic distribution. The asymptotic variance v_2 can be estimated using a HAC estimator and replacing l_t by $\widehat{l}_t$:

$$\widehat{l}_t = 2y_t\left(\widehat{\theta}_{HJ}\right)\widehat{u}_t - \widehat{u}_t^2 + \hat{\delta}^2 + \psi\left(\hat{\nu}_1\right)$$

where $\widehat{u}_t = \widehat{\nu}_1'\overline{r}_t/N$ and $\hat{\nu}_1 = \widehat{\Sigma}_\gamma^{-1}\widehat{\alpha}$. A formal proof of the HAC estimator consistency could be done along the line of that of Lemma 3.

When taking the difference of HJ distances for two asset pricing models, we obtain the following results.

Proposition 7. *Let $\hat{\delta}_\gamma^2$ be the regularized Hansen-Jagannathan distance with Ridge or Tikhonov regularization. Suppose Assumptions 1-4 and 6 are satisfied, $y_a \neq y_b$, and $\delta_a^2, \delta_b^2 \neq 0$. As T, N go to infinity and if γ is chosen such that $\gamma T \to \infty$ and $\gamma^2 T \to 0$,*

$$\sqrt{T}\left((\hat{\delta}_{a\gamma}^2 - \hat{\delta}_{b\gamma}^2) - (\delta_a^2 - \delta_b^2)\right) \xrightarrow{d} \mathcal{N}(0, v_3),$$

where

$$v_3 = \lim_{N,T\to\infty} var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\tilde{l}_{at} - \tilde{l}_{bt}\right)\right],$$

where $\tilde{l}_{m,t} = 2y_{m,t}\tilde{u}_{m,t} - \tilde{u}_{m,t}^2 + \delta_m^2 + \psi\left(\nu_{m,1}\right)$ *for* $m = a, b$. $y_{m,t}$ *is the SDF of model* m *and* $\nu_{m,1} = \Sigma^{-1}\alpha_m$ *where* α_m *represents the pricing errors of model* m. *Moreover, under* $H_0 : \delta_a^2 = \delta_b^2$, *we have the following simplification*

$$\sqrt{T}\left(\hat{\delta}_{a\gamma}^2 - \hat{\delta}_{b\gamma}^2\right) \xrightarrow{d} \mathcal{N}(0, v_4)$$

where v_4 *is the long-run variance of* $\tilde{l}_{at} - \tilde{l}_{bt} = 2y_{a,t}\tilde{u}_{a,t} - \tilde{u}_{a,t}^2 - 2y_{b,t}\tilde{u}_{b,t} + \tilde{u}_{b,t}^2 + \psi\left(\nu_{a,1}\right) - \psi\left(\nu_{b,1}\right)$.

When using Proposition 7 to compare two asset pricing models, one should use the same value of the regularization parameter. We can use Proposition 7 to construct a so-called Normal test of $H_0 : \delta_a^2 = \delta_b^2$ of the form $\sqrt{T}\left(\hat{\delta}_{a\gamma}^2 - \hat{\delta}_{b\gamma}^2\right)\widehat{v}_4^{-1/2}$, where $\widehat{v}_4$ is a HAC estimator of v_4. This statistics converges asymptotically to a standard normal distribution. Note the difference of rates of $\left(\hat{\delta}_{a\gamma}^2 - \hat{\delta}_{b\gamma}^2\right)$ which is $\sqrt{T}$ here instead of T in the previous sections. Why such a difference? The $\sqrt{T}$ rate comes from the rate of $\hat{\delta}_\gamma^2$ obtained in Proposition 6. It would be expected to have the same rate for the difference $\left(\hat{\delta}_{a\gamma}^2 - \hat{\delta}_{b\gamma}^2\right)$ in all scenarios. However, when $y_{at} = y_{bt}$, we have $\alpha_a = \alpha_b$ and $\nu_{a,1} = \nu_{b,1}$ so that $\tilde{l}_{at} - \tilde{l}_{bt} = 0$ and $v_3 = 0$. Because of this simplification, the rate of convergence of $\left(\hat{\delta}_{a\gamma}^2 - \hat{\delta}_{b\gamma}^2\right)$ is faster than the standard $\sqrt{T}$ rate when the SDFs coincide.

Given $\hat{\delta}_{a\gamma}^2 - \hat{\delta}_{b\gamma}^2$ has a different asymptotic distribution depending on whether the model is correctly specified or not, it is not obvious to know which test to use. Kan and Robotti (2009) propose to apply a sequential test inspired from Vuong (1989). However, Vuong's test has been shown to exhibit large size distortions. Shi (2015) and Schennach and Wilhelm (2017) propose one-step tests which have good size and power. The first approach relies on simulated critical values and the second on sample splitting. Implementing these procedures in our setting would be interesting but goes beyond the scope of the present paper.

4.3 Multiple comparison

In this section, we present a comparison test of multiple models. The test is based on the work of Wolak (1963) and Wolak (1989), see also Gospodinov, Kan, and Robotti (2013). Suppose we have $p+1$ models with HJ-distances given by $\delta_i, \quad i = 1, \ldots, p+1$. We are interested in testing whether a benchmark model (Model 1) has an aggregate pricing error as low as the other p models. Let $d_i = \delta_1^2 - \delta_i^2$, $i = 2, \ldots, p+1$ be the difference between the HJ-distance of the benchmark and those of the remaining models and $d = (d_2, \cdots, d_{p+1})'$. The null hypothesis of the test is $H_0 : d \leq 0_p$. The test uses the sample counterpart of d, $\hat{d}_\gamma = \left(\hat{d}_{2\gamma}, \cdots, \hat{d}_{p+1\gamma}\right)'$ for a given value of γ. To have the same framework as Wolak (1989), we rely on the fact that as $N, T \to \infty$ and $\gamma \to 0$,

$$\sqrt{T}(\hat{d}_\gamma - d) \xrightarrow{d} \mathcal{N}(0_p, \Omega_d)$$

using Proposition 7. The latter is valid only when the models have distinct SDFs and they are misspecified, that is $\delta_i > 0$. This seems to be a reasonable assumption given asset pricing models can be viewed as approximations of the reality and usually include different factors. Let $\tilde{d}_\gamma$ be the optimal solution of the following quadratic programming problem

$$\min_d (\hat{d}_\gamma - d)'\hat{\Omega}_{d,\gamma}^{-1}(\hat{d}_\gamma - d) \quad s.t.\ d \leq 0_p,$$

where $\hat{\Omega}_{d,\gamma}$ is a consistent estimator of Ω_d when $N, T \to \infty$ and $\gamma \to 0$. The likelihood ratio statistic of the null hypothesis is

$$LR_\gamma = T(\hat{d}_\gamma - \tilde{d}_\gamma)' \hat{\Omega}_{d,\gamma}^{-1} (\hat{d}_\gamma - \tilde{d}_\gamma). \tag{17}$$

The distribution of the previous statistics is obtained under the least favorable value, i.e. $d = 0_p$. We have $LR_\gamma \xrightarrow{d} \sum_{i=0}^{p} w_{p-i}(\Omega_d)\chi^2(i)$, where the weights w_i sum up to one[3] and $\chi^2(i)$ are independent Chi-square random variables with i degrees of freedom and $\chi^2(0)$ is a dirac at 0. See Kudo (1987) and Wolak (1963) for more details.

5 Monte Carlo Simulations

In this section, we run several Monte Carlo simulations to showcase the advantage of the regularization schemes described previously. The parameters for the simulations are calibrated on real data.

5.1 SDF parameter estimates

In this subsection, we analyze the small sample properties of the SDF parameter test. We are interested in testing whether a particular factor is priced in the returns. This corresponds to testing whether a SDF parameter is null using a t-statistics. We compare the small-samples size properties of our test with that of Kan and Robotti (2008).

We consider the three factor model of Fama and French (1993) (FF3), where the risk factors are the market excess return (MKT, f_1), the return difference between portfolios of small and large stocks (SMB, f_2), and the return difference between portfolios of high and low book-to-market ratios (HML, f_3). Our simulation strategy follows closely that of Fan, Fan, and Lv (2008).

First, we estimate the factor model

$$r_t = \mu_2 + \beta_1 \tilde{f}_{1t} + \beta_2 \tilde{f}_{2t} + \beta_3 \tilde{f}_{3t} + \varepsilon_t \tag{18}$$

using for r_t the excess returns of 202 combined portfolios going from July 1963 to January 2026 extracted from the Kenneth French's Website; these portfolios are the same equity portfolios used in Giglio and Xiu (2021). The portfolios list is presented in Table 8 in Appendix C. The factors are also from Kenneth French's website. So we obtain a 202×3 matrix B of estimators of $(\beta_1, \beta_2, \beta_3)$, we compute their 3×1 mean vector μ_b and 3×3 variance matrix Σ_b. We also estimate the mean μ_1 and variance V_{11} of the factors. We compute the standard deviation (std) of the residuals of the regression (18) and obtain their sample mean and std. In the sequel, we will draw the

[3] Appendix C of Gospodinov, Kan, and Robotti (2013) gives the procedure to compute $w_i(\Omega_d)$ and the p-value of the test.

standard deviations from a Gamma distribution. To calibrate the two parameters of the Gamma distribution, we use the method of moments in order to match the sample mean and sample std of the residual std with the theoretical ones.

For each simulation, we proceed as follows:

(a) We first generate a random sample of $f = (f_1, f_2, f_3)'$ of size T from the normal distribution $N(\mu_1, V_{11})$.

(b) Then, we generate a random sample of N factor loadings $(b_1, b_2, ..., b_N)$ from $N(\mu_b, \Sigma_b)$.

(c) We generate a random sample of N standard deviations $\sigma_1, ..., \sigma_N$ from a Gamma distribution which two parameters have been estimated as previously explained.

(d) Once, we have N standard deviations of the errors, we generate a random sample of $\varepsilon = (\varepsilon_1, ..., \varepsilon_N)'$ of size T from $N\left(0, diag\left(\sigma_1^2, ..., \sigma_N^2\right)\right)$.

(e) Then, we get a random sample $r = (r_1, ..., r_T)$ from (18). If $N > 202$, we construct μ_2 by concatenating several μ_2.

As the mean μ_2 is not constrained, the SDF is typically misspecified. For our sample with $N = 202$ test assets and $T = 751$, the (non-regularized) Hansen-Jagannathan distance equals 1.25.

We consider $N = 50$, 300, and 1000 and $T = 150$, 350, and 650. We run 10000 simulations to evaluate the empirical level and power of the tests. In all cases, the regularization parameter γ is chosen using the procedure described in Section 3.4 where c is chosen in a log-scale grid going from 0.0001 to 0.01. Standard errors and covariance matrices are robust to misspecification but not to autocorrelation. Table 1 reports the empirical size of the t-test on the individual θ coefficients in the SDF. To evaluate the size of the test on the jth coefficient, we generate the model without the jth factor and estimate a full model including it. We compare our test with that of Kan and Robotti (2008). When $N > T$, the matrix $\widehat{V}_{22}$ is singular and their test is not directly applicable. In that case, we use the Moore-Penrose inverse of $\widehat{V}_{22}$. This test coincides with our test when setting $\gamma = 0$.

When the number of test assets, N, is small, $N = 50$, we observe that the size of both tests is roughly the same and they both tend to overrejects slightly. When N increases, the size of the Moore-Penrose based test decreases rapidly while that of the regularized test remains around or slightly above the nominal size.

We now turn our attention to the empirical power of our t-test. Table 2 presents the rejection rate of testing $H_0 : \theta_j = 0$ when the SDF parameter is non null. The Moore-Penrose based test (Panel A) has reasonable power for $N = 50$ but its power falls below the empirical size for $N = 300$. This is due to the instability of the Moore-Penrose inverse mentioned in Subsection 2.2 and the fact that for N close to T, the sample covariance and hence the estimators and their std are badly estimated. To gain more insight, Table 13 in Appendix C reports some statistics regarding the estimators of the FF3 parameters when $N = 300$ for both methods. The statistics are the mean bias (using the pseudo-true value), median bias, and standard deviation calculated over 10000 simulations. For $T = 150$, the Moore-Penrose based estimators exhibit a slightly larger bias than that for the regularized estimator and a much larger variance. This large variance and the fact that, in absence of the regularization, the standard errors of the estimators are poorly estimated explain the low power of the Moore-Penrose based test. Interestingly, we observe a slight improvement of its power when $N = 1000$ (see Table 2). This echoes the double-descent phenomenon discussed in Hastie, Montanari, Rosset, and Tibshirani (2022) and Liao, Ma, Neuhierl, and Shi (2026). On the other hand, the Tikhonov based test (Panel B) exhibits a good power throughout the values of N and T. Its power increases with both T and N, suggesting that augmenting the number of test assets is beneficial in terms of power. When comparing the test performance between the three factors, we observe that the rejection rate of the market is higher than for the two other factors. The values of γ obtained by cross-validation in the different simulations are reported in Appendix C. We observe that as expected γ decreases with T which makes sense because the covariance matrix becomes less ill-conditioned when T increases. Interestingly, γ tends to increase slightly with N for $T = 150$ but remains almost constant for larger T.

In summary, regularization is necessary for the t-test to maintain power when N increases. A well-chosen regularization parameter permits to take advantage from a large number of test assets.

Table 1: Empirical size of t-tests under misspecification

N	50	300	1000	50	300	1000	50	300	1000
Panel A: Test using Moore Penrose inverse									
	θ_{MKT}			θ_{SMB}			θ_{HML}		
$T = 150$	0.066	0.025	0.013	0.068	0.040	0.011	0.068	0.039	0.012
$T = 350$	0.061	0.058	0.018	0.058	0.058	0.016	0.064	0.058	0.018
$T = 650$	0.058	0.053	0.016	0.060	0.057	0.028	0.062	0.055	0.032
Panel B: Test using Tikhonov regularized inverse									
$T = 150$	0.065	0.060	0.066	0.074	0.065	0.054	0.079	0.062	0.052
$T = 350$	0.058	0.057	0.057	0.061	0.059	0.044	0.068	0.056	0.044
$T = 650$	0.057	0.051	0.050	0.059	0.054	0.046	0.058	0.055	0.046

Note: The theoretical size is 5%.

Table 2: Empirical power of t-tests under misspecification

N	50	300	1000	50	300	1000	50	300	1000
Panel A: Test using Moore Penrose inverse									
	θ_{MKT}			θ_{SMB}			θ_{HML}		
$T = 150$	0.410	0.042	0.089	0.064	0.033	0.082	0.131	0.015	0.050
$T = 350$	0.942	0.000	0.080	0.177	0.000	0.091	0.381	0.000	0.044
$T = 650$	1.000	0.001	0.047	0.318	0.000	0.069	0.595	0.000	0.021
Panel B: Test using Tikhonov regularized inverse									
$T = 150$	0.907	0.966	0.983	0.581	0.723	0.821	0.634	0.793	0.871
$T = 350$	0.997	1.000	1.000	0.612	0.734	0.847	0.737	0.889	0.960
$T = 650$	1.000	1.000	1.000	0.722	0.850	0.997	0.846	0.988	1.000

Note: The theoretical size is 5%.

5.2 Model comparison tests

In this section, we investigate the finite sample behavior of the pairwise and multiple comparison tests. Table 3 presents the results. One must keep in mind that when comparing models, it is essential to use the same penalization parameter. In this section, we will compare the Fama-French's FF3 model with a three factor model which we call q3. This model includes three factors: the market and the first two factors of q5 from Hou, Mo, Xue, and Zhang (2021), namely the size factor (ME) and the investment factor (IA). For calibration, the dataset was downloaded from Lu Zhang's webpage[4] and goes from January 1967 to December 2024. Finally, we also consider the polynomial type of model used in Dittmar (2002). The SDF of the model is given by

$$y = 1 - \theta_{MKT} \left(r_{MKT} - E\left[r_{MKT} \right] \right) - \theta_{mkt,2} \left(r_{MKT}^2 - E\left[r_{MKT}^2 \right] \right) - \theta_{mkt,3} \left(r_{MKT}^3 - E\left[r_{MKT}^3 \right] \right)$$

where the quadratic and cubic terms are supposed to reflect the preference of agents for higher moments. Without being fully nonparametric, these terms permit to incorporate nonlinearity in the SDF.

Panel A presents the tests developed in Proposition 5. The latter tests the equality of two non-nested SDFs. The simulated data are from FF3 and q3 models. To evaluate the size, we set the mean of the returns such that the non-overlapping factors have null SDF parameters and the two models are misspecified (so the DGP under the null contains only the market). The Wald test tests the nullity of the coefficients of the non-market factors in an encompassing complete model containing the 5 factors of q3 and FF3. On the other hand, the Weighted χ^2 test uses (15) based on the estimation of two separate models (q3 and FF3). To analyze the power, we generate a factor model which includes both the factors from FF3 and from q3, so that SDF parameters of the non-overlapping factors are non-zero. We choose γ by two-fold cross-validation applied to the complete model which includes all 5 factors. The results show that the two tests tend to overreject slightly in small samples. They both exhibit a strong power which increases in both N and T. This would not be the case if one would use the approach of Kan and Robotti (2008) as this test tends to underreject when N is large.

Panel B presents the test of equality of the HJ-distances of two models when $y_a \neq y_b$. The test uses the Normal pairwise test statistic of Proposition 7. To evaluate the size of the test, we compare two misspecified models with three factors each. The two models have in common SMB and HML factors. For each model, we also include an extra factor obtained by adding to MKT a normally distributed error with mean 0 and variance 20% of the market variance. This guarantees that the models have different SDFs and the same HJ-distance. Under the null, the DGP is FF3 model. We observe that the test is very conservative. This is not the case in absence of regularization when N is small as shown in Gospodinov, Kan, and Robotti (2013). In fact, the Tikhonov regularization tends to shrink the HJ distance; this may explain the small size of the test based on differences of regularized HJ distances. To evaluate the power, the DGP is a

[4] https://global-q.org/factors.html

one-factor model with 3rd factor of q5 while the two competing models are one factor model with squared MKT and a two-factor model with SMB and HML. We observe that the test has good power. Interestingly, the power does not necessarily increase with N, this seems to be due to the fact that γ is optimized for each pairs (T, N).

Panel C shows the finite sample behavior of the comparison test of multiple models. The test uses Wolak statistic (17). To evaluate the size, we repeat the same procedure as in Panel B. For $p = 1$, we use two FF3s and for $p = 2$, three FF3s. To evaluate the power, we simulate a model with IA of q5 as DGP. The benchmark is the squared market and the alternative is a two-factor model with SMB and HML for $p = 1$, whereas the alternative models are a one-factor model with IA plus noise and a two-factor model with SMB and HML for $p = 2$. We employ the γ of the benchmark model to run the tests. The average regularized squared HJ-distance for $N = 50$ and $T = 150$ of the benchmark is 0.956, of the model with SMB and HML is 0.878, and of the model with noisy IA is 0.894. The results show that Wolak test is conservative but exhibits high empirical power. Particularly, the pairwise test ($p = 1$) has a higher empirical power than the Normal pairwise test of Panel B which makes sense because it is a one-sided test.

