Hansen-Jagannathan distance with many assets

Marine Carrasco and Cheikh Nokho *
Université de Montréal

October 2024

Abstract

This paper examines the evaluation of asset pricing models with many test assets. The models are specified through a linear stochastic discount factor (SDF). We implement two interpretable regularization schemes to extend the Hansen-Jagannathan distance in a framework of a data-rich environment. These regularizations are shown to yield a relaxation of the Fundamental Equation of Asset Pricing and, therefore, take into account the global misspecified nature of models in finance. We derive the asymptotic properties of the SDF parameter estimator and implement comparison tests of asset pricing models. All results are obtained under the double asymptotic where the number of assets and the number of time series increase to infinity.

*Carrasco gratefully aknowledges partial financial support from FRQSC and SSHRC. Address: Université de Montréal, Departement de Sciences Economiques, CP 6128, succ Centre Ville, Montreal, QC H3C3J7, Canada. Email: marine.carrasco@umontreal.ca.

1 Introduction

Dynamic Asset Pricing Models mainly strive to understand the difference in expected returns among assets. Models differ according to the researcher's systemic risk: for example, CAPM proposes the market portfolio as the main relevant risk factor. Several alternative models (anomalies) have been tested in the literature following the rejection of the CAPM. These models can always be obtained by the relationship between the stochastic discount factor (SDF), pricing kernel, and the proposed risk factors.

A well-known measure of model misspecification is the Hansen-Jagannathan (HJ) distance, which measures the distance between a proposed pricing kernel and the closest valid one (see Hansen and Jagannathan (1997)). The distance is similar to the GMM one except for the weighting matrix which is equal to the inverse of the second moment matrix of the returns. With this distance, the specification test of models (whether the HJ-distance is null) is often rejected (Hodrick and Zhang (2001); Ludvigson (2013)). Therefore, the misspecification of models is usually assumed. In addition, the distance is used to estimate a parameter of the SDF and evaluate whether a risk factor is a priced source of risk.

Even when the models are considered misspecified, one would like to compare the performance of competing asset pricing models. This task is difficult as many asset pricing models seem to perform very well in explaining the well-known 25 portfolios sorted on size (S) and book-to-market (B-M) of Fama and French (1992). As pointed out by Daniel and Titman (2012), this is chiefly due to the characteristics of the formed portfolios which cover a restricted dimension of the returns. Lewellen, Nagel, and Shanken (2010) mention the strong covariance structure of the S/B-M portfolios and suggest increasing the number of test assets, among other recommendations. Kan and Robotti (2009) augment the dataset with the 49 US industry portfolios and compare the HJ distance of several asset pricing models. However, they could not differentiate them due to the high variability of the data.

With the HJ distance, test assets cannot be expanded infinitely without worrying about the weighting matrix. The latter's estimation is quickly unreliable and unstable as the return covariance is near singular or downright non-invertible when the number of assets exceeds the length of the time-series. Cochrane (2005) advanced that a number of assets larger than 1/10 of the time period frequently leads to a near singular covariance matrix. Using this weighting matrix is equivalent to testing asset pricing models with a particular portfolio built from the original returns or test assets. However, a near-singular matrix produces exceptionally leveraged portfolios that are economically not reasonable. Therefore, one ends up focusing on uninteresting portfolios. The situation is exacerbated when, for example, researchers use a considerable amount of individual returns as test assets.

The same issue arises frequently, and the well-known generalized least-squares (GLS) is another example as pointed out by Cochrane (2005). In the presence of heteroscedasticity, OLS estimates are still consistent; however, GLS will be more efficient. Nevertheless, inaccurate estimation or modeling of the errors' covariance matrix leads to a deterioration of the GLS results. Therefore, it is sometimes even better to stop at the OLS level of estimation. Furthermore, standard GMM presents the same issue as Jagannathan, Skoulakis, and Wang (2010) discussed. Therefore, the first step GMM may be more robust than the one with the optimal matrix.

This paper examines the evaluation and comparison of asset pricing models with many test assets, therefore an unstable covariance matrix. First, relying on the inverse problem literature (see Carrasco, Florens, and Renault (2007)), we extend the HJ distance to account for many test assets while assuming that all models are inherently misspecified. Specifically, we implement Tikhonov and Ridge regularizations of the inverse of the covariance matrix in the HJ distance. We show that these regularizations relax the Fundamental Equation of Asset Pricing. In addition, the new misspecification measures can be interpreted as the distance between a proposed pricing kernel and the closest valid SDF pricing returns with controlled errors. All these methods depend on a regularization parameter that controls the level of misspecification. Second, we provide the asymptotic distribution of SDF parameters obtained by minimizing the regularized distance. This permits to determine whether a particular factor is a priced source of risk in the returns and is essential to compare models. In our setting, we allow the number of assets to be higher than the number of time series data. Third, to compare models in the most general manner, we derive the distribution of the regularized distance. All the results are derived under the double asymptotics where the number of assets N and the number of observations T go to infinity simultaneously.

Our work is related 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)). 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. As we are interested in estimating the parameters that minimize the HJ distance under misspecification (pseudo-true value), this paper is also related to Antoine, Proulx, and Renault (2020). However, unlike their approach, we employ the unconditional version of the HJ distance with many assets. Several papers also propose methods to either stabilize or improve the estimation of covariance matrices (Carrasco and Rossi (2016); Carrasco, Kone, and Noumon (2019); Ledoit and Wolf (2003); Ledoit and Wolf (2020)). Kozak, Nagel, and Santosh (2020) consider a