Table 3. Model comparison tests under misspecification

N	50	300	1000	50	300	1000	50	300	1000	50	300	1000
	Panel A: Pairwise tests of equality of two SDFs											
	Wald test						Weighted χ^2 test					
T	Size			Power			Size			Power		
150	0.078	0.052	0.038	0.759	0.905	0.969	0.062	0.068	0.063	0.606	0.789	0.875
350	0.060	0.044	0.036	0.862	0.969	0.998	0.060	0.061	0.056	0.759	0.910	0.970
650	0.052	0.036	0.033	0.937	0.996	1.000	0.054	0.056	0.051	0.858	0.957	0.998
	Panel B: Normal pairwise test of equality of two HJ-distances (with different SDFs)											
T	Size			Power								
150	0.0021	0.000	0.000	0.048	0.149	0.067						
350	0.004	0.003	0.001	0.416	0.501	0.735						
650	0.014	0.016	0.008	0.751	0.705	0.908						
	Panel C: Multiple comparison test (Wolak test)											
	Wolak test ($p = 1$)						Wolak test ($p = 2$)					
T	Size			Power			Size			Power		
150	0.002	0.001	0.000	0.145	0.307	0.255	0.001	0.000	0.000	0.126	0.288	0.217
350	0.009	0.010	0.003	0.618	0.611	0.868	0.004	0.004	0.001	0.611	0.669	0.902
650	0.019	0.020	0.014	0.829	0.757	0.921	0.013	0.013	0.009	0.832	0.822	0.954

Note: The theoretical size is 5%.

6 Empirical application

For the empirical application, we compare four popular models, namely the Fama and French (1993) model (FF3), the q5 of Hou, Mo, Xue, and Zhang (2021), the nonlinear model of Dittmar (2002), and the Fama and French (2015) model (FF5). The latter add two new factors to the FF3: the profitability (RMW) and investment (CMA) factors. The q5 model has the following factors: the market, the size factor (ME) and the investment factor (IA), the equity factor (ROE), and the expected growth factor (EG).

For this analysis, we use the large set of test assets put together by Giglio and Xiu (2021)[5]. This set of $N = 647$ portfolios includes 202 equity portfolios from Kenneth French's website, treasury bonds, corporate bonds, currencies, and US equities sorted by a large number of characteristics, see Table 8 in Appendix C. The data spans the period from March 1976 to January 2010, so that we have $T = 407$. In this sample, T is quite a bit smaller than N so that regularization is necessary.

Table 4 presents the estimation of the SDF parameters of the four models. For the FF3, we note that all the factors are priced, while for FF5, HML is not priced, so the value factor disappears. This outcome is in line with the results of Fama and French (2015), who argue that the value factor is redundant as the model with the five factors does not improve upon the model with just four factors without HML. For the nonlinear model, the coefficient of MKT^2 is not significant. Finally, in the q5 model, the equity and investment factors are unpriced. This model exhibits the lowest pricing errors. For this table, the value of the regularization parameter has been optimized separately for each model, therefore the squared HJ distances $\hat{\delta}^2_\gamma$ are not comparable across the four models because they use different values of γ.

[5]The data was downloaded https://stefanogiglio.org/#data. We thank Dacheng Xiu and Stefano Giglio for letting us use this dataset.

Table 4: SDF parameter estimates

	FF3				Nonlinear model		
Factors	θ_{MKT}	θ_{SMB}	θ_{HML}		θ_{MKT}	$\theta_{MKT\hat{}2}$	$\theta_{MKT\hat{}3}$
SDF	0.035***	0.026***	0.061***		-0.124***	0.001	0.001**
t-ratio	5.795	5.347	3.790		-2.760	0.063	2.378
p-value	0.000	0.000	0.000		0.006	0.950	0.017
$\hat{\delta}^2_\gamma$	0.055	γ	0.0347		0.038	γ	0.0347
				FF5 model			
Factors	θ_{MKT}	θ_{SMB}	θ_{HML}	θ_{RMW}	θ_{CMA}		
SDF	0.060***	0.061***	-0.024	0.113***	0.162***		
t-ratio	8.988	5.763	-1.027	5.910	5.157		
p-value	0.000	0.000	0.304	0.000	0.000		
$\hat{\delta}^2_\gamma$	0.372	γ	0.0003				
				q5 model			
Factors	θ_{MKT}	θ_{ME}	θ_{IA}	θ_{ROE}	θ_{EG}		
SDF	0.077***	0.081***	0.060	-0.033	0.352***		
t-ratio	9.056	8.975	1.575	-0.788	2.731		
p-value	0.000	0.000	0.115	0.431	0.006		
$\hat{\delta}^2_\gamma$	0.010	γ	0.0347				

***,**,* indicate that the null hypothesis of unpriced source of risk is rejected at the 1%, 5%, and 10% levels. The standard errors are robust to both misspecification and autocorrelation (they use Newey-West HAC estimator).

We also examine whether the models exhibit different explanatory power. To compare models, it is essential to keep the same level of penalization. We use a penalization level $\gamma = 0.00034$ which is optimal for a model including the factors of FF5 and q5 together. For this value of γ, the regularized squared HJ distances of the different models are given by 0.270 (q5), 0.372 (FF5), 0.440 (FF3), and 0.466 (nonlinear).

First, we perform pairwise comparison tests. Table 5 reports the results. For the nested models FF3 and FF5, we use a Wald test of the joint significance of the additional factors RMW and CMA. The null hypothesis is strongly rejected, indicating that FF5 provides a better fit than FF3.

For all other pairs of models, we apply the asymptotic normal test of Proposition 7, based on differences in regularized squared HJ distances. This test requires that the competing models have distinct SDFs and are misspecified. The first condition is satisfied since the models differ in their factor structures; in particular, equality of SDFs would require that all models reduce to a function of the market factor alone, which is not supported by the data as evidenced in Table 4. The second condition is also plausible in this setting, as all models considered are parsimonious and do not fully span the cross-section of returns.

The results can be summarized as follows. FF3 and the nonlinear model do not exhibit statistically significant differences in pricing performance, as indicated by the high p-value. In contrast, all other model pairs display significant differences in their HJ distances, suggesting meaningful variation in their ability to explain returns.

Table 5: Pairwise HJ-distance comparison tests

	FF3	Nonlinear	q5
FF5	63.319	7.997	6.033
	(0.000)	(0.005)	(0.014)
FF3		1.144	12.087
		(0.285)	(0.000)
Nonlinear			24.118
			(0.000)

P-values are in brackets. Covariance matrices use Newey-West HAC estimator.

Finally, to determine which model is best, we implement the multiple model comparison test of Wolak (1989). The null hypothesis of this test is that the squared HJ-distance of a benchmark model is smaller than those of two or more alternative models. In turn, we consider each model as a benchmark and compare it against the others. For each test, we remove alternative models nested by the benchmark model as H_0 is verified by construction (the benchmark has already lower pricing errors (HJ)). Within the remaining alternatives, we also remove models nested by others. Finally, we remove alternative models that nest the benchmark as the asymptotic normality assumption on d_i does not hold under the null of $d_i = 0$. For example, to compare the FF3 against the other models, we remove FF5 from the alternative.

Table 6 presents the results of these comparisons. Each line represents the benchmark model. For FF3, the small p-value of 0.000 suggests that its pricing performance is not as good as those of the alternatives (q5 and the nonlinear model). Similarly, for FF5 and the nonlinear models, the low p-values indicate that these models might be dominated by one of the alternatives. Finally, the null hypothesis cannot be rejected for q5. In conclusion, q5 has significantly smaller pricing error than FF3, FF5, and the nonlinear models.

Table 6: Multiple model comparison tests

Benchmark	Alternatives	$\hat{\delta}^2_\gamma$	LR	p-value
FF3	q5 and nonlinear	0.440	6.043	0.000
FF5	q5 and nonlinear	0.372	3.016	0.016
Nonlinear	FF5 and q5	0.466	12.708	0.000
q5	FF5 and nonlinear	0.270	0.000	0.633

Note: Covariance matrices use Newey-West HAC estimator.

7 Conclusion

This paper develops a framework for evaluating and comparing asset pricing models in settings with a large number of test assets. We introduce regularized versions of the Hansen–Jagannathan distance based on Tikhonov and ridge schemes to stabilize the inversion of the return covariance matrix. This allows the SDF parameters to be estimated even when the number of assets exceeds the number of time observations. In addition to improving numerical stability, these regularizations admit an economic interpretation in terms of allowing controlled pricing errors, which can be viewed as arising from transaction costs or market frictions. The regularization parameter is selected using a data-driven procedure based on out-of-sample performance.

Building on this framework, we propose a set of tests for comparing competing asset pricing models. Our approach focuses on differences in regularized HJ distances, providing a natural way to rank models under misspecification. For linear SDF specifications, the resulting statistics admit simple analytical forms and are straightforward to implement. Monte Carlo evidence shows that the proposed regularization substantially improves the finite-sample performance of standard procedures, correcting the lack of power observed when using the Moore–Penrose inverse.

Several directions for future research remain. First, it would be useful to develop inference procedures that are robust to weak factors, i.e., factors with low correlation with returns. Second, our analysis is restricted to linear SDFs; extending the framework to nonlinear models, for example along the lines of Gospodinov, Kan, and Robotti (2013), would be of clear interest. Third, since the HJ distance is based on second moments, alternative discrepancy measures incorporating higher moments, as in Almeida and Garcia (2012), could provide additional insights. Finally, our empirical analysis relies on a set of test portfolios. Expanding the cross-section using transformations of firm characteristics, as in Didisheim, Ke, Kelly, and Malamud (2024), would generate a much richer set of moment conditions and provide a natural setting where regularization becomes even more critical.

8 Appendix A: Short review on regularization

In this section, we introduce some notations and results on norms and operators. Then, we explain how Ridge and Tikhonov regularizations act on eigenvalues to stabilize the inverse of the covariance matrix. Finally, we compare the two regularizations.

Definition 2.

1. *For a vector $v \in \mathbb{R}^N$, $\|v\| = \sqrt{v'v}$ is the euclidean norm and $\|v\|_N = \sqrt{v'v/N}$.*
2. *For an arbitrary $K \times N$ matrix V, the operator norm of V is $\|V\|_{op} = \sup_{\|\phi\|_N = 1} \|V\phi\|_N$. Therefore, for any vector $u \in \mathbb{R}^N$, $\|Vu\|_N \leq \|V\|_{op} \| \|u\|_N$.*
3. *We define the Frobenius norm as $\|V\|_F = \left(tr(V^{'}V)\right)^{\frac{1}{2}}$. We have $\|V\| \leq \|V\|_F$ and for any vector $u \in \mathbb{R}^N$, $\|Vu\| \leq \|V\|_F \|u\|$.*
4. *If $\|v\|_N < \infty$ when $N \to \infty$, we note $\|v\|_\infty$ its limit value.*

Operators and spectral decomposition

Recall that $\mathbb{R}^N$ is endowed by the norm $\| \phi \|_N^2 = \phi'\phi/N$. Let $\mathbb{R}^T$ be endowed with norm $\| v \|_T^2 = \dfrac{v^{'}v}{T}$ induced by inner product $\langle v_1, v_2 \rangle_T = \dfrac{v_1^{'}v_2}{T}$. Let H be the operator from $\mathbb{R}^N$ to $\mathbb{R}^T$ defined by $H\phi = \dfrac{\bar{R}\phi}{N}$ and its adjoint H^* is the operator from $\mathbb{R}^T$ to $\mathbb{R}^N$ such that $H^*v = \dfrac{\bar{R}'v}{T}$. With that, we have the operator $H^*H\phi = \dfrac{\bar{R}'\bar{R}}{NT}\phi = \hat{\Sigma}\phi$ which goes from $\mathbb{R}^N$ to $\mathbb{R}^N$. Let $\left\{\sqrt{\hat{\lambda}_j}, \hat{\phi}_j, \hat{v}_j\right\}$, $j = 1, 2, ...$ be the singular value decomposition of H such that $H\widehat{\phi}_j = \sqrt{\hat{\lambda}_j}\hat{v}_j$ and $H^*\widehat{v}_j = \sqrt{\hat{\lambda}_j}\hat{\phi}_j$. Note that $\left\{\hat{\lambda}_j, \hat{\phi}_j\right\}$, $j = 1, 2, ..., min(N, T)$ are the non zero eigenvalues and eigenvectors of $\hat{\Sigma}$.

Regularity space

Notice that as $\hat{\beta} = \dfrac{\bar{R}'P_F}{T}$, where $P_F = \bar{F}(\dfrac{\bar{F}'\bar{F}}{T})^{-1} = \begin{bmatrix} P_F^1 & \cdots & P_F^K \end{bmatrix}$. Then, $\hat{\beta}$ can be rewritten as $\hat{\beta} = \begin{bmatrix} H^*P_F^1 & \cdots & H^*P_F^K \end{bmatrix}$. Therefore, $\hat{\beta}_k \in \mathcal{R}(H^*), k = 1, \cdots, K$. From Proposition 6.2 of Carrasco, Florens, and Renault (2007), $\mathcal{R}(H^*) = \mathcal{H}(\hat{\Sigma}) = \mathcal{R}(\hat{\Sigma}^{\frac{1}{2}})$ where $\mathcal{H}(\hat{\Sigma})$ is the Reproducing Kernel Hilbert Space (RKHS) of $\hat{\Sigma}$. In Assumption 3, we impose the stronger assumption $\beta \in \mathcal{R}\left(\Sigma^2\right)$.

Relationship between Ridge and Tikhonov

For Ridge regularization, $\widehat{\Sigma}^{-1}$ is replaced by $\widehat{\Sigma}_R^{-1} = \left(\widehat{\Sigma} + \gamma I_N\right)^{-1}$. So this regularization method replaces the explosive eigenvalues of $\widehat{\Sigma}^{-1}$, $\dfrac{1}{\widehat{\lambda}_j}$, $j = 1, \cdots, N$ by $\dfrac{1}{\widehat{\lambda}_j + \gamma}$ which is bounded by $1/\gamma$. On the other hand, for Tikhonov regularization, $\widehat{\Sigma}^{-1}$ is replaced by $\widehat{\Sigma}_K^{-1} = \left(\widehat{\Sigma}^2 + \gamma I_N\right)^{-1}\widehat{\Sigma}$.

Therefore, the eigenvalues $\dfrac{1}{\widehat{\lambda}_j}$ are replaced by $\dfrac{\widehat{\lambda}_j}{\widehat{\lambda}_j^2 + \gamma}$ which is bounded by $\widehat{\lambda}_j/\gamma$ and is exactly zero when $\widehat{\lambda}_j = 0$. It may seem like quite different regularization techniques, however the difference is minor in our application. Indeed, in the formulas of $\widehat{\delta}_\gamma^2$ and $\hat{\theta}$, $\widehat{\Sigma}^{-1}$ is applied to either $\hat{\beta} = \dfrac{\overline{R}'\overline{F}}{T}\left(\dfrac{\overline{F}'\overline{F}}{T}\right)^{-1}$ or $\widehat{\mu}_2 = \overline{R}' 1_T/T$. Both are elements of the RKHS of $\hat{\Sigma}$. This yields simplification as explained below.

Let the $N \times 1$ vector v be an element of the RKHS of $\hat{\Sigma}$, i.e. there exists a $T \times 1$ vector w such that $v = \overline{R}' w/T$.

1. **Ridge regularization**

$$\begin{aligned}
\widehat{\Sigma}_R^{-1} v &= \sum_j \frac{1}{\hat{\lambda}_j + \gamma} \left\langle v, \widehat{\phi}_j \right\rangle_N \widehat{\phi}_j \\
&= \sum_j \frac{1}{\hat{\lambda}_j + \gamma} \left\langle \overline{R}' w, \widehat{\phi}_j \right\rangle_N \widehat{\phi}_j \\
&= \sum_j \frac{\sqrt{\hat{\lambda}_j}}{\hat{\lambda}_j + \gamma} \left\langle w, \widehat{\phi}_j \right\rangle_T \widehat{\phi}_j \\
&= \sum_j \frac{q(\gamma, \widehat{\lambda}_j)}{\sqrt{\hat{\lambda}_j}} \left\langle w, \widehat{\phi}_j \right\rangle_T \widehat{\phi}_j
\end{aligned}$$

where $q(\gamma, \mu) = \dfrac{\mu}{\gamma + \mu}$.

2. **Tikhonov regularization**

$$\begin{aligned}
\widehat{\Sigma}_K^{-1} v &= \sum_j \frac{\hat{\lambda}_j}{\hat{\lambda}_j^2 + \gamma} \left\langle v, \widehat{\phi}_j \right\rangle_N \widehat{\phi}_j \\
&= \sum_j \frac{\hat{\lambda}_j}{\hat{\lambda}_j^2 + \gamma} \left\langle \overline{R}' w, \widehat{\phi}_j \right\rangle_N \widehat{\phi}_j \\
&= \sum_j \frac{\sqrt{\hat{\lambda}_j}\hat{\lambda}_j}{\hat{\lambda}_j^2 + \gamma} \left\langle w, \widehat{\phi}_j \right\rangle_T \widehat{\phi}_j \\
&= \sum_j \frac{q(\gamma, \widehat{\lambda}_j^2)}{\sqrt{\hat{\lambda}_j}} \left\langle w, \widehat{\phi}_j \right\rangle_T \widehat{\phi}_j.
\end{aligned}$$

Therefore, Tikhonov and Ridge regularizations have the same formula except for $q(\gamma, \widehat{\lambda}_j^2)$ replaced by $q(\gamma,\widehat{\lambda}_j)$. So both regularizations give basically the same results (the only difference is that the optimal rate for γ may be different).

9 Appendix B: Penalization on the Lagrange multiplier

In Section 3.1, we provided an interpretation of the regularization in terms of a penalization on the norm of the pricing error. Here, we show that it involves a penalization on the Lagrange multipliers. To get insights on the mechanism behind the penalization in (6), we consider the dual of the optimization problem (6) (see the proof of Proposition 1 in the supplementary appendix):

$$\delta_R^2 = \underset{\nu_1\in\mathbb{R}^N,\nu_2\in\mathbb{R}}{max} E\left\{2y\nu_1^{'}\frac{r}{N} - \frac{\nu_1^{'}rr^{'}\nu_1}{N^2} - \nu_2^2 - 2\frac{\nu_1^{'}r\nu_2}{N} - \frac{\gamma}{N}\parallel \nu_1 \parallel^2\right\}. \tag{19}$$

Remark that Equation (19) is the penalized version of the dual of (5) with a penalization applied to ν_1, the Lagrange multiplier associated with $E\,(mr)$.

The first order condition with respect to ν_1 gives

$$\alpha - \frac{E\,(rr')}{N}\nu_1 - \nu_2\mu_2 - \gamma\nu_1 = 0. \tag{20}$$

The first order condition with respect to ν_2 gives $\nu_2 = -\nu_1'\mu_2/N$. Replacing ν_2 in Equation (20) gives

$$\nu_1 = (\Sigma+\gamma I_N)^{-1}\alpha.$$

And the solution is the ridge regularized Hansen-Jagannathan distance

$$\delta_R^2 = \frac{\alpha'(\Sigma+\gamma I_N)^{-1}\alpha}{N}.$$

For Tikhonov regularization, the dual is given by

$$\delta_K^2 = \underset{\nu_1\in\mathbb{R}^N,\nu_2\in\mathbb{R}}{max} E\left\{2y\nu_1^{'}\frac{r}{N} - \frac{\nu_1^{'}rr^{'}\nu_1}{N^2} - \nu_2^2 - 2\frac{\nu_1^{'}r\nu_2}{N} - \frac{\gamma}{N}\nu_1'\Sigma^{-1}\nu_1\right\}. \tag{21}$$

Solving in ν_1 and ν_2 yields $\nu_1 = [\Sigma^2+\gamma I_N]^{-1}\Sigma E[ry] = [\Sigma^2+\gamma I_N]^{-1}\Sigma\alpha$, $\nu_2 = -\nu_1'\mu_2/N$, and

$$\delta_K^2 = \frac{\alpha^{'}[\Sigma^2+\gamma I_N]^{-1}\Sigma\alpha}{N}.$$

So for both regularizations, the penalization on $\|E\,(mr)\|_N^2$ translates into a penalization on the Lagrange multiplier ν_1 and hence relaxes the condition.

10 Appendix C: Extra information for the simulations and empirical application

Table 7: List of test portfolios for simulations

25 portfolios formed on size and book-to-market
17 industry portfolios
25 portfolios formed on operating profitability and investment
25 portfolios formed on size and variance
35 portfolios formed on size and net share issues
25 portfolios formed on size and accruals
25 portfolios formed on size and market beta
25 portfolios formed on size and momentum

Table 8: List of test portfolios for the empirical application

202 portfolios from Table 7
10 maturity-sorted government bond portfolios
10 corporate bond portfolios sorted on yield spread
6 currency portfolios sorted on interest rate differentials
6 currency portfolios sorted on currency momentum
413 anomaly portfolios based on 103 characteristics

For more details, see Section III.1. of the supplementary appendix of Giglio and Xiu (2021).

Table 9: Regularization parameter γ used for Table 1 when testing MKT

	$N = 50$	$N = 300$	$N = 1000$
$T = 150$	0.002 (0.002)	0.001 (0.001)	0.001 (0.001)
$T = 350$	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
$T = 650$	0.001 (0.001)	0.000 (0.000)	0.000 (0.000)

Notes: The values in the cells correspond to the mean of γ over simulations and its standard deviation in parenthesis.

Table 10: Regularization parameter γ used for Table 1 when testing SMB

	$N = 50$	$N = 300$	$N = 1000$
$T = 150$	0.031 (0.036)	0.031 (0.037)	0.034 (0.038)
$T = 350$	0.016 (0.018)	0.016 (0.018)	0.016 (0.019)
$T = 650$	0.010 (0.011)	0.009 (0.011)	0.009 (0.011)

Notes: The values in the cells correspond to the mean of γ over simulations and its standard deviation in parenthesis.

Table 11: Regularization parameter used for Table 1 when testing HML

	$N = 50$	$N = 300$	$N = 1000$
$T = 150$	0.052 (0.064)	0.055 (0.066)	0.059 (0.067)
$T = 350$	0.027 (0.031)	0.027 (0.032)	0.029 (0.033)
$T = 650$	0.017 (0.019)	0.016 (0.019)	0.017 (0.020)

Notes: The values in the cells correspond to the mean of γ over simulations and its standard deviation in parenthesis.

Table 12: Regularization parameter used for Table 2

	$N = 50$	$N = 300$	$N = 1000$
$T = 150$	0.055 (0.063)	0.057 (0.064)	0.061 (0.065)
$T = 350$	0.027 (0.031)	0.028 (0.031)	0.028 (0.031)
$T = 650$	0.016 (0.018)	0.016 (0.018)	0.016 (0.018)

Notes: The values in the cells correspond to the mean of γ over simulations and its standard deviation in parenthesis.

Table 13: Bias and standard deviation of the θ estimators for $N = 300$

	mean bias	med bias	std	mean bias	med bias	std	mean bias	med bias	std
Panel A: Estimation using Moore Penrose inverse									
	θ_{MKT}			θ_{SMB}			θ_{HML}		
$T = 150$	0.027	-0.012	11.22	-0.260	0.004	33.34	0.156	0.003	34.83
$T = 350$	0.001	0.000	0.011	-0.000	-0.000	0.016	0.000	0.000	0.021
$T = 650$	0.000	0.000	0.003	-0.000	-0.000	0.004	0.000	-0.000	0.007
Panel B: Estimation using Tikhonov regularized inverse									
$T = 150$	0.001	0.001	0.019	-0.000	-0.000	0.026	-0.000	-0.000	0.026
$T = 350$	0.001	0.000	0.010	-0.000	-0.000	0.014	-0.000	-0.000	0.015
$T = 650$	-0.000	0.000	0.003	-0.000	-0.000	0.004	-0.000	-0.000	0.007

References

Almeida, C., and R. Garcia (2012): "Assessing misspecified asset pricing models with empirical likelihood estimators," *Journal of Econometrics*, 170(2), 519–537.

——— (2017): "Economic Implications of Nonlinear Pricing Kernels," *Management Science*, 63, 3361–3380.

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

Antoine, B., K. Proulx, and E. Renault (2020): "Pseudo-True SDFs in Conditional Asset Pricing Models," *Journal of Financial Econometrics*, 18(4), 656–714.

Babii, A., E. Ghysels, and J. Striaukas (2024): "High-Dimensional Granger Causality Tests with an Application to VIX and News," *Journal of Financial Econometrics*, 22(3), 605–635.

Barillas, F., and J. Shanken (2018): "Comparing Asset Pricing Models," *The Journal of Finance*, 73(2), 715–754.

Barras, L. (2019): "A large-scale approach for evaluating asset pricing models," *Journal of Financial Economics*, 134, 549–569.

Bauschke, H. H., and P. L. Combettes (2017): *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics. Springer International Publishing, Cham.

Bernstein, D. S. (2009): *Matrix Mathematics: Theory, Facts, and Formulas (Second Edition)*. Princeton University Press.

Borwein, J. M. (1993): "On the failure of maximum entropy reconstruction of Fredholm equations and other infinite systems," *Mathematical Programming*, 61, 251–261.

Borwein, J. M., and A. S. Lewis (1992): "Partially finite convex programming, Part I: Quasi relative interiors and duality theory," *Mathematical Programming*, 57(1-3), 15–48.

Brodie, J., I. Daubechies, C. De Mol, and I. Loris (2009): "Sparse and Stable Markowitz Portfolios," *Proceedings of the National Academy of Science*, 106, 12267–12272.

Callot, L., M. Caner, O. Onder, and E. Ulasan (2021): "A Nodewise Regression Approach to Estimating Large Portfolios," *Journal of Business and Economic Statistics*, 39, 520–531.

Carrasco, M. (2012): "A regularization approach to the many instruments problem," *Journal of Econometrics*, 170(2), 383–398.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Chapter 77 Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, vol. 6, pp. 5633–5751. Elsevier.

Carrasco, M., N. Kone, and N. Noumon (2019): "Optimal Portfolio Selection using Regularization," .

Carrasco, M., and B. Rossi (2016): "In-Sample Inference and Forecasting in Misspecified Factor Models," *Journal of Business & Economic Statistics*, 34(3), 313–338.

Chamberlain, G., and M. Rothschild (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51(5), 1281–1304, Publisher: [Wiley, Econometric Society].

Chang, J., S. X. Chen, and X. Chen (2015): "High dimensional generalized empirical likelihood for moment restrictions with dependent data," *Journal of Econometrics*, 185, 283–304.

Chernov, M., B. T. Kelly, S. Malamud, and J. Schwab (2025): "A Test of the Efficiency of a Given Portfolio in High Dimensions," Working Paper 33565, National Bureau of Economic Research.

Cochrane, J. H. (1996): "A cross-sectional test of an investment-based asset pricing," *Journal of Political Economy*, 104(3), 572.

——— (2005): *Asset pricing*. Princeton University Press, Princeton, N.J.

Daniel, K., and S. Titman (2012): "Testing Factor-Model Explanations of Market Anomalies," *Critical Finance Review*, 1(1), 103–139, Publisher: Now Publishers, Inc.

Davydov, Y. A. (1968): "Convergence of Distributions Generated by Stationary Stochastic Processes," *Theory of Probability & Its Applications*, 13(4), 691–696.

Didisheim, A., S. B. Ke, B. T. Kelly, and S. Malamud (2024): "APT or AIPT? The Surprising Dominance of Large Factor Models," Working Paper 33012, National Bureau of Economic Research.

Dittmar, R. F. (2002): "Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns," *The Journal of Finance*, 57(1), 369–403.

D′Hondt, C., R. De Winne, E. Ghysels, and S. Raymond (2020): "Artificial Intelligence Alter Egos: Who might benefit from robo-investing?," *Journal of Empirical Finance*, 59, 278–299.

Fama, E. F., and K. R. French (1992): "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47(2), 427–465.

——— (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33(1), 3–56.

——— (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116(1), 1–22.

Fan, J., Y. Fan, and J. Lv (2008): "High dimensional covariance matrix estimation using a factor model," *Journal of Econometrics*, 147(1), 186–197.

Fan, J., Y. Liao, and M. Mincheva (2013): "Large covariance estimation by thresholding principal orthogonal complements," *Journal of the Royal Statistical Society B*, 75, 603–680.

Ferson, W. (2019): "Empirical asset pricing: Models and methods," Publisher: MIT Press.

Francq, C., and J.-M. Zakoïan (2005): "A Central Limit Theorem for Mixing Triangular Arrays of Variables Whose Dependence Is Allowed to Grow with the Sample Size," *Econometric Theory*, 21(6), 1165–1171, Publisher: Cambridge University Press.

Gagliardini, P., E. Ossola, and O. Scaillet (2016): "Time-varying risk premium in large cross-sectional equity data sets," *Econometrica*, 84, 985–1046.

Gagliardini, P., and D. Ronchetti (2020): "Comparing Asset Pricing Models by the Conditional Hansen-Jagannathan Distance," *Journal of Financial Econometrics*, 18(2), 333–394.

Giglio, S., and D. Xiu (2021): "Asset Pricing with Omitted Factors," *Journal of Political Economy*, 129(7), 1947–1990.

Gospodinov, N., R. Kan, and C. Robotti (2013): "Chi-squared tests for evaluation and comparison of asset pricing models," *Journal of Econometrics*, 173(1), 108–125.

——— (2016): "On the properties of the constrained Hansen-Jagannathan distance," *Journal of Empirical Finance*, 36, 121–150.

Goyal, A. (2012): "Empirical cross-sectional asset pricing: a survey," *Financial Markets and Portfolio Management*, 26(1), 3–38.

Hall, A., and D. Pelletier (2011): "Nonnested testing in models estimated via generalized method of moments," *Econometric Theory*, 27, 443–456.

Hall, A. R., and A. Inoue (2003): "The large sample behaviour of the generalized method of moments estimator in misspecified models," *Journal of Econometrics*, 114(2), 361–394.

Hansen, L. P., J. Heaton, and E. G. J. Luttmer (1995): "Econometric Evaluation of Asset Pricing Models," *The Review of Financial Studies*, 8(2), 237–274, Publisher: Oxford Academic.

HANSEN, L. P., AND R. JAGANNATHAN (1997): "Assessing Specification Errors in Stochastic Discount Factor Models," *The Journal of Finance*, 52(2), 557–590.

HASTIE, T., A. MONTANARI, S. ROSSET, AND R. J. TIBSHIRANI (2022): "Surprises in high-dimensional ridgeless least squares interpolation," *The Annals of Statistics*, 50, 949Ů986.

HODRICK, R. J., AND X. ZHANG (2001): "Evaluating the specification errors of asset pricing models," *Journal of Financial Economics*, 62(2), 327–376.

HOU, K., H. MO, C. XUE, AND L. ZHANG (2021): "An Augmented q-Factor Model with Expected Growth*," *Review of Finance*, 25(1), 1–41.

KAN, R., AND C. ROBOTTI (2008): "Specification tests of asset pricing models using excess returns," *Journal of Empirical Finance*, 15(5), 816–838.

——— (2009): "Model Comparison Using the Hansen-Jagannathan Distance," *The Review of Financial Studies*, 22(9), 3449–3490, Publisher: Oxford Academic.

KLEIBERGEN, F., AND Z. ZHAN (2020): "Robust inference for consumption-based asset pricing," *Journal of Finance*, 75, 507–5502.

KORSAYE, S. A., A. QUAINI, AND F. TROJANI (2025): "Smart Stochastic Discount Factors," *Management Science*.

KOZAK, S., S. NAGEL, AND S. SANTOSH (2020): "Shrinking the cross-section," *Journal of Financial Economics*, 135, 271–292.

KRESS, R. (2014): *Linear Integral Equations*, Applied Mathematical Sciences. Springer-Verlag, New York, 3 edn.

KUDO, A. (1987): "An exact test for multiple inequality and equality constraints in the linear regression model," *Journal of the American Statistical Association*, 82, 782–793.

LEDOIT, O., AND M. WOLF (2003): "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *Journal of Empirical Finance*, 10(5), 603–621.

——— (2017): "Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks," *The Review of Financial Studies*, 30, 4349–4388.

LEWELLEN, J., S. NAGEL, AND J. SHANKEN (2010): "A skeptical appraisal of asset pricing tests," *Journal of Financial Economics*, 96(2), 175–194.

LI, J., AND Z. LIAO (2020): "Uniform nonparametric inference for time series," *Journal of Econometrics*, 219(1), 38–51.

LIAO, Y., X. MA, A. NEUHIERL, AND Z. SHI (2026): "Benign Overfitting in Economic Forecasting via Noise Regularization," Working paper.

LUDVIGSON, S. C. (2013): "Advances in Consumption-Based Asset Pricing: Empirical Tests," in *Handbook of the Economics of Finance*, vol. 2, pp. 799–906. Elsevier.

LUENBERGER, D. G. (1969): *Optimization by vector space methods*. N.Y., Wiley, N.Y.

LÖNN, R., AND P. SCHOTMAN (2024): "Empirical Asset Pricing with Many Test Assets," *Journal of Financial Econometrics*, 22, 1236–1263.

MANRESA, E., F. PANARANDA, AND E. SENTANA (2023): "Empirical evaluation of overspecified asset pricing models," *Journal of Financial Economics*, 147, 338–351.

NEWEY, W. K., AND R. J. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72(1), 219–255.

PESARAN, M. H., AND R. SMITH (2021): "Factor Strengths, Pricing Errors, and Estimation of Risk Premia," *CESifo working paper*.

——— (2023): "Arbitrage pricing theory, the stochastic discount factor and estimation of risk premia from portfolios," *Econometrics and Statistics*, 26, 17–30.

PHILLIPS, P. C. B., AND H. R. MOON (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67(5), 1057–1111.

QIU, C., AND T. OTSU (2022): "Information theoretic approach to high-dimensional multiplicative models: Stochastic discount factor and treatment effect," *Quantitative Economics*, 13, 63–94.

RAPONI, V., C. ROBOTTI, AND P. ZAFFARONI (2020): "Testing Beta-Pricing Models Using Large Cross-Sections," *The Review of Financial Studies*, 33(6), 2796–2842.

REN, Y., AND K. SHIMOTSU (2009): "Improvement in finite sample properties of the HansenŰ-Jagannathan distance test," *Journal of Empirical Finance*, 16(3), 483–506.

RIO, E. (1993): "Covariance inequalities for strongly mixing processes," in *Annales de l'IHP Probabilités et statistiques*, vol. 29, pp. 587–597. Issue: 4.

SCHENNACH, S., AND D. WILHELM (2017): "A Simple Parametric Model Selection Test," *Journal of the American Statistical Association*, 112, 1663–1674.

SHANKEN, J. (1985): "Multivariate tests of the zero-beta CAPM," *Journal of Financial Economics*, 14(3), 327–348.

Shanken, J., and G. Zhou (2007): "Estimating and testing beta pricing models: Alternative methods and their performance in simulations," *Journal of Financial Economics*, 84(1), 40–86.

Shi, X. (2015): "A nondegenerate Vuong test," *Quantitative Economics*, 6, 85–121.

Vuong, Q. (1989): "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," *Econometrica*, 57, 307–333.

Wolak, F. A. (1963): "A multivariate analogue of the one-sided test," *Biometrika*, 50, 403–418.

——— (1989): "Testing inequality constraints in linear econometric models," *Journal of Econometrics*, 41(2), 205–235.

11 Supplementary Appendix for "Hansen-Jagannathan distance with many assets" by Marine Carrasco and Cheikh Nokho

11.1 Proof of lemma 1

1. We have

$$\begin{aligned} V_{22} &= E\left[(r_t-\mu_2)(r_t-\mu_2)'\right] \\ &= E\left[\left(\beta\tilde{f}_t+\epsilon_t\right)\left(\beta\tilde{f}_t+\epsilon_t\right)'\right] \\ &= \beta E\left(\tilde{f}_t\tilde{f}_t'\right)\beta'+E\left(\epsilon_t\epsilon_t'\right). \end{aligned}$$

Therefore,

$$tr(\Sigma)=tr\left(\frac{\beta V_{11}\beta'}{N}+\frac{\Sigma_\epsilon}{N}\right).$$

From Assumption 1, $tr(\Sigma_\epsilon)/N=O(1)$, $tr(\beta'\beta)/N=tr(\beta\beta')/N=\sum_i \beta_i'\beta_i/N=O(1)$. Hence,

$$\frac{tr(\beta V_{11}\beta')}{N}=\frac{tr(V_{11}\beta'\beta)}{N}\leq\frac{|trV_{11}|\,|tr(\beta'\beta)|}{N}$$

where we used the fact $|tr(AB)|\leq trA\,trB$ when A and B are positive semi-definite matrices (see Bernstein (2009)). As a result $tr(\Sigma)=O(1)$.

2. We have

$$E\|r_t\|_N^2=E(r_t'r_t)/N=tr(\Sigma)+\mu_2'\mu_2/N.$$

Using $\mu_2=\alpha+\beta\lambda$, we obtain

$$\frac{\mu_2'\mu_2}{N}=\frac{\alpha'\alpha}{N}+2\frac{(\beta\lambda)'\alpha}{N}+\frac{\lambda'\beta'\beta\lambda}{N}.$$

$$\frac{\lambda'\beta'\beta\lambda}{N}=\frac{tr(\beta'\beta\lambda\lambda')}{N}\leq\frac{tr(\beta'\beta)}{N}tr(\lambda\lambda')=O(1)$$

by Assumption 1. Moreover, using Cauchy-Schwarz

$$
\begin{aligned}
\frac{\left|(\beta\lambda)'\alpha\right|}{N} &\leq \sqrt{\frac{(\beta\lambda)'(\beta\lambda)}{N}}\sqrt{\frac{\alpha'\alpha}{N}} \\
&= \sqrt{\frac{tr\left(\beta'\beta\lambda\lambda'\right)}{N}}\sqrt{\frac{\alpha'\alpha}{N}} \\
&\leq \sqrt{\frac{tr\left(\beta'\beta\right)}{N}}\sqrt{tr\left(\lambda\lambda'\right)}\sqrt{\frac{\alpha'\alpha}{N}} \\
&= O\left(1\right).
\end{aligned}
$$

Therefore, $\mu_2'\mu_2/N = O(1)$ and hence $E\left\|r_t\right\|_N^2 = O(1)$.

11.2 Proof of Proposition 1

We transform the primal problem to be able to use the Fenchel-Rockafellar Duality, see Chapter 15 of Bauschke and Combettes (2017) or Borwein and Lewis (1992) as well as Korsaye, Quaini, and Trojani (2025).

Let X be the $(N+1)$ vector defined as $X = \begin{bmatrix} \frac{2r'}{N} & 2 \end{bmatrix}'$.

Let $f_y : L^2 \to \mathbb{R}$ be the function defined by $f_y(x) = E[(x-y)^2]$ and $A : L^2 \to \mathbb{R}^{N+1}$ be the operator such that $A(m) = E[mX]$.

Let $g : \mathbb{R}^{N+1} \to (-\infty, +\infty]$ be defined by $g(x) = h(x_1) + \chi_{\{2\}}(x_{-1})$ where $x = (x_1', x_{-1})' \in \mathbb{R}^N \times \mathbb{R}$, $\chi_{\{2\}}$ is

$$
\chi_{\{2\}}(x) = \begin{cases} 0 & \text{if } x = 2, \\ \infty & \text{otherwise,} \end{cases}
$$

and $h(x) = \frac{N}{4\gamma}\left\|x\right\|^2$.

First, we prove the result for Ridge regularization. Problem (6) can be rewritten as

$$
\delta_R^2 = \inf_{m \in L^2}\left\{f_y(m) + g(A(m))\right\}.
$$

The functions f_y and g are convex as $x \mapsto x^2$ is convex. Moreover, A is a linear bounded function. From Theorem 4.2 of Borwein and Lewis (1992), strong duality holds if $(ri\, dom(g)) \cap (ri\, A(dom(f_y))) \neq \emptyset$.[6]

To check this condition, it suffices to find an element of $(ri\, dom(g)) \cap (ri\, A(dom(f_y)))$. Let $m_0 \in L^2$ be a candidate SDF[7] with finite pricing error, i.e. $E(m_0) = 1$ and $\| E[m_0 r] \|_N^2 < \infty$.

[6] For a convex set $S \subseteq \mathbb{R}^N$, $ri\, S$ is its relative interior. The latter is the interior with respect to the affine hull of S, $aff\, S$. Specifically, $ri\, S = \{x \in S : B_\epsilon(x) \cap aff\, S \subseteq S\}$, where $aff\, S = \{\theta_1 x_1 + \cdots + \theta_k x_k : x_1, \ldots, x_k \in S, \theta_1 + \cdots + \theta_k = 1\}$ and $B_\epsilon(x) = \left\{y \in \mathbb{R}^N : \| y - x \| < \epsilon\right\}$.

[7] Such a random variable exists, for instance $m_0 = 1$ satisfy the conditions.

As $m_0 \in L^2$ and $E(m_0 - y)^2 < \infty$, then $m_0 \in dom(f_y)$ and $A(m_0) \in ri\, A(dom(f_y))$. In addition, because $E\,[2m_0] = 2$ and $\parallel E[m_0 r] \parallel_N^2 < \infty$, $g(A(m_0)) = \frac{N}{4\gamma} \parallel E[m_0 \frac{2r}{N}] \parallel^2 = \frac{1}{\gamma} \parallel E[m_0 r] \parallel_N^2 < \infty$, therefore $A(m_0) \in ri\, dom(g)$. In conclusion, $A(m_0)$ is an element of $(ri\, dom(g)) \cap (ri\, A(dom(f_y)))$ and hence $(ri\, dom(g)) \cap (ri\, A(dom(f_y))) \neq \emptyset$.

Then by Theorem 4.2 of Borwein and Lewis (1992), we have[8]

$$\delta_R^2 = \max_{\nu \in \mathbb{R}^{N+1}} \left\{ -f_y^*(A^*(-\nu)) - g^*(\nu) \right\},$$

where f_y^* and g^* are the conjugate functions of f_y and g respectively and A^* is the adjoint of A.

Let us determine the relevant conjugate functions.

$f_y^*(z) = E\left\{ \underset{w \in L^2}{sup}\, zw - (w-y)^2 \right\} = E\left\{ zy + \frac{1}{4} z^2 \right\}$[9], $A^* : \mathbb{R}^{N+1} \to L^2$ and $A^*(\theta) = X'\theta$. Finally $g^*(\nu) = h^*(\nu_1) + \chi^*_{\{2\}}(\nu_2)$ where $\nu = (\nu_1', \nu_2) \in \mathbb{R}^{N+1}$. The conjugates[10] are given by $\chi^*_{\{2\}}(\nu_2) = \underset{x \in \{2\}}{sup}\, x\nu_2 = 2\nu_2$ and $h^*(\nu_1) = \frac{\gamma}{N} \|\nu_1\|^2$. So, $g^*(\nu) = 2\nu_2 + \frac{\gamma}{N} \|\nu_1\|^2$.

Therefore

$$\delta_R^2 = \max_{\nu_1 \in \mathbb{R}^N, \nu_2 \in \mathbb{R}} E\left\{ 2y\nu_1' \frac{r}{N} + 2y\nu_2 - \frac{\nu_1' r r' \nu_1}{N^2} - \nu_2^2 - 2\frac{\nu_1' r \nu_2}{N} - 2\nu_2 - \frac{\gamma}{N} \parallel \nu_1 \parallel^2 \right\}.$$

Now, using the fact that $E[y] = 1$, we obtain

$$\delta_R^2 = \max_{\nu_1 \in \mathbb{R}^N, \nu_2 \in \mathbb{R}} E\left\{ 2y\nu_1' \frac{r}{N} - \frac{\nu_1' r r' \nu_1}{N^2} - \nu_2^2 - 2\frac{\nu_1' r \nu_2}{N} - \frac{\gamma}{N} \parallel \nu_1 \parallel^2 \right\}.$$

Solving in ν_1 and ν_2 gives

$$\nu_1 = (\Sigma + \gamma I)^{-1} \alpha \text{ and } \nu_2 = -\nu_1' \mu_2 / N,$$

and hence $\delta_R^2 = \alpha' \left(\Sigma + \gamma I\right)^{-1} \alpha$.

We can do the same for Tikhonov by setting $h(x) = \frac{N}{4\gamma} \|x\|_\Sigma^2$. The latter can be rewritten as $h(x) = \frac{N}{2\gamma} n(x)$ with $n(x) = \frac{1}{2} \|x\|_\Sigma^2$. Therefore the convex conjugate of h is $h^*(z) = \frac{N}{2\gamma} n^* \left(\frac{z}{\frac{N}{2\gamma}} \right)$ with

$$n^*(z) = \sup_{w \in R^N} \left\{ w' z - \frac{w' \Sigma w}{2} \right\}.$$

[8] For convenience, we replaced ν by $-\nu$ in Borwein and Lewis' formula. This does affect the final result.

[9] This comes from the definition of the functional conjugate of a convex function p.196 of Luenberger (1969) and the use of Riesz Theorem in the L^2 space equipped with the usual inner product.

[10] To determine the conjugate of h, note that the conjugate of $\frac{1}{2} \|x\|^2$ is again $\frac{1}{2} \|x\|^2$. In addition, if $f(x) = ag(x) + b$, then $f^*(x) = ag^* \left(\frac{x}{a} \right) + b$.

The expression in brackets is maximized at $w = \Sigma^{-1}z$. Therefore, $n^*(z) = \frac{z'\Sigma^{-1}z}{2}$. As a result, $h^*(z) = \frac{\gamma}{N}\|z\|^2_{\Sigma^{-1}}$.

11.3 Proof of Proposition 2

The proof of Proposition 2 uses the following lemmas.

Lemma 2. *Suppose Assumptions 1 and 2 are satisfied. Then,* $\left\|\epsilon'\bar{F}\right\|_F^2 = O_p(NT)$, *where* ϵ is a $T \times N$ *matrix with (t,i) element* ϵ_{ti}.

Proof of Lemma 2.
Let us denote $Y_{\epsilon f,t} = \epsilon_t \bar{f}_t'$.
First,

$$
\begin{aligned}
E\left[\left\|\frac{\epsilon'\bar{F}}{T}\right\|_F^2\right] &= E\left[tr\left\{\left(\frac{1}{T}\sum_{t=1}^{T}\epsilon_t\bar{f}_t'\right)'\left(\frac{1}{T}\sum_{t=1}^{T}\epsilon_t\bar{f}_t'\right)\right\}\right] \\
&= E\left[tr\left\{\left(\frac{1}{T^2}\sum_{t=1}^{T}Y_{\epsilon f,t}'Y_{\epsilon f,t} + \frac{1}{T^2}\sum_{t\neq s}^{T}Y_{\epsilon f,t}'Y_{\epsilon f,s}\right)\right\}\right] \\
&= \frac{1}{T}E\left[tr(Y_{\epsilon f,1}'Y_{\epsilon f,1})\right] + \frac{2}{T}\sum_{l=1}^{T}(1-\frac{l}{T})E\left[tr(Y_{\epsilon f,1}'Y_{\epsilon f,1+l})\right]
\end{aligned}
$$

We have

$$
\begin{aligned}
trE[Y_{\epsilon f,t}'Y_{\epsilon f,t}] &= E[tr(\bar{f}_t\epsilon_t'\epsilon_t\bar{f}_t')] \\
&= trE[\bar{f}_t'\bar{f}_t\epsilon_t'\epsilon_t]
\end{aligned}
$$

From Cauchy-Schwarz, $\left|E[\bar{f}_t'\bar{f}_t\epsilon_t'\epsilon_t]\right| \leq \sqrt{E[\left\|\bar{f}_t\right\|^4]E[\|\epsilon_t\|^4]} = O(N)$. Therefore, $\frac{1}{T}E\left[tr(Y_{\epsilon f,1}'Y_{\epsilon f,1})\right] = O(\frac{N}{T})$.

Using Davydov's inequality (Davydov (1968), Rio (1993)) [11] (with $q = r = 2 + \rho$),

$$
\begin{aligned}
trE\left[(Y_{\epsilon f,1}^{'} Y_{\epsilon f,1+l})\right] &= \sum_{i=1}^{N}\sum_{k=1}^{K} E\left[\left(\bar{f}_{k1}\epsilon_{i1}\right)\left(\bar{f}_{k1+l}\epsilon_{i1+l}\right)\right] \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} cov(\bar{f}_{k1}\epsilon_{i1}, \bar{f}_{k1+l}\epsilon_{i1+l}) \\
&\leq 12\sum_{i=1}^{N}\sum_{k=1}^{K} \alpha_x(l)^{\frac{\rho}{2+\rho}} E[(\bar{f}_{kt}\epsilon_{it})^{2+\rho}]^{\frac{2}{2+\rho}}.
\end{aligned}
$$

As a result,

$$
\begin{aligned}
\frac{2}{T}\sum_{l=1}^{T}(1-\frac{l}{T})E\left[tr(Y_{\epsilon f,1}^{'} Y_{\epsilon f,1+l})\right] &\leq \frac{24}{T}\sum_{i=1}^{N}\sum_{k=1}^{K} E[(\bar{f}_{kt}\epsilon_{it})^{2+\rho}]^{\frac{2}{2+\rho}} \sum_{l=1}^{T}(1-\frac{l}{T})\alpha_x(l)^{\frac{\rho}{2+\rho}} \\
&\leq \frac{24}{T}\sum_{i=1}^{N}\sum_{k=1}^{K} E[(\bar{f}_{kt}\epsilon_{it})^{2+\rho}]^{\frac{2}{2+\rho}} \sum_{l=1}^{T}\alpha_x(l)^{\frac{\rho}{2+\rho}}.
\end{aligned}
$$

From Assumption 2(iii) and Cauchy-Schwarz, $\left|E[(\bar{f}_{kt}\epsilon_{it})^{2+\rho}]\right| \leq E[\bar{f}_{kt}^{4+2\rho}]^{\frac{1}{2}} E[\epsilon_{it}^{4+2\rho}]^{\frac{1}{2}} \leq c^{\frac{1}{2}} E[\bar{f}_{kt}^{4+2\rho}]^{\frac{1}{2}}$. So,

$$
\frac{2}{T}\sum_{l=1}^{T}(1-\frac{l}{T})E\left[tr(Y_{\epsilon f,1}^{'} Y_{\epsilon f,1+l})\right] = O(\frac{N}{T}).
$$

Hence, $E\left[\left\|\frac{\epsilon'\bar{F}}{T}\right\|_F^2\right] = O(\frac{N}{T})$. In conclusion,

$$
\left\|\epsilon'\bar{F}\right\|_F^2 = O_p(NT).
$$

Lemma 3. *Suppose Assumptions 1 and 2 are satisfied. Then,* $\left\|\hat{\beta} - \beta\right\|_F^2 = O_p\left(\frac{N}{T}\right)$.

[11] For any positive real numbers p, q, r such that $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$, the covariance between two r.vs X and Y is bounded as follows: $cov(X,Y) \leq 12\alpha(\sigma(X),\sigma(Y))^{\frac{1}{p}} E[\mid X \mid^q]^{\frac{1}{q}} E[\mid Y \mid^r]^{\frac{1}{r}}$, where $\sigma(X)$ is the sigma algebra generated by X and α is the strong mixing coefficient.

Proof of Lemma 3. Using the fact that $\left\|\epsilon'\bar{F}\right\|_F^2 = O_p(NT)$, we have

$$\begin{aligned}
\left\|\hat{\beta}-\beta\right\|_F^2 &= \left\|\frac{\epsilon'\bar{F}}{T}\hat{V}_{11}^{-1}\right\|_F^2 \\
&= \left\|\frac{\epsilon'\bar{F}}{T}V_{11}^{-1}+\frac{\epsilon'\bar{F}}{T}(\hat{V}_{11}^{-1}-V_{11}^{-1})\right\|_F^2 \\
&\leq 2\left\|\frac{\epsilon'\bar{F}}{T}V_{11}^{-1}\right\|_F^2+2\left\|\frac{\epsilon'\bar{F}}{T}(\hat{V}_{11}^{-1}-V_{11}^{-1})\right\|_F^2 \\
&\leq 2\left\|\frac{\epsilon'\bar{F}}{T}V_{11}^{-1}\right\|_F^2+2\left\|\frac{\epsilon'\bar{F}}{T}\right\|_F^2\cdot\left\|\hat{V}_{11}^{-1}-V_{11}^{-1}\right\|_F^2 \\
&= O_p\left(\frac{NT}{T^2}\right)+O_p\left(\frac{NT}{T^2}\cdot\frac{1}{T}\right) \\
&= O\left(\frac{N}{T}\right).
\end{aligned}$$

Therefore,

$$\left\|\hat{\beta}-\beta\right\|_F^2 = O_p\left(\frac{N}{T}\right)$$

Lemma 4. *Suppose Assumptions 1 and 2 are satisfied. For* $k = 1, \cdots, K$, $\left\|\hat{\beta}_k-\beta_k\right\|_N = O_p\left(\frac{1}{\sqrt{T}}\right)$.

Proof of Lemma 4. As $\left\|\hat{\beta}-\beta\right\|_F^2 = O_p\left(\frac{N}{T}\right)$, we have $\frac{1}{N}tr\left[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)\right] = O_p\left(\frac{1}{T}\right)$.

$$(\hat{\beta}-\beta)(\hat{\beta}-\beta)' = \sum_{k=1}^{K}((\hat{\beta}_k-\beta_k)(\hat{\beta}_k-\beta_k)').$$

As a result,

$$\frac{1}{N}tr\left[\sum_{k=1}^{K}((\hat{\beta}_k-\beta_k)(\hat{\beta}_k-\beta_k)')\right] = \sum_{k=1}^{K}\parallel\hat{\beta}_k-\beta_k\parallel_N^2 = O_p(\frac{1}{T})$$

In conclusion, $\|\hat{\beta}_k-\beta_k\|_N = O_p\left(\frac{1}{\sqrt{T}}\right)$.

Lemma 5. *Suppose Assumptions 1, 2, and 4(i), we have*

$$\left\|\widehat{\Sigma}-\Sigma\right\|_{HS} = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Proof of Lemma 5. $\widehat{\Sigma}$ and Σ are elements of the Hilbert space of Hilbert-Schmidt operators with inner product $\langle A, B\rangle = tr\,(A'B)$ and norm $\|A\|_{HS}^2 = tr\,(A'A)$.

Using the notations $\tilde{r}_t = r_t - \mu_2$ and $X_t = \tilde{r}_t\tilde{r}_t'/N - \Sigma$, we have by Davydov's inequality and the mixing propriety of r_t

$$\begin{aligned} E\left[\left\|\frac{1}{T}\sum_{t=1}^{T} X_t\right\|_{HS}^2\right] &= \frac{1}{T}E\,\|X_1\|_{HS}^2 + \frac{2}{T}\sum_{l=1}^{T} E\,\langle X_1, X_{1+l}\rangle \\ &\leq \frac{1}{T}E\,\|X_1\|_{HS}^2 + \frac{24}{T}\left(E\left(\|X_1\|_{HS}^{2+\rho}\right)\right)^{\frac{2}{2+\rho}}\sum_{l=1}^{T}\left(1-\frac{l}{T}\right)\alpha\,(l)^{\frac{\rho}{2+\rho}}. \end{aligned}$$

Under the mixing property $\sum_l \alpha\,(l)^{\frac{\rho}{2+\rho}} < \infty$ and under the assumption 4(i), we have $E\left(\|X_1\|_{HS}^{2+\rho}\right) < \infty$. Moreover, $\widehat{\Sigma} - \Sigma = \sum_{t=1}^{T} X_t/T + o_p\,(1/T)$. It follows that $\left\|\widehat{\Sigma} - \Sigma\right\|_{HS}^2 = O_p\,(1/T)$ by Markov's inequality.

Lemma 6. *Suppose that Assumptions 1 to 3. If* $\beta_k \in \mathcal{R}\left(\Sigma^{1+\omega}\right)$ *for some* $1 \geq \omega > 0$ *(instead of the stronger condition* $\beta_k \in \mathcal{R}\left(\Sigma^2\right)$*), we have the following result:*

$$\left\|\hat{\Sigma}_\gamma^{-1}\hat{\beta}_k - \Sigma^{-1}\beta_k\right\|_N^2 = O_p\left(\frac{1}{\gamma T}\right) + O(\gamma^\omega).$$

Proof of Lemma 6. We consider Tikhonov only. We follow the proof of Lemma 3 of Carrasco (2012). We have the following decomposition.

$$\begin{aligned} \left\|\hat{\Sigma}_\gamma^{-1}\hat{\beta}_k - \Sigma^{-1}\beta_k\right\|_N^2 &\leq 3\left\|\hat{\Sigma}_\gamma^{-1}\left(\hat{\beta}_k - \beta_k\right)\right\|_N^2 & (22)\\ &\quad +3\left\|\left(\hat{\Sigma}_\gamma^{-1} - \Sigma_\gamma^{-1}\right)\beta_k\right\|_N^2 & (23)\\ &\quad +3\left\|\left(\Sigma_\gamma^{-1} - \Sigma^{-1}\right)\beta_k\right\|_N^2 & (24) \end{aligned}$$

First, consider (22). Let $v = \hat{\beta}_k - \beta_k$. So,

$$\begin{aligned} \left\|\widehat{\Sigma}_\gamma^{-1}v\right\|_N^2 &= \left\|\left(\widehat{\Sigma}^2 + \gamma I\right)^{-1}\widehat{\Sigma}v\right\|_N^2 \\ &= \sum_{j=1}^{\infty}\frac{\widehat{\lambda}_j^2}{\left(\widehat{\lambda}_j^2 + \gamma\right)^2}\left\langle\hat{\phi}_j, v\right\rangle_N^2 \\ &\leq \frac{1}{\gamma}\|v\|_N^2 = O_p\left(\frac{1}{\gamma T}\right) \end{aligned}$$

where we used $\|v\|_N^2 = \left\|\hat{\beta}_k - \beta_k\right\|_N^2 = O_p\left(1/T\right)$ by Lemma 4. Hence (22) is $O_p\left(\frac{1}{\gamma T}\right)$.

For (23), we will use the relation

$$
\begin{aligned}
& \left(\hat{\Sigma}^2 + \gamma I\right)^{-1} \hat{\Sigma} - \left(\Sigma^2 + \gamma I\right)^{-1} \Sigma \\
= \; & \left(\hat{\Sigma}^2 + \gamma I\right)^{-1} \left[\hat{\Sigma}\left(\Sigma - \hat{\Sigma}\right)\Sigma + \gamma\left(\hat{\Sigma} - \Sigma\right)\right]\left(\Sigma^2 + \gamma I\right)^{-1}
\end{aligned}
$$

As $\beta_k \in \mathcal{R}\left(\Sigma^{1+\omega}\right)$, there exists ϕ such that $\|\phi\|_N < \infty$ and $\beta_k = \Sigma^{1+\omega}\phi$. We have

$$
\begin{aligned}
\left\|\left(\hat{\Sigma}_\gamma^{-1} - \Sigma_\gamma^{-1}\right)\beta_k\right\|_N^2 \; \leq \; & \\
& \left\|\left(\hat{\Sigma}^2 + \gamma I\right)^{-1} \hat{\Sigma}\left(\Sigma - \hat{\Sigma}\right)\Sigma\left(\Sigma^2 + \gamma I\right)^{-1}\beta_k\right\|_N^2 \qquad (25) \\
& + \left\|\left(\hat{\Sigma}^2 + \gamma I\right)^{-1} \gamma\left(\hat{\Sigma} - \Sigma\right)\left(\Sigma^2 + \gamma I\right)^{-1}\beta_k\right\|_N^2 \qquad (26)
\end{aligned}
$$

Regarding (25), we have

$$
\begin{aligned}
& \left\|\left(\hat{\Sigma}^2 + \gamma I\right)^{-1} \hat{\Sigma}\left(\Sigma - \hat{\Sigma}\right)\Sigma\left(\Sigma^2 + \gamma I\right)^{-1}\beta_k\right\|_N^2 \\
\leq \; & \left\|\left(\hat{\Sigma}^2 + \gamma I\right)^{-1} \hat{\Sigma}\right\|_{op,N}^2 \left\|\Sigma - \hat{\Sigma}\right\|_{op,N}^2 \left\|\Sigma\left(\Sigma^2 + \gamma I\right)^{-1}\beta_k\right\|_N^2 \\
= \; & O_p\left(\frac{1}{\gamma T}\right)
\end{aligned}
$$

because $\left\|\Sigma - \hat{\Sigma}\right\|_{op,N}^2 = O_p\left(1/T\right)$.

Regarding (26), we have

$$
\begin{aligned}
\left\|\left(\hat{\Sigma}^2 + \gamma I\right)^{-1} \gamma\left(\hat{\Sigma} - \Sigma\right)\left(\Sigma^2 + \gamma I\right)^{-1}\beta_k\right\|_N^2 \; & \leq \; \left\|\left(\hat{\Sigma} - \Sigma\right)\left(\Sigma^2 + \gamma I\right)^{-1}\beta_k\right\|_N^2 \\
& = \; O_p\left(\frac{1}{\gamma T}\right).
\end{aligned}
$$

In summary, the rate of (23) is $O_p\left(\frac{1}{\gamma T}\right)$.

Finally, the term (24) satisfies

$$
\begin{aligned}
& \left\|\left(\Sigma_\gamma^{-1}-\Sigma^{-1}\right)\beta_k\right\|_N^2 \\
= & \left\|\left(\Sigma_\gamma^{-1}-\Sigma^{-1}\right)\Sigma^{1+\omega}\phi\right\|_N^2 \\
= & \left\|\left(\left(\Sigma^2+\gamma I\right)^{-1}\Sigma-\Sigma^{-1}\right)\Sigma\Sigma^{\omega}\phi\right\|_N^2 \\
= & \left\|\left(\left(\Sigma^2+\gamma I\right)^{-1}\Sigma^2-I\right)\Sigma^{\omega}\phi\right\|_N^2 \\
= & \left\|\left(\Sigma^2+\gamma I\right)^{-1}\left(\Sigma^2-\Sigma^2-\gamma\right)\Sigma^{\omega}\phi\right\|_N^2 \\
= & \gamma^2\left\|\left(\Sigma^2+\gamma I\right)^{-1}\Sigma^{\omega}\phi\right\|_N^2 \\
= & \gamma^2\left\|\left(\Sigma^2+\gamma I\right)^{-1+\omega/2}\left(\Sigma^2+\gamma I\right)^{-\omega/2}\Sigma^{\omega}\phi\right\|_N^2 \\
= & O\left(\gamma^{\omega}\right).
\end{aligned}
$$

This completes the proof of Lemma 6.

Lemma 7. *Let*

$$X_{T,N}=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}<\tilde{r}_t,u>_N=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_{t,T,N},$$

where $u\in\mathbb{R}^N$ is not random and $\|u\|_N=O(1)$. If Assumptions 2(i) and 4(i) hold, then

$$X_{T,N}\xrightarrow{d}\mathcal{N}(0,\omega^2)$$

when T,N go simultaneously to ∞, where $\omega^2=\lim\limits_{N,T\to\infty} var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_{t,T,N})>0$.

Proof of Lemma 7.

First, consider the case when $\{r_t\}_{t=1,\cdots,T}$ are independent. From Assumptions 4(iii), $E\left[\| r_t \|_N^{2+\rho}\right]=O(1)$ when N goes to infinity for $\rho>0$. To establish the central limit theorem, we need to verify the Lindeberg condition for a double indexed process of Phillips and Moon (1999)(see their Theorem 2). In our setting, this condition can be rewritten as

$$\lim_{N,T\to\infty}\frac{1}{\sigma_{T,N}^2}\sum_{t=1}^{T}E\left[Y_{t,T,N}^2 1_{|Y_{t,T,N}|>\sigma_{T,N}\varepsilon}\right]\to 0$$

for every $\varepsilon>0$, with $\sigma_{T,N}^2=T.var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_{t,T,N}\right)=T.\frac{u'\Sigma u}{N}\equiv T\sigma_N^2>0$.

To see that this condition is satisfied, observe that when $\left|\frac{Y_{t,T,N}}{\sigma_{T,N}\varepsilon}\right| > 1$,

$$\varepsilon^{\rho}\frac{Y_{t,T,N}^2}{\sigma_{T,N}^2} \leq \frac{Y_{t,T,N}^{2+\rho}}{\sigma_{T,N}^{2+\rho}}.$$

Therefore,

$$\varepsilon^{\rho}E\left[\frac{Y_{t,T,N}^2}{\sigma_{T,N}^2}1_{|Y_{t,T,N}|>\sigma_{T,N}\varepsilon}\right] \leq E\left[\frac{Y_{t,T,N}^{2+\rho}}{\sigma_{T,N}^{2+\rho}}1_{|Y_{t,T,N}|>\sigma_{T,N}\varepsilon}\right] \leq E\left[\frac{Y_{t,T,N}^{2+\rho}}{\sigma_{T,N}^{2+\rho}}\right].$$

Moreover

$$\lim_{N,T\to\infty}\sum_{t=1}^{T}E\left[\frac{Y_{t,T,N}^{2+\rho}}{\sigma_{T,N}^{2+\rho}}\right] = \lim_{N,T\to\infty}\frac{1}{T^{\rho/2}}\frac{1}{\sigma_N^{2+\rho}}E\left[< \tilde{r}_t, u >_N{}^{2+\rho}\right] = 0$$

as $\sigma_N = O(1)$, $E\left[|\langle\tilde{r}_t, u\rangle_N|^{2+\rho}\right] \leq E\left[\|\tilde{r}_t\|_N^{2+\rho}\right]\|u\|_N^{2+\rho} = O(1)$ by Assumption 4(iii). Hence, the Lindeberg condition is satisfied.

For the dependent case**,** by Davydov's inequality (Davydov (1968), Rio (1993)) (with $q = r = 2+\rho$), we have

$$\begin{aligned} var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_{t,T,N}) &= E[Y_{1,T,N}^2] + 2\sum_{l=1}^{T}(1-\frac{l}{T})E\left[Y_{1,T,N}Y_{1+l,T,N}\right] \\ &\leq E[Y_{1,T,N}^2] + 24\left(E[|Y_{t,T,N}|^{2+\rho}]\right)^{\frac{2}{2+\rho}}\sum_{l=1}^{T}(1-\frac{l}{T})\alpha_x(l)^{\frac{\rho}{2+\rho}} \\ &\leq E[Y_{1,T,N}^2] + 24\left(E[|Y_{t,T,N}|^{2+\rho}]\right)^{\frac{2}{2+\rho}}\sum_{l=1}^{T}\alpha_x(l)^{\frac{\rho}{2+\rho}}. \end{aligned}$$

As a result, $0 < \underset{N,T\to\infty}{lim}\ var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_{t,T,N}\right) < \infty$. In addition, the central limit theorem of Francq and Zakoïan (2005) applies because the Lindeberg condition (iii) of page 1168 still applies when N goes to infinity, see also Chang, Chen, and Chen (2015) p.288 for a similar application of this result. Hence, $X_{T,N}$ asymptotically converges to a normal distribution when T, N go to ∞.

Lemma 8. *Suppose Assumptions 2(i) and 4 are satisfied. For any $u, v \in \mathbb{R}^N$ with $\|u\|_N < \infty$ and $\|v\|_N < \infty$, as N and T go to ∞, if*

$$\underset{N,T\to\infty}{lim}\ var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\langle\tilde{r}_t, v\rangle_N\langle\tilde{r}_t, u\rangle_N - E\left(\langle\tilde{r}_t, v\rangle_N\langle\tilde{r}_t, u\rangle_N\right)\right)\right] \equiv \sigma_{u,v}^2 > 0,$$

then $\sqrt{T}\left\langle\left(\hat{\Sigma}-\Sigma\right)v,u\right\rangle_N$ *converges to a Gaussian distribution of mean* 0 *and variance* $\sigma^2_{u,v}$.

Proof of Lemma 8. We have the following decomposition of $\hat{\Sigma}-\Sigma$:

$$\begin{aligned}
\hat{\Sigma}-\Sigma &= \frac{1}{NT}\sum_{t=1}^{T}\left(r_t-\hat{\mu}_2\right)\left(r_t-\hat{\mu}_2\right)^{'}-\frac{1}{N}E[\tilde{r}_t\tilde{r}_t^{'}] \\
&= \frac{1}{NT}\sum_{t=1}^{T}\left(r_t-\mu_2+\mu_2-\hat{\mu}_2\right)\left(r_t-\mu_2+\mu_2-\hat{\mu}_2\right)^{'}-\frac{1}{N}E[\tilde{r}_t\tilde{r}_t^{'}] \\
&= \frac{1}{NT}\sum_{t=1}^{T}\left(\tilde{r}_t\tilde{r}_t^{'}-E[\tilde{r}_t\tilde{r}_t^{'}]\right)+\left(\mu_2-\hat{\mu}_2\right)\frac{1}{NT}\sum_{t=1}^{T}\left(r_t-\mu_2\right)^{'} \\
&\quad +\frac{1}{NT}\sum_{t=1}^{T}\left(r_t-\mu_2\right)\left(\mu_2-\hat{\mu}_2\right)^{'} \\
&\quad +\frac{1}{N}\left(\mu_2-\hat{\mu}_2\right)\left(\mu_2-\hat{\mu}_2\right)^{'}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left\langle\sqrt{T}\left[\hat{\Sigma}-\Sigma\right]v,u\right\rangle_N &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\left\langle\tilde{r}_t,v\right\rangle_N\left\langle\tilde{r}_t,u\right\rangle_N-E\left(\left\langle\tilde{r}_t,v\right\rangle_N\left\langle\tilde{r}_t,u\right\rangle_N\right)\right\} \\
&\quad +\left\langle v,\left(\mu_2-\hat{\mu}_2\right)\right\rangle_N\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\langle\tilde{r}_t,u\right\rangle_N \\
&\quad +\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\langle v,\tilde{r}_t\right\rangle_N\left\langle\left(\mu_2-\hat{\mu}_2\right),u\right\rangle_N \\
&\quad +\sqrt{T}\left\langle v,\left(\mu_2-\hat{\mu}_2\right)\right\rangle_N\left\langle\left(\mu_2-\hat{\mu}_2\right),u\right\rangle_N.
\end{aligned}$$

Following Lemma 7, $\left\langle v,\left(\mu_2-\hat{\mu}_2\right)\right\rangle_N\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\langle\tilde{r}_t,u\right\rangle_N=O_p\left(\frac{1}{\sqrt{T}}\right)$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\langle v,\tilde{r}_t\right\rangle_N\left\langle\left(\mu_2-\hat{\mu}_2\right),u\right\rangle_N=O_p\left(\frac{1}{\sqrt{T}}\right)$, and $\sqrt{T}\left\langle v,\left(\mu_2-\hat{\mu}_2\right)\right\rangle_N\left\langle\left(\mu_2-\hat{\mu}_2\right),u\right\rangle_N=O_p\left(\frac{1}{\sqrt{T}}\right)$. As a result,

$$\left\langle\sqrt{T}\left[\hat{\Sigma}-\Sigma\right]v,u\right\rangle_N=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\left\langle\tilde{r}_t,v\right\rangle_N\left\langle\tilde{r}_t,u\right\rangle_N-E\left(\left\langle\tilde{r}_t,v\right\rangle_N\left\langle\tilde{r}_t,u\right\rangle_N\right)\right\}+O_p\left(\frac{1}{\sqrt{T}}\right).$$

From here on, the proof is similar to that of Lemma 7. To apply the central limit theorem of Francq and Zakoïan (2005), we need

$$\lim_{N\to\infty}\sup_t E[(\left\langle v,\tilde{r}_t\right\rangle_N\left\langle\tilde{r}_t,u\right\rangle_N)^{2+\rho}]<\infty$$

for some $\rho > 0$. This condition is met because of Assumption 4(iii).

Proof of Proposition 2

Consistency:

Recall that by Equation (9), we have

$$\begin{aligned}\hat{\theta}^{\gamma}_{HJ} - \theta_{HJ} = & (\hat{V}_{11}^{-1} - V_{11}^{-1})\lambda \\ & + \hat{V}_{11}^{-1}(\hat{\mu}_1 - \mu_1) \\ & + \hat{V}_{11}^{-1}\left(\frac{\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}\hat{\beta}}{N}\right)^{-1}\left[\frac{\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}(\beta - \hat{\beta})}{N}(\lambda + \hat{\mu}_1 - \mu_1)\right. \\ & \left. + \frac{\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}\alpha}{N} + \frac{\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}\bar{\epsilon}}{N}\right].\end{aligned} \tag{27}$$

For the first two rows, $(\hat{V}_{11}^{-1} - V_{11}^{-1})\lambda$ and $\hat{V}_{11}^{-1}(\hat{\mu}_1 - \mu_1)$ converge to 0 in probability by the law of large numbers and Assumption 2 (i).

$$\begin{aligned}\frac{\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}\hat{\beta}}{N} &= \left[\frac{\hat{\beta}'_{k_1}\hat{\Sigma}_{\gamma}^{-1}\hat{\beta}_{k_2}}{N}\right]_{k_1,k_2=1,\cdots,K} \\ &= \left\langle \hat{\beta}_{k_1}, \hat{\Sigma}_{\gamma}^{-1}\hat{\beta}_{k_2}\right\rangle_{N;k_1,k_2=1,\cdots,K}.\end{aligned}$$

We have

$$\begin{aligned}\left\langle \hat{\beta}_{k_1}, \hat{\Sigma}_{\gamma}^{-1}\hat{\beta}_{k_2}\right\rangle_N &= \left\langle \hat{\beta}_{k_1} - \beta_{k_1}, \hat{\Sigma}_{\gamma}^{-1}\hat{\beta}_{k_2} - \Sigma^{-1}\beta_{k_2}\right\rangle_N + & (28)\\ & \left\langle \hat{\beta}_{k_1} - \beta_{k_1}, \Sigma^{-1}\beta_{k_2}\right\rangle_N + & (29)\\ & \left\langle \beta_{k_1}, \hat{\Sigma}_{\gamma}^{-1}\hat{\beta}_{k_2} - \Sigma^{-1}\beta_{k_2}\right\rangle_N + & (30)\\ & \left\langle \beta_{k_1}, \Sigma^{-1}\beta_{k_2}\right\rangle_N - C_{k_1,k_2} & (31)\\ & + C_{k_1,k_2}. & (32)\end{aligned}$$

where C_{k_1,k_2} is the (k_1, k_2) element of the matrix C defined in Assumption 3.

$|(28)| \leq \left\|\hat{\beta}_{k_1} - \beta_{k_1}\right\|_N \left\|\hat{\Sigma}_{\gamma}^{-1}\hat{\beta}_{k_2} - \Sigma^{-1}\beta_{k_2}\right\|_N \to 0$ as $N, T \to \infty$, and $\gamma T \to \infty$ using Lemma 6.

For (29), we have

$$\left|\left\langle \hat{\beta}_{k_1} - \beta_{k_1}, \Sigma^{-1}\beta_{k_2}\right\rangle_N\right| \leq \left\|\hat{\beta}_{k_1} - \beta_{k_1}\right\|_N \left\|\Sigma^{-1}\beta_{k_2}\right\|_N \to 0$$

as $N, T \to \infty$, by Assumption 3(i) and Lemma 4.

By Lemma 6, $\|(30)\| = o_p(1)$.

Finally, using Assumption 3, (31) goes to 0 as N goes to ∞.

In conclusion, $\frac{\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\hat{\beta}}{N} \to C$ as $N, T \to \infty$, $\gamma T \to \infty$, and $\gamma \to 0$.

Using the same argument as before, we have $\frac{\hat{\beta}'\hat{\Sigma}_\gamma^{-1}(\beta - \hat{\beta})}{N} = \frac{\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\beta}{N} - \frac{\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\hat{\beta}}{N} \xrightarrow{P} 0$ as $N, T \to \infty$, $\gamma T \to \infty$, and $\gamma \to 0$.

For $\frac{\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\alpha}{N}$, we have $\hat{\Sigma}_\gamma^{-1}\hat{\beta} \xrightarrow{P} \Sigma^{-1}\beta$ when as $N, T \to \infty$, and $\gamma T \to \infty$. Moreover, $\beta'\Sigma^{-1}\alpha = 0$ by the first order condition of (2). Therefore when as $N, T \to \infty$, and $\gamma T \to \infty$, $\frac{\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\alpha}{N} \xrightarrow{P} 0$.

The same is true for $\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\bar{\epsilon}$, which converges in probability to 0 as $N, T \to \infty$, and $\gamma T \to \infty$. The consistency of $\hat{\theta}_{HJ}^{\gamma}$ follows.

Distribution:

We detail the proof of the asymptotic normality for Tikhonov estimator. The result for ridge could be shown similarly. We analyze the decomposition (27) using the following results:

- $\hat{\beta} - \beta = \frac{1}{T}\sum_{t=1}^{T} \epsilon_t \bar{f}_t{}' \hat{V}_{11}^{-1}$
- Note that $\hat{C}_\beta = \frac{\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\hat{\beta}}{N}$ and we already have shown that $\hat{C}_\beta - C_\beta \xrightarrow{P} 0_{K,K}$.
- $\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\alpha = (\hat{\beta} - \beta)'\hat{\Sigma}_\gamma^{-1}\alpha + \beta'(\hat{\Sigma}_\gamma^{-1} - \Sigma^{-1})\alpha$ using the fact that $\beta'\Sigma^{-1}\alpha = 0$ by the population first-order condition of (2).
- We have

$$\begin{aligned}
\beta'(\hat{\Sigma}_\gamma^{-1} - \Sigma^{-1})\alpha &= \beta'(\hat{\Sigma}_\gamma^{-1} - \Sigma_\gamma^{-1} + \Sigma_\gamma^{-1} - \Sigma^{-1})\alpha \\
&= \beta'(\hat{\Sigma}_\gamma^{-1} - \Sigma_\gamma^{-1})\alpha + \beta'(\Sigma_\gamma^{-1} - \Sigma^{-1})\alpha
\end{aligned}$$

  and

$$\begin{aligned}
\beta'(\hat{\Sigma}_\gamma^{-1} - \Sigma_\gamma^{-1})\alpha &= \beta'\left(\hat{\Sigma}^2 + \gamma I\right)^{-1}\left[\hat{\Sigma}\left(\Sigma - \hat{\Sigma}\right)\Sigma + \gamma\left(\hat{\Sigma} - \Sigma\right)\right]\left(\Sigma^2 + \gamma I\right)^{-1}\alpha \\
&= \beta'\hat{\Sigma}_\gamma^{-1}\left(\Sigma - \hat{\Sigma}\right)\Sigma_\gamma^{-1}\alpha + \gamma\beta'\left(\hat{\Sigma}^2 + \gamma I\right)^{-1}\left(\hat{\Sigma} - \Sigma\right)\left(\Sigma^2 + \gamma I\right)^{-1}\alpha
\end{aligned}$$

Therefore,

$$
\begin{aligned}
&\sqrt{T}\left(\hat{\theta}^{\gamma}_{HJ}-\theta_{HJ}\right)\\
&+\sqrt{T}\hat{V}_{11}^{-1}\hat{C}_{\beta}^{-1}\left\{\gamma\beta'\left(\hat{\Sigma}^2+\gamma I\right)^{-1}\left(\hat{\Sigma}-\Sigma\right)\left(\Sigma^2+\gamma I\right)^{-1}\frac{\alpha}{N}+\beta'(\Sigma_{\gamma}^{-1}-\Sigma^{-1})\frac{\alpha}{N}\right\} &(33)\\
&=\hat{V}_{11}^{-1}\Bigg\{-\sqrt{T}(\hat{V}_{11}-V_{11})V_{11}^{-1}\lambda &(34)\\
&+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(f_t-\mu_1) &(35)\\
&+\hat{C}_{\beta}^{-1}\Bigg[-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}\epsilon_t\bar{f}_t'}{N}\hat{V}_{11}^{-1}(\lambda+\hat{\mu}_1-\mu_1) &(36)\\
&-\sqrt{T}\beta'\hat{\Sigma}_{\gamma}^{-1}(\hat{\Sigma}-\Sigma)\Sigma_{\gamma}^{-1}\frac{\alpha}{N} &(37)\\
&+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{V}_{11}^{-1}\frac{\bar{f}_t\epsilon_t'\hat{\Sigma}_{\gamma}^{-1}\alpha}{N} &(38)\\
&+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}\epsilon_t}{N}\Bigg]\Bigg\}. &(39)
\end{aligned}
$$

We need to prove the asymptotic normality of each component of (34) to (39) to get the result of Proposition 2.

- Regarding (34), note that

$$
\begin{aligned}
\sqrt{T}(\hat{V}_{11}-V_{11})V_{11}^{-1}\lambda &= \sqrt{T}\hat{V}_{11}\theta_{HJ}-\sqrt{T}\lambda\\
\sqrt{T}\hat{V}_{11} &= \frac{1}{\sqrt{T}}\sum_t\left(f_t-\widehat{\mu}_1\right)\left(f_t-\widehat{\mu}_1\right)'\\
&= \frac{1}{\sqrt{T}}\sum_t\tilde{f}_t\tilde{f}_t'+\frac{1}{\sqrt{T}}\sum_t\tilde{f}_t\left(\mu_1-\widehat{\mu}_1\right)'\\
&\quad+\frac{1}{\sqrt{T}}\sum_t\left(\mu_1-\widehat{\mu}_1\right)\tilde{f}_t'+\sqrt{T}\left(\mu_1-\widehat{\mu}_1\right)\left(\mu_1-\widehat{\mu}_1\right)'\\
&= \frac{1}{\sqrt{T}}\sum_t\tilde{f}_t\tilde{f}_t'+o_p\left(1\right).
\end{aligned}
$$

because $\sqrt{T}\left(\mu_1-\widehat{\mu}_1\right)=O_p\left(1\right)$ and $\frac{1}{\sqrt{T}}\sum_t\tilde{f}_t=O_p\left(1\right)$. Hence

$$
\sqrt{T}(\hat{V}_{11}-V_{11})V_{11}^{-1}\lambda=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\tilde{f}_t\tilde{f}_t'\theta_{HJ}-\lambda\right)+o_p(1).
$$

- Normality of (35) comes from Assumption 2.
- Consider term (36):

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}\epsilon_t\bar{f}_t'}{N}\hat{V}_{11}^{-1}(\lambda+\hat{\mu}_1-\mu_1) &= \frac{\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}}{N}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\tilde{f}_t'V_{11}^{-1}\lambda\right. &(40)\\ &+ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\bar{f}_t'(\hat{V}_{11}^{-1}-V_{11}^{-1})\lambda &(41)\\ &- \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t(\hat{\mu}_1-\mu_1)'V_{11}^{-1}\lambda &(42)\\ &+ \left.\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\bar{f}_t'\hat{V}_{11}^{-1}(\hat{\mu}_1-\mu_1)\right]. &(43)\end{aligned}$$

The term (41) can be rewritten as $T^{-\frac{1}{2}}\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}\frac{\epsilon'\bar{F}}{N}(\hat{V}_{11}^{-1}-V_{11}^{-1})\lambda$. From Lemma 2, $\left\|\epsilon'\bar{F}\right\|_F^2 = O_p(NT)$. So $\left\|\epsilon'\bar{F}\right\|^2 = O_p(NT)$. From Lemma 6, $\left\|\hat{\Sigma}_{\gamma}^{-1}\hat{\beta}_k-\Sigma^{-1}\beta_k\right\|_N \to 0,$ as $N,T\to\infty,$ and $\gamma T\to\infty$. In addition, $\left\|(\hat{V}_{11}^{-1}-V_{11}^{-1})\lambda\right\|^2 = O_p\left(\frac{1}{T}\right)$. Therefore, (41) is $o_p(1)$ when $N,T\to\infty$ and $\gamma T\to\infty$.

Using the fact that $\theta_{HJ}=V_{11}^{-1}\lambda$, we can rewrite (42) as

$$-(\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}-\beta'\Sigma^{-1})\frac{1}{N\sqrt{T}}\sum_t\epsilon_t(\hat{\mu}_1-\mu_1)'\theta_{HJ}-\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\langle\Sigma^{-1}\beta,\epsilon_t\right\rangle_N\right)(\hat{\mu}_1-\mu_1)'\theta_{HJ}.$$

Using a proof similar to that of Lemma 2, we can show that $\left\|\sum_t\epsilon_t\right\|^2 = \|\epsilon'1\|^2 = O_p(NT)$. Moreover, $\|\hat{\mu}_1-\mu_1\|^2 = O_P\left(\frac{1}{T}\right)$. As a result, $(\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}-\beta'\Sigma^{-1})\frac{1}{N\sqrt{T}}\sum_t\epsilon_t(\hat{\mu}_1-\mu_1)'\theta_{HJ}$ is $o_p(1)$ when $N,T\to\infty,$ and $\gamma T\to\infty$.

Using Lemma 7, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\langle\Sigma^{-1}\beta,\epsilon_t\right\rangle_N$ is $O_p(1)$ as $\left\|\Sigma^{-1}\beta\right\|_N<\infty$ and ϵ_t has the same characteristics as $\tilde{r}_t$. Then the second term is $o_p(1)$ when $N,T\to\infty$. Therefore, (42) is $o_p(1)$.

Regarding (43), we can rewrite it as

$$\frac{\sqrt{T}}{N}(\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}-\beta'\Sigma^{-1})(\hat{\beta}-\beta)(\hat{\mu}_1-\mu_1)+\frac{\sqrt{T}}{N}\beta'\Sigma^{-1}(\hat{\beta}-\beta)(\hat{\mu}_1-\mu_1).$$

From Lemmas 3 and 6, this expression is $o_p(1)$ as $N,T\to\infty,$ and $\gamma T\to\infty$.

Finally, (40) is equal to

$$\frac{1}{N\sqrt{T}}\left(\hat{\beta}'\hat{\Sigma}_\gamma^{-1}-\beta'\Sigma^{-1}\right)\epsilon'\tilde{F}V_{11}^{-1}\lambda+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\langle\Sigma^{-1}\beta,\epsilon_t\tilde{f}_t'V_{11}^{-1}\lambda\right\rangle_N.$$

The first part is $o_P(1)$ when $N,T\to\infty$, and $\gamma T\to\infty$. For the second part, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\langle\Sigma^{-1}\beta,\epsilon_t\tilde{f}_t'V_{11}^{-1}\lambda\right\rangle_N$ converges to a normal distribution by virtue of Lemma 7 applied to $u=\Sigma^{-1}\beta$ and replacing r_t by $\epsilon_t\tilde{f}_t'$.

- Consider (37). As $N,T\to\infty$, $\gamma\to0$ and $\gamma T\to\infty$, the normality of $\sqrt{T}\beta'\hat{\Sigma}_\gamma^{-1}(\hat{\Sigma}-\Sigma)\Sigma_\gamma^{-1}\frac{\alpha}{N}$ comes from Lemma 8. Indeed,

$$\sqrt{T}\beta'\hat{\Sigma}_\gamma^{-1}(\hat{\Sigma}-\Sigma)\Sigma_\gamma^{-1}\frac{\alpha}{N}=\sqrt{T}\beta'\Sigma^{-1}(\hat{\Sigma}-\Sigma)\frac{\Sigma^{-1}\alpha}{N}+o_p(1).$$

Using the proof of Lemma 8 , we can rewrite $\left\langle\Sigma^{-1}\beta,\sqrt{T}(\hat{\Sigma}-\Sigma)\Sigma^{-1}\alpha\right\rangle_N$ as

$$\begin{aligned}\left\langle\Sigma^{-1}\beta,\sqrt{T}\left[\hat{\Sigma}-\Sigma\right]\Sigma^{-1}\alpha\right\rangle_N &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\left\langle\tilde{r}_t,\Sigma^{-1}\alpha\right\rangle_N\left\langle\tilde{r}_t,\Sigma^{-1}\beta\right\rangle_N\right.\\ &\quad -E[\left\langle\tilde{r}_t,\Sigma^{-1}\alpha\right\rangle_N\left\langle\tilde{r}_t,\Sigma^{-1}\beta\right\rangle]+o_p(1).\end{aligned}$$

This term is asymptotically Gaussian.

- Consider (38). $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{V}_{11}^{-1}\frac{\bar{f}_t\epsilon_t'\hat{\Sigma}_\gamma^{-1}\alpha}{N}$ has a Gaussian distribution by using the following decomposition

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{V}_{11}^{-1}\frac{\bar{f}_t\epsilon_t'\hat{\Sigma}_\gamma^{-1}\alpha}{N} &= V_{11}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\tilde{f}_t\epsilon_t'\hat{\Sigma}_\gamma^{-1}\alpha}{N} &(44)\\ &+\frac{1}{N\sqrt{T}}\left(\hat{V}_{11}^{-1}-V_{11}^{-1}\right)\bar{F}'\epsilon\hat{\Sigma}_\gamma^{-1}\alpha &(45)\\ &+\frac{1}{N\sqrt{T}}V_{11}^{-1}(\mu_1-\hat{\mu}_1)\sum_{t=1}^{T}\epsilon_t'\hat{\Sigma}_\gamma^{-1}\alpha. &(46)\end{aligned}$$

By Lemma 7, (44) converges to a Gaussian distribution. Using Lemmas 2 and 6, (45) and (46) are $o_p(1)$ when $N,T\to\infty$, and $\gamma T\to\infty$.

- The term in (33) can be dealt with using the same approach as in the proof of Lemma 6 with

$\omega = 1$. We obtain the rate $O_p\left(\sqrt{T}\gamma\right)$. Therefore, the term (33) is $o_p(1)$ as $N, T, \gamma T \to \infty$ and $\gamma^2 T \to 0$.

Using the previous results, we have

$$\sqrt{T}\left(\hat{\theta}^{\gamma}_{HJ} - \theta_{HJ}\right) - \hat{V}_{11}^{-1}.A = o_p(1),$$

where

$$\begin{aligned}
A &= \left(-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{f}_t\tilde{f}_t'\theta_{HJ} + \sqrt{T}\lambda\right) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{f}_t + C_{\beta}^{-1}\left[-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\beta'\Sigma^{-1}\frac{\epsilon_t\tilde{f}_t'\theta_{HJ}}{N}\right. \\
&\quad \left. + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_{11}^{-1}\tilde{f}_t\frac{\epsilon_t'\Sigma^{-1}\alpha}{N} - \beta'\Sigma^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\tilde{r}_t\tilde{r}_t'}{N^2} - \frac{\Sigma}{N}\right)\right)\Sigma^{-1}\alpha + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\beta'\Sigma^{-1}\epsilon_t}{N}\right] \\
&\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}h_t
\end{aligned}$$

and

$$\begin{aligned}
h_t &= -\tilde{f}_t\tilde{f}_t'\theta_{HJ} + \lambda + \tilde{f}_t - C_{\beta}^{-1}\beta'\Sigma^{-1}\frac{\epsilon_t\tilde{f}_t'\theta_{HJ}}{N} + C_{\beta}^{-1}V_{11}^{-1}\tilde{f}_t\frac{\epsilon_t'\Sigma^{-1}\alpha}{N} \\
&\quad - C_{\beta}^{-1}\beta'\Sigma^{-1}\frac{\tilde{r}_t\tilde{r}_t'}{N^2}\Sigma^{-1}\alpha + C_{\beta}^{-1}\frac{\beta'\Sigma^{-1}\epsilon_t}{N}.
\end{aligned}$$

As all the components of A are normally distributed, we have

$$\sqrt{T}\left(\hat{\theta}^{\gamma}_{HJ} - \theta_{HJ}\right) \xrightarrow{d} \mathcal{N}(0_K, V_{11}^{-1}\Omega V_{11}^{-1}),$$

where $\Omega = \lim_{N,T\to\infty} var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}h_t\right]$.

Using the fact that $y_t = 1 - \tilde{f}_t'\theta_{HJ}$ and noting $\tilde{u}_t = \frac{\tilde{r}_t'\Sigma^{-1}\alpha}{N}$, h_t can be rewritten as follows

$$h_t = \tilde{f}_t y_t + \lambda + \frac{C_{\beta}^{-1}\beta'\Sigma^{-1}}{N}(\epsilon_t y_t - \tilde{r}_t\tilde{u}_t) + C_{\beta}^{-1}V_{11}^{-1}\tilde{f}_t\frac{\epsilon_t'\Sigma^{-1}\alpha}{N}.$$

11.4 Proof of Proposition 3

Let

$$S_T = \sum_{|j|<T}\Gamma_j$$

where $\Gamma_j = \dfrac{1}{T}\sum_{t=j+1}^{T} E\left(h_t h_{t-j}'\right)$ for $j \geq 0$ and $\Gamma_j = \Gamma_{-j}'$. Note that $S = \lim_{T\to\infty} S_T$.

Let

$$\widetilde{S}_T = \sum_{|j|<T} \omega\left(\frac{j}{m_T}\right)\widetilde{\Gamma}_j$$

where $\widetilde{\Gamma}_j = \dfrac{1}{T}\sum_{t=j+1}^{T} h_t h_{t-j}'$ for $j \geq 0$ and $\widetilde{\Gamma}_j = \widetilde{\Gamma}_{-j}'$ and $h_t = h_t(\xi_1, \xi_2)$ depends on the (pseudo) true values of the parameters. On the other hand, $\widehat{h}_t = h_t\left(\hat{\xi}_1, \hat{\xi}_2\right)$ depends on the estimators.

We have

$$\hat{S}_T - S = \hat{S}_T - \widetilde{S}_T + \widetilde{S}_T - S_T + S_T - S$$

where $S_T - S \xrightarrow{P} 0$ by construction.

Note that under Assumption 5, the process $\{h_t\}$ satisfies the conditions of Theorem 1(a) of Andrews (1991). It follows from this theorem that $\widetilde{S}_T - S_T \to 0$ provided $m_T/\sqrt{T} \to 0$.

Now we focus on the term $\hat{S}_T - \widetilde{S}_T$. To simplify the notations, we can assume that there is only one factor, so that h_t and S are scalar, see the argument of Andrews (1991) (p.852) to justify that this can be done without loss of generality. The set of finite parameters ξ_1 includes parameters which estimators converge at the usual $\sqrt{T}$ rate, namely μ_1, V_{11}, and θ, and one parameter, C_β, which estimator converges at a slower rate. Indeed, we have

$$\begin{aligned}
\left|\widehat{C}_\beta - C_\beta\right| &= \left|\frac{\hat{\beta}'\hat{\Sigma}_\gamma^{-1}\hat{\beta}}{N} - \frac{\beta'\Sigma^{-1}\beta}{N}\right| \\
&= \left|\frac{\left(\hat{\beta}-\beta\right)'\Sigma^{-1}\beta}{N} + \frac{\beta'\left(\hat{\Sigma}_\gamma^{-1}\hat{\beta} - \Sigma^{-1}\beta\right)}{N} + \frac{\left(\hat{\beta}-\beta\right)'\left(\hat{\Sigma}_\gamma^{-1}\hat{\beta} - \Sigma^{-1}\beta\right)}{N}\right|
\end{aligned}$$

where we used the fact that $\widehat{a}\widehat{b} - ab = (\widehat{a}-a)\,b + a\left(\widehat{b}-b\right) + (\widehat{a}-a)\left(\widehat{b}-b\right)$. By Cauchy-Schwarz, we have

$$\begin{aligned}
\left|\frac{\left(\hat{\beta}-\beta\right)'\Sigma^{-1}\beta}{N}\right| &\leq \left\|\hat{\beta}-\beta\right\|_N \left\|\Sigma^{-1}\beta\right\|_N = O_p\left(\frac{1}{\sqrt{T}}\right), \\
\left|\frac{\beta'\left(\hat{\Sigma}_\gamma^{-1}\hat{\beta} - \Sigma^{-1}\beta\right)}{N}\right| &\leq \|\beta\|_N \left\|\hat{\Sigma}_\gamma^{-1}\hat{\beta} - \Sigma^{-1}\beta\right\|_N = O_p\left(\frac{1}{\sqrt{\gamma T}} + \sqrt{\gamma}\right), \\
\left|\frac{\left(\hat{\beta}-\beta\right)'\left(\hat{\Sigma}_\gamma^{-1}\hat{\beta} - \Sigma^{-1}\beta\right)}{N}\right| &\leq \left\|\hat{\beta}-\beta\right\|_N \left\|\hat{\Sigma}_\gamma^{-1}\hat{\beta} - \Sigma^{-1}\beta\right\|_N = O_p\left(\frac{1}{T\sqrt{\gamma}} + \frac{\sqrt{\gamma}}{\sqrt{T}}\right)
\end{aligned}$$

where we used Lemma 6. It follows that

$$\left|\widehat{C}_\beta - C_\beta\right| = O_p\left(\frac{1}{\sqrt{\gamma T}} + \sqrt{\gamma}\right).$$

Regarding the estimators of the elements of ξ_2, we have the following rates

$$\begin{aligned}
\left\|\hat{\Sigma}_\gamma^{-1}\hat{\beta} - \Sigma^{-1}\beta\right\|_N &= O_p\left(\frac{1}{\sqrt{\gamma T}} + \sqrt{\gamma}\right), \\
\left\|\hat{\Sigma}_\gamma^{-1}\widehat{\alpha} - \Sigma^{-1}\alpha\right\|_N &= O_p\left(\frac{1}{\sqrt{\gamma T}} + \sqrt{\gamma}\right), \\
\left\|\widehat{\mu}_2 - \mu_2\right\|_N &= O_p\left(\frac{1}{\sqrt{T}}\right), \\
\left\|\widehat{\beta} - \beta\right\|_N &= O_p\left(\frac{1}{\sqrt{T}}\right)
\end{aligned}$$

where the first rate follows again from Lemma 6 and the second rate can be established using a similar proof as that of Lemma 6.

We have

$$\hat{S}_T - \widetilde{S}_T = \sum_{|j|<T} \omega\left(\frac{j}{m_T}\right)\left(\tilde{\Gamma}_j\left(\widehat{\xi}_1, \widehat{\xi}_2\right) - \tilde{\Gamma}_j\left(\xi_1, \xi_2\right)\right)$$

and

$$\begin{aligned}
\tilde{\Gamma}_j\left(\widehat{\xi}_1, \widehat{\xi}_2\right) - \tilde{\Gamma}_j\left(\xi_1, \xi_2\right) &= \tilde{\Gamma}_j\left(\widehat{\xi}_1, \widehat{\xi}_2\right) - \tilde{\Gamma}_j\left(\hat{\xi}_1, \xi_2\right) \\
&\quad + \tilde{\Gamma}_j\left(\hat{\xi}_1, \xi_2\right) - \tilde{\Gamma}_j\left(\xi_1, \xi_2\right).
\end{aligned}$$

Using the mean-value theorem, we have

$$\tilde{\Gamma}_j\left(\hat{\xi}_1, \xi_2\right) - \tilde{\Gamma}_j\left(\xi_1, \xi_2\right) = \frac{\partial \tilde{\Gamma}_j\left(\overline{\xi}_1, \xi_2\right)}{\partial \xi_1}\left(\hat{\xi}_1 - \xi_1\right)$$

where $\overline{\xi}_1$ is between $\widehat{\xi}_1$ and ξ_1. In addition,

$$\begin{aligned}
\sup_{j\geq 1}\left\|\frac{\partial \tilde{\Gamma}_j\left(\overline{\xi}_1, \xi_2\right)}{\partial \xi_1}\right\| &= \sup_{j\geq 1}\left\|\frac{1}{T}\sum_{t=j+1}^{T} h_t\left(\overline{\xi}_1, \xi_2\right)\frac{\partial h_{t-j}\left(\overline{\xi}_1, \xi_2\right)}{\partial \xi_1} + h_{t-j}\left(\overline{\xi}_1, \xi_2\right)\frac{\partial h_t\left(\overline{\xi}_1, \xi_2\right)}{\partial \xi_1}\right\| \\
&\leq 2\left(\frac{1}{T}\sum_{t=1}^{T}\sup_{\xi_1\in E} h_t\left(\xi_1, \xi_2\right)^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\sup_{\xi_1\in E}\left\|\frac{\partial h_t\left(\xi_1, \xi_2\right)}{\partial \xi_1}\right\|^2\right)^{1/2} \\
&= O_p(1)
\end{aligned}$$

by Assumption (5) and Markov's inequality.

Then, we obtain

$$\begin{aligned}
\left|\sum_{|j|<T}\omega\left(\frac{j}{m_T}\right)\left(\tilde{\Gamma}_j\left(\hat{\xi}_1,\xi_2\right)-\tilde{\Gamma}_j\left(\xi_1,\xi_2\right)\right)\right| &\leq \sum_{|j|<T}\left|\omega\left(\frac{j}{m_T}\right)\right|\left\|\frac{\partial\tilde{\Gamma}_j\left(\overline{\xi}_1,\xi_2\right)}{\partial\xi_1}\right\|\left\|\hat{\xi}_1-\xi_1\right\| \\
&\leq \frac{1}{m_T}\sum_{|j|<T}\left|\omega\left(\frac{j}{m_T}\right)\right|\sup_j\left\|\frac{\partial\tilde{\Gamma}_j\left(\overline{\xi}_1,\xi_2\right)}{\partial\xi_1}\right\|m_T\left\|\hat{\xi}_1-\xi_1\right\| \\
&= O_p\left(\frac{m_T}{\sqrt{\gamma T}}+m_T\sqrt{\gamma}\right)
\end{aligned}$$

where we used the fact that $\frac{1}{m_T}\sum_{|j|<T}\left|\omega\left(\frac{j}{m_T}\right)\right|\to\int|\omega(x)|\,dx<\infty$ (see Andrews (1991), p.852).

Now we turn our attention on the term $\tilde{\Gamma}_j\left(\widehat{\xi}_1,\widehat{\xi}_2\right)-\tilde{\Gamma}_j\left(\hat{\xi}_1,\xi_2\right)$. We have

$$\begin{aligned}
&\tilde{\Gamma}_j\left(\widehat{\xi}_1,\widehat{\xi}_2\right)-\tilde{\Gamma}_j\left(\hat{\xi}_1,\xi_2\right) \\
=\;& \frac{1}{T}\sum_{t=j+1}^{T}h_t\left(\widehat{\xi}_1,\widehat{\xi}_2\right)h_{t-j}\left(\widehat{\xi}_1,\widehat{\xi}_2\right)-h_t\left(\widehat{\xi}_1,\xi_2\right)h_{t-j}\left(\widehat{\xi}_1,\xi_2\right) \\
=\;& \frac{1}{T}\sum_{t=j+1}^{T}\left(h_t\left(\widehat{\xi}_1,\widehat{\xi}_2\right)-h_t\left(\widehat{\xi}_1,\xi_2\right)\right)h_{t-j}\left(\widehat{\xi}_1,\xi_2\right) \\
&+\frac{1}{T}\sum_{t=j+1}^{T}h_t\left(\widehat{\xi}_1,\xi_2\right)\left(h_{t-j}\left(\widehat{\xi}_1,\widehat{\xi}_2\right)-h_{t-j}\left(\widehat{\xi}_1,\xi_2\right)\right) \\
&+\frac{1}{T}\sum_{t=j+1}^{T}\left(h_t\left(\widehat{\xi}_1,\widehat{\xi}_2\right)-h_t\left(\widehat{\xi}_1,\xi_2\right)\right)\left(h_{t-j}\left(\widehat{\xi}_1,\widehat{\xi}_2\right)-h_{t-j}\left(\widehat{\xi}_1,\xi_2\right)\right).
\end{aligned}$$

Observe that $h_t(\xi_1,\xi_2)$ is linear in $\Sigma^{-1}\beta$, $\Sigma^{-1}\alpha$, β and is quadratic in μ_2. The most complicated term is the one involving $\Sigma^{-1}\alpha$. We will focus on this one, the rest of the proof is similar. Let us denote $\Sigma^{-1}\alpha$ by ν and let $\xi_2=(\nu',\xi_{2\nu}')'$.

$$\begin{aligned}
&\frac{1}{T}\sum_{t=j+1}^{T}\left(h_t\left(\widehat{\xi}_1,\widehat{\xi}_2\right)-h_t\left(\widehat{\xi}_1,\xi_2\right)\right)h_{t-j}\left(\widehat{\xi}_1,\xi_2\right) \\
=\;& \frac{1}{T}\sum_{t=j+1}^{T}\left(h_t\left(\widehat{\xi}_1,\hat{\nu},\widehat{\xi}_{2\nu}\right)-h_t\left(\widehat{\xi}_1,\nu,\widehat{\xi}_{2\nu}\right)+h_t\left(\widehat{\xi}_1,\nu,\widehat{\xi}_{2\nu}\right)-h_t\left(\widehat{\xi}_1,\xi_2\right)\right)h_{t-j}\left(\widehat{\xi}_1,\xi_2\right).
\end{aligned}$$

We have

$$\begin{aligned}
& \frac{1}{T}\sum_{t=j+1}^{T}\left(h_t\left(\widehat{\xi}_1,\hat{\nu},\widehat{\xi}_{2\nu}\right)-h_t\left(\widehat{\xi}_1,\nu,\widehat{\xi}_{2\nu}\right)\right)h_{t-j}\left(\widehat{\xi}_1,\xi_2\right) \\
= & -\hat{C}_{\beta}^{-1}\frac{\hat{\beta}'\hat{\Sigma}_{\gamma}^{-1}}{N}\left(\frac{1}{T}\sum_{t=j+1}^{T}\overline{r}_t\overline{r}_t'h_{t-j}\left(\widehat{\xi}_1,\xi_2\right)\right)\frac{\hat{\nu}-\nu}{N}+\hat{C}_{\beta}^{-1}\widehat{V}_{11}^{-1}\left(\frac{1}{T}\sum_{t=j+1}^{T}\overline{f}_t h_{t-j}\left(\widehat{\xi}_1,\xi_2\right)\widehat{\epsilon}_t'\right)\frac{\hat{\nu}-\nu}{N}
\end{aligned} \tag{47}$$

The rate of convergence is not affected by the fact that ξ_1 and $\xi_{2\nu}$ are estimated, hence to simplify notations, we drop the hats on $\widehat{\xi}_1$ and $\widehat{\xi}_{2\nu}$. Consider first the second term of (47).

$$\begin{aligned}
& \left|C_{\beta}^{-1}V_{11}^{-1}\left(\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{f}_t h_{t-j}\epsilon_t'\right)\frac{\hat{\nu}-\nu}{N}\right| \\
= & \left|C_{\beta}^{-1}V_{11}^{-1}\left\langle\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{f}_t h_{t-j}\epsilon_t,\hat{\nu}-\nu\right\rangle_N\right| \\
\leq & C_{\beta}^{-1}V_{11}^{-1}\left\|\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{f}_t h_{t-j}\epsilon_t\right\|_N\|\hat{\nu}-\nu\|_N \\
\leq & C_{\beta}^{-1}V_{11}^{-1}\left(\frac{1}{T}\sum_{t=j+1}^{T}h_{t-j}^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=j+1}^{T}\left\|\widetilde{f}_t\epsilon_t\right\|_N^2\right)^{1/2}\|\hat{\nu}-\nu\|_N \\
= & O_p\left(\frac{1}{\sqrt{\gamma T}}+\sqrt{\gamma}\right).
\end{aligned}$$

Now consider the first term of (47). To simplify notations, denote $\Sigma_\gamma^{-1}\beta$ by τ. We have

$$
\begin{aligned}
& C_\beta^{-1}\frac{\tau'}{N}\left(\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{r}_t\widetilde{r}_t{}'h_{t-j}\right)\frac{\hat{\nu}-\nu}{N} \\
= \;& C_\beta^{-1}\left\langle \tau, \left(\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{r}_t\widetilde{r}_t{}'h_{t-j}\right)\frac{\hat{\nu}-\nu}{N}\right\rangle_N \\
\leq \;& C_\beta^{-1}\left\|\tau\right\|_N \left\|\left(\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{r}_t\widetilde{r}_t{}'h_{t-j}\right)\frac{\hat{\nu}-\nu}{N}\right\|_N \\
\leq \;& C_\beta^{-1}\left\|\tau\right\|_N \left\|\left(\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{r}_t\widetilde{r}_t{}'h_{t-j}\right)\right\|_{op}\frac{1}{N}\left\|\hat{\nu}-\nu\right\|_N \\
\leq \;& C_\beta^{-1}\left\|\tau\right\|_N \left(\frac{1}{N^2T}\sum_{t=j+1}^{T}\left\|\widetilde{r}_t\right\|^4\right)^{1/2}\left(\frac{1}{T}\sum_{t=j+1}^{T}h_{t-j}^2\right)^{1/2}\left\|\hat{\nu}-\nu\right\|_N \\
= \;& O_p\left(\frac{1}{\sqrt{\gamma T}}+\sqrt{\gamma}\right)
\end{aligned}
$$

where we used the fact that $E\left\|\widetilde{r}_t\right\|^4/N^2<\infty$ by Assumption 4. The other terms can be dealt with similarly and details are omitted.

11.5 Proof of Propositions 4 and 5

The proof will use the following lemmas.

Lemma 9. *In the case where Model b nests Model a (case $K_2 = 0$), $\delta_a^2 - \delta_b^2 = 0 \Leftrightarrow \theta_{b2} = 0_{K_3}$ where θ_{b2} is the coefficient of f_{3t} in Model b.*

Lemma 10. *In the case of non-nested models with overlapping factors f_{1t}, $y_a = y_b \Leftrightarrow \theta_{a2} = 0_{K_2}$ and $\theta_{b2} = 0_{K_3}$ where θ_{a2} is the coefficient of f_{2t} in Model a.*

Proof of Lemmas 9 and 10.

The proof is similar to that of Lemmas 2 and 3 of Kan and Robotti (2009). The main difference is that they work with gross returns while we use excess returns.

In the case of nested models, $\delta_a^2 - \delta_b^2 = 0 \Leftrightarrow y_a = y_b$. So we can see Lemma 9 as a special case of Lemma 10 when $K_2 = 0$. Therefore, we focus on proving Lemma 10.

It is easy to see that $y_a = y_b \Leftrightarrow \theta_{a1} = \theta_{b1}$, $\theta_{a2} = 0_{K_2}$ and $\theta_{b2} = 0_{K_3}$. So it suffices to show that $\theta_{a2} = 0_{K_2}$ and $\theta_{b2} = 0_{K_3}$ imply $\theta_{a1} = \theta_{b1}$.

We have

$$\theta_m = \left(V_{m,12}V_{22}^{-1}V_{m,21}\right)^{-1} V_{m,12}V_{22}^{-1}\mu_2,\ m = a, b.$$

It follows that

$$\left(V_{m,12}V_{22}^{-1}V_{m,21}\right)\theta_m = V_{m,12}V_{22}^{-1}\mu_2. \tag{48}$$

Using the decomposition $V_{m,21} = \left[V_{m,21}^1\ V_{m,21}^2\right]$ where $V_{m,21}^1$ is a $N \times K_1$ matrix and $V_{m,21}^2$ is a $N \times K_m$ matrix with $K_m = K_2$ for Model a and $K_m = K_3$ for Model b, we obtain

$$\begin{bmatrix} V_{m,12}^1 V_{22}^{-1} V_{m,21}^1 & V_{m,12}^1 V_{22}^{-1} V_{m,21}^2 \\ V_{m,12}^2 V_{22}^{-1} V_{m,21}^1 & V_{m,12}^2 V_{22}^{-1} V_{m,21}^2 \end{bmatrix} \begin{bmatrix} \theta_{m1} \\ \theta_{m2} \end{bmatrix} = \begin{bmatrix} V_{m,12}^1 V_{22}^{-1}\mu_2 \\ V_{m,12}^2 V_{22}^{-1}\mu_2 \end{bmatrix}.$$

The first row block gives

$$V_{m,12}^1 V_{22}^{-1} V_{m,21}^1 \theta_{m1} + V_{m,12}^1 V_{22}^{-1} V_{m,21}^2 \theta_{m2} = V_{m,12}^1 V_{22}^{-1}\mu_2.$$

When $\theta_{a2} = 0_{K_2}$ and $\theta_{b2} = 0_{K_3}$, we have

$$\theta_{m1} = \left(V_{m,12}^1 V_{22}^{-1} V_{m,21}^1\right)^{-1} V_{m,12}^1 V_{22}^{-1}\mu_2,\ m = a, b.$$

It follows from $V_{a,12}^1 = E\left(\widetilde{r}_t f_{1t}'\right) = V_{b,12}^1$, that $\theta_{a1} = \theta_{b1}$.

Lemma 11. *Consider two non-nested models a,b with overlapping factors as described in Section 4.2, then we have*

$$\delta_a^2 - \delta_b^2 = [\theta_{a2}'\ \theta_{b2}'] \begin{bmatrix} -C_{a,22}^{-1} & 0 \\ 0 & C_{b,22}^{-1} \end{bmatrix} \begin{bmatrix} \theta_{a2} \\ \theta_{b2} \end{bmatrix}$$

where C_m, $m = a, b$ is defined in (13).

Proof of Lemma 11.

Consider the Model 0 which includes only the factor 1: $y_t = 1 - f_{1t}'\theta_0$. Using Equation (12), we have for $m = a, b$,

$$\begin{aligned}\delta_m^2 - \delta_0^2 &= \mu_2' V_{22}^{-1} \mu_2 - \mu_2' V_{22}^{-1} V_{m,21} (V_{m,12} V_{22}^{-1} V_{m,21})^{-1} V_{m,12} V_{22}^{-1} \mu_2 \\ &\quad - \mu_2' V_{22}^{-1} \mu_2 + \mu_2' V_{22}^{-1} V_{0,21} (V_{0,12} V_{22}^{-1} V_{0,21})^{-1} V_{0,12} V_{22}^{-1} \mu_2 \\ &= \mu_2' V_{22}^{-1} V_{0,21} (V_{0,12} V_{22}^{-1} V_{0,21})^{-1} V_{0,12} V_{22}^{-1} \mu_2 \\ &\quad - \mu_2' V_{22}^{-1} V_{m,21} (V_{m,12} V_{22}^{-1} V_{m,21})^{-1} V_{m,12} V_{22}^{-1} \mu_2.\end{aligned}$$

Now using the fact that $V_{0,21} = V_{m,21}^1 = E\left(\widetilde{r}_t f_{1t}'\right)$ and $V_{m,21} = \left[V_{m,12}^1 \ V_{m,12}^2\right]$, we have

$$V_{0,21}(V_{0,12}V_{22}^{-1}V_{0,21})^{-1}V_{0,12} = V_{m,21}\begin{bmatrix} (V_{m,12}^1 V_{22}^{-1} V_{m,21}^1)^{-1} & 0 \\ 0 & 0 \end{bmatrix} V_{m,12}.$$

Exploiting (48), we have

$$\delta_m^2 - \delta_0^2 = \theta_m'(V_{m,12}V_{22}^{-1}V_{m,21})\begin{bmatrix} (V_{m,12}^1 V_{22}^{-1} V_{m,21}^1)^{-1} & 0 \\ 0 & 0 \end{bmatrix}(V_{m,12}V_{22}^{-1}V_{m,21})\theta_m.$$

Using

$$\begin{aligned}(V_{m,12}V_{22}^{-1}V_{m,21}) &= \begin{bmatrix} V_{m,12}^1 \\ V_{m,12}^2 \end{bmatrix} V_{22}^{-1}\left[V_{m,21}^1 \ V_{m,21}^2\right] \\ &= \begin{bmatrix} V_{m,12}^1 V_{22}^{-1} V_{m,21}^1 & V_{m,12}^1 V_{22}^{-1} V_{m,21}^2 \\ V_{m,12}^2 V_{22}^{-1} V_{m,21}^1 & V_{m,12}^2 V_{22}^{-1} V_{m,21}^2 \end{bmatrix}\end{aligned}$$

and combining the different expressions, we obtain

$$\begin{aligned}\delta_m^2 - \delta_0^2 &= -\theta_{m2}'\left[V_{m,12}^2 V_{22}^{-1} V_{m,21}^2 - V_{m,12}^2 V_{22}^{-1} V_{m,21}^1\left(V_{m,12}^1 V_{22}^{-1} V_{m,21}^1\right)^{-1} V_{m,12}^1 V_{22}^{-1} V_{m,21}^2\right]\theta_{m2} \\ &= -\theta_{m2}' C_{m,22}^{-1}\theta_{m2}.\end{aligned}$$

Therefore, $\delta_a^2 - \delta_b^2 = \delta_a^2 - \delta_0^2 - \delta_b^2 + \delta_0^2 = \theta_{b2}' C_{b,22}^{-1}\theta_{b2} - \theta_{a2}' C_{a,22}^{-1}\theta_{a2}$. This concludes the proof of Lemma 11.

Proof of Proposition 4.

When the models are nested, $\theta_{a2} = 0$ and by Lemma 11, we have $\delta_a^2 - \delta_b^2 = \theta_{b2}' C_{b,22}^{-1}\theta_{b2}$.

Given $C_{b,22}^{-1}$ is full rank, $\delta_a^2 - \delta_b^2 = 0$ if and only if $\theta_{b2} = 0$. This is the first result of Proposition 4.

For Tikhonov, under the hypothesis $\theta_{b2} = 0$, $z = \sqrt{T}V(\hat{\theta}_{b2}^{\gamma})^{-\frac{1}{2}}\hat{\theta}_{b2}^{\gamma} \xrightarrow{d} \mathcal{N}(0, I_{K_2+K_3})$ as T, N and γT go to infinity and $\gamma^2 T$ goes to zero, by Proposition 2.

$T(\hat{\delta}_{a,\gamma}^2 - \hat{\delta}_{b,\gamma}^2) = T\hat{\theta}_{b2}^{\gamma\prime}\hat{C}_{b,22,\gamma}^{-1}\hat{\theta}_{b2}^{\gamma} = z' V(\hat{\theta}_{b2}^{\gamma})^{\frac{1}{2}}\hat{C}_{b,22,\gamma}^{-1}V(\hat{\theta}_{b2}^{\gamma})^{\frac{1}{2}}z$. $\hat{C}_{b,22,\gamma}^{-1}$ converges in probability to $C_{b,22}^{-1}$ as $\hat{C}_{b,22}^{-1}$ is a continuous function of $\hat{C}_b$ which converges to C_b when N, T and $\gamma T \to \infty$ and

$\gamma \to 0$ (see the proof of the consistency of $\hat{\theta}_{HJ}^{\gamma}$). Therefore

$$T(\delta_{a,\gamma}^2 - \delta_{b,\gamma}^2) \xrightarrow{d} z' V(\hat{\theta}_{b2}^{\gamma})^{\frac{1}{2}} C_{b,22}^{-1} V(\hat{\theta}_{b2}^{\gamma})^{\frac{1}{2}} z.$$

The result follows from the singular value decomposition of $V(\hat{\theta}_{b2}^{\gamma})^{-\frac{1}{2}} C_{b,22}^{-1} V(\hat{\theta}_{b2}^{\gamma})^{-\frac{1}{2}}$.

Proof of Proposition 5

The proof of Point 1 follows from Lemma 10.

Now, consider Point 2. For Tikhonov, under the hypothesis $\begin{bmatrix}\theta_{a2}\\ \theta_{b2}\end{bmatrix} = 0$, $z = \sqrt{T} V\left(\begin{bmatrix}\hat{\theta}_{a2}^{\gamma}\\ \hat{\theta}_{b2}^{\gamma}\end{bmatrix}\right)^{-\frac{1}{2}} \begin{bmatrix}\hat{\theta}_{a2}^{\gamma}\\ \hat{\theta}_{b2}^{\gamma}\end{bmatrix} \xrightarrow{d} \mathcal{N}(0, I_{K_3})$ as T, N and $\gamma T \to \infty$, $\gamma^2 T \to 0$, and γ goes to zero. Moreover, by Lemma 11,

$$\begin{aligned}
T(\hat{\delta}_{a,\gamma}^2 - \hat{\delta}_{b,\gamma}^2) &= T \begin{bmatrix}\hat{\theta}_{a2}^{\gamma}\\ \hat{\theta}_{b2}^{\gamma}\end{bmatrix}' \begin{bmatrix} -\hat{C}_{a,22,\gamma}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & \hat{C}_{b,22,\gamma}^{-1}\end{bmatrix} \begin{bmatrix}\hat{\theta}_{a2}^{\gamma}\\ \hat{\theta}_{b2}^{\gamma}\end{bmatrix} \\
&= z' V\left(\begin{bmatrix}\hat{\theta}_{a2}^{\gamma}\\ \hat{\theta}_{b2}^{\gamma}\end{bmatrix}\right)^{\frac{1}{2}} \begin{bmatrix} -\hat{C}_{a,22,\gamma}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & \hat{C}_{b,22,\gamma}^{-1}\end{bmatrix} V\left(\begin{bmatrix}\hat{\theta}_{a2}^{\gamma}\\ \hat{\theta}_{b2}^{\gamma}\end{bmatrix}\right)^{\frac{1}{2}} z.
\end{aligned}$$

The result follows from the singular value decomposition of

$$V\left(\begin{bmatrix}\hat{\theta}_{a2}^{\gamma}\\ \hat{\theta}_{b2}^{\gamma}\end{bmatrix}\right)^{\frac{1}{2}} \begin{bmatrix} -\hat{C}_{a,22,\gamma}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & \hat{C}_{b,22,\gamma}^{-1}\end{bmatrix} V\left(\begin{bmatrix}\hat{\theta}_{a2}^{\gamma}\\ \hat{\theta}_{b2}^{\gamma}\end{bmatrix}\right)^{\frac{1}{2}}.$$

11.6 Proof of Proposition 6

For Ridge, $q_t(\eta_\gamma) = 2y_t \frac{\nu_{1\gamma}' \tilde{r}_t}{N} - \frac{\nu_{1\gamma}' \tilde{r}_t \tilde{r}_t' \nu_{1\gamma}}{N^2} + 2\frac{\nu_{1\gamma}' \mu_2}{N} - \gamma \frac{\nu_{1\gamma}' \nu_{1\gamma}}{N}$.

For Tikhonov, $q_t(\eta_\gamma) = 2y_t \frac{\nu_{1\gamma}' \tilde{r}_t}{N} - \frac{\nu_{1\gamma}' \tilde{r}_t \tilde{r}_t' \nu_{1\gamma}}{N^2} + 2\frac{\nu_{1\gamma}' \mu_2}{N} - \gamma \frac{\nu_{1\gamma}' \Sigma^{-1} \nu_{1\gamma}}{N}$.

We use the following decomposition of $\sqrt{T}\left(\hat{\delta}_\gamma^2 - \delta^2\right)$:

$$\begin{aligned}
\sqrt{T}\left(\hat{\delta}_\gamma^2 - \delta^2\right) &= \sqrt{T}\left(\hat{\delta}_\gamma^2 - \delta_\gamma^2\right) + \sqrt{T}\left(\delta_\gamma^2 - \delta^2\right) \\
&= \sqrt{T}\left(\hat{E}\left[q_t(\widehat{\eta}_\gamma)\right] - \hat{E}\left[q_t(\eta_\gamma)\right]\right) && (49)\\
&+ \sqrt{T}\left(\hat{E}\left[q_t(\eta_\gamma)\right] - E\left[q_t(\eta_\gamma)\right]\right) && (50)\\
&+ \sqrt{T}\left(\delta_\gamma^2 - \delta^2\right) && (51)
\end{aligned}$$

First, we show that the term (49) is negligible. By a fixed point argument, as $\hat{\theta}_\gamma$ minimizes the objective function for a fixed $\widehat{\nu}_{1\gamma}$, we have

$$\hat{E}\left[q_t(\widehat{\eta}_\gamma)\right] \le \hat{E}\left[q_t(\theta_\gamma, \widehat{\nu}_{1\gamma})\right].$$

Therefore,

$$\begin{aligned} \sqrt{T}\left(\hat{E}\left[q_t(\widehat{\eta}_\gamma)\right] - \hat{E}\left[q_t(\eta_\gamma)\right]\right) &\le \sqrt{T}\left(\hat{E}\left[q_t(\theta_\gamma, \widehat{\nu}_{1\gamma})\right] - \hat{E}\left[q_t(\eta_\gamma)\right]\right) \\ &\le \sqrt{T}\nabla_{\nu_1}\hat{E}\left[q_t(\eta_\gamma)\right]'(\hat{\nu}_{1\gamma} - \nu_{1\gamma}) \end{aligned}$$

by the concavity of q_t in $\nu_{1\gamma}$.

The term $\nabla_{\nu_1}\hat{E}\left[q_t(\eta_\gamma)\right]$ is the $N \times 1$ vector corresponding to the derivative of $\hat{E}\left[q_t(\eta_\gamma)\right]$ with respect to $\nu_{1\gamma}$:

$$\nabla_{\nu_1}\hat{E}\left[q_t(\eta_\gamma)\right] = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial q_t(\eta_\gamma)}{\partial \nu_{1\gamma}} = \frac{2}{T}\sum_{t=1}^{T}\left\{y_t\frac{\tilde{r}_t}{N} - \frac{\tilde{r}_t\tilde{r}_t'\nu_{1\gamma}}{N^2} + \frac{\mu_2}{N} - \frac{\gamma}{N}\nu_{1\gamma}\right\}.$$

Therefore, for Ridge and Tikhonov,

$$\begin{aligned} & \sqrt{T}\nabla_{\nu_1}\hat{E}\left[q_t(\eta_\gamma)\right]'(\hat{\nu}_{1\gamma} - \nu_{1\gamma}) \\ = \; & \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left[y_t\frac{\tilde{r}_t}{N} - E[y_t\frac{\tilde{r}_t}{N}]\right]'(\hat{\nu}_{1\gamma} - \nu_{1\gamma}) \qquad (52) \\ & -\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left[\nu_{1\gamma}'\frac{\tilde{r}_t\tilde{r}_t'}{N^2} - E[\nu_{1\gamma}'\frac{\tilde{r}_t\tilde{r}_t'}{N^2}]\right](\hat{\nu}_{1\gamma} - \nu_{1\gamma}). \qquad (53) \end{aligned}$$

We have

$$\begin{aligned} \|\hat{\nu}_{1\gamma} - \nu_{1\gamma}\|_N &= \left\|\hat{\Sigma}_\gamma^{-1}\widehat{\alpha} - \Sigma_\gamma^{-1}\alpha\right\|_N \\ &= \left\|\hat{\Sigma}_\gamma^{-1}\widehat{\alpha} - \Sigma^{-1}\alpha + \Sigma^{-1}\alpha - \Sigma_\gamma^{-1}\alpha\right\|_N \\ &\le \left\|\hat{\Sigma}_\gamma^{-1}\widehat{\alpha} - \Sigma^{-1}\alpha\right\|_N \\ &\quad + \left\|\Sigma^{-1}\alpha - \Sigma_\gamma^{-1}\alpha\right\|_N = o_p(1) \end{aligned}$$

when N, T, $\gamma T \to \infty$ and $\gamma \to 0$. So, (52) is equal to

$$\begin{aligned} \left|\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left(y_t\frac{\tilde{r}_t}{N} - E\left[y_t\frac{\tilde{r}_t}{N}\right]\right)'(\hat{\nu}_{1\gamma} - \nu_{1\gamma})\right| &\le \left\|\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left(y_t\tilde{r}_t - E\left[y_t\tilde{r}_t\right]\right)'\right\|_N \|\hat{\nu}_{1\gamma} - \nu_{1\gamma}\|_N \\ &= o_p(1) \end{aligned}$$

while (53) is bounded by

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\nu_{1\gamma}^{'}\frac{\tilde{r}_t\tilde{r}_t^{'}}{N^2}-E[\nu_{1\gamma}^{'}\frac{\tilde{r}_t\tilde{r}_t^{'}}{N^2}]\right](\hat{\nu}_{1\gamma}-\nu_{1\gamma}) \\
= & \left|\left\langle\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_{1\gamma},\hat{\nu}_{1\gamma}-\nu_{1\gamma}\right\rangle_N\right| \\
\leq & \left\|\sqrt{T}(\hat{\Sigma}-\Sigma)\right\|_{op}\|\nu_{1\gamma}\|_N\|\hat{\nu}_{1\gamma}-\nu_{1\gamma}\|_N \\
\leq & \|\sqrt{T}(\hat{\Sigma}-\Sigma)\|_{op}\left\|\Sigma^{-1}\alpha\right\|_N\|\hat{\nu}_{1\gamma}-\nu_{1\gamma}\|_N \\
= & o_p(1),
\end{aligned}
$$

when N, T, $\gamma T\to\infty$ and $\gamma\to 0$. Therefore, as N, T, $\gamma T\to\infty$ and $\gamma\to 0$, $\sqrt{T}\nabla\hat{E}\left[q_t(\eta_\gamma)\right](\hat{\nu}_{1\gamma}-\nu_{1\gamma})=o_p(1)$ and hence $(49)=o_p(1)$.

Term (50) can be rewritten as

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[2y_t\nu_{1\gamma}^{'}\frac{\tilde{r}_t}{N}-\nu_{1\gamma}^{'}\frac{\tilde{r}_t\tilde{r}_t^{'}}{N^2}\nu_{1\gamma}-E\left[2y_t\nu_{1\gamma}^{'}\frac{\tilde{r}_t}{N}-\nu_{1\gamma}^{'}\frac{\tilde{r}_t\tilde{r}_t^{'}}{N^2}\nu_{1\gamma}\right]\right] &= \\
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[2y_t\nu_1^{'}\frac{\tilde{r}_t}{N}-E\left[2y_t\nu_1^{'}\frac{\tilde{r}_t}{N}\right]\right]& \\
+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[2y_t\frac{\tilde{r}_t^{'}}{N}(\nu_{1\gamma}-\nu_1)-E\left[2y_t\frac{\tilde{r}_t^{'}}{N}(\nu_{1\gamma}-\nu_1)\right]\right]& \\
-\left\langle\sqrt{T}(\hat{\Sigma}-\Sigma)(\nu_{1\gamma}-\nu_1),\nu_{1\gamma}-\nu_1\right\rangle_N& \\
-2\left\langle\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_1,\nu_{1\gamma}-\nu_1\right\rangle_N& \\
-\left\langle\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_1,\nu_1\right\rangle_N.&
\end{aligned}
$$

As N, $T\to\infty$ and $\gamma\to 0$, $\left\langle\sqrt{T}(\hat{\Sigma}-\Sigma)(\nu_{1\gamma}-\nu_1),\nu_{1\gamma}-\nu_1\right\rangle_N\leq\left\|\sqrt{T}(\hat{\Sigma}-\Sigma)\right\|_N\|\nu_{1\gamma}-\nu_1\|_N^2\to 0$ and $\left\langle\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_1,\nu_{1\gamma}-\nu_1\right\rangle_N\leq\|\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_1\|_N\|\nu_{1\gamma}-\nu_1\|_N\to 0$. So (50) is equivalent to $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}2y_t\nu_1^{'}\frac{\tilde{r}_t}{N}+\left\langle\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_1,\nu_1\right\rangle_N$ which converges to a normal distribution using Lemmas 7 and 8. Therefore, Equation (50) converges to a normal distribution with variance

$$
\underset{N,T\to\infty}{lim}\ var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{2y_t\nu_1^{'}\frac{\tilde{r}_t}{N}-\nu_1\frac{\tilde{r}_t\tilde{r}_t^{'}\nu_1}{N^2}-E\left[2y_t\nu_1^{'}\frac{\tilde{r}_t}{N}-\nu_1\frac{\tilde{r}_t\tilde{r}_t^{'}\nu_1}{N^2}\right]\right\}\right].
$$

Finally, regarding (51), we have

$$\begin{aligned}\left|\sqrt{T}\left(\delta_\gamma^2-\delta^2\right)\right|^2 &= T\left\langle\alpha,\left(\Sigma_\gamma^{-1}-\Sigma^{-1}\right)\alpha\right\rangle_N^2 . \\ &\leq T\left\langle\Sigma_\gamma^{-1}\alpha,\left(\Sigma-\Sigma_\gamma\right)\Sigma^{-1}\alpha\right\rangle_N^2 .\end{aligned}$$

For Ridge, we have $T\left\langle\Sigma_\gamma^{-1}\alpha,\left(\Sigma-\Sigma_\gamma\right)\Sigma^{-1}\alpha\right\rangle_N^2 = \gamma^2 T\left\langle\Sigma_\gamma^{-1}\alpha,\Sigma^{-1}\alpha\right\rangle_N^2 \leq \gamma^2 T\left\|\Sigma^{-1}\alpha\right\|_N^2 = O(\gamma^2 T)$.

For Tikhonov, we have

$$\begin{aligned}\left\langle\Sigma_\gamma^{-1}\alpha,\left(\Sigma-\Sigma_\gamma\right)\Sigma^{-1}\alpha\right\rangle_N &= \sum_j\left(\lambda_j-\frac{\lambda_j^2+\gamma}{\lambda_j}\right)\left\langle\phi_j,\Sigma_\gamma^{-1}\alpha\right\rangle_N\left\langle\phi_j,\Sigma^{-1}\alpha\right\rangle_N \\ &= -\gamma\sum_j\frac{1}{\lambda_j}\left\langle\phi_j,\Sigma_\gamma^{-1}\alpha\right\rangle_N\left\langle\phi_j,\Sigma^{-1}\alpha\right\rangle_N \\ \left|\left\langle\Sigma_\gamma^{-1}\alpha,\left(\Sigma-\Sigma_\gamma\right)\Sigma^{-1}\alpha\right\rangle_N\right|^2 &\leq \gamma^2\sum_j\frac{1}{\lambda_j}\left\langle\phi_j,\Sigma_\gamma^{-1}\alpha\right\rangle_N^2\sum_j\frac{1}{\lambda_j}\left\langle\phi_j,\Sigma^{-1}\alpha\right\rangle_N^2 = O(\gamma^2)\end{aligned}$$

as $\alpha\in\Phi_2$. Therefore, $\left|\sqrt{T}\left(\delta_\gamma^2-\delta^2\right)\right|^2 = O(\gamma^2 T)$.

11.7 Proof of Proposition 7

The distribution of the difference of HJ-distances follows from Proposition 6, which gives the distribution for each model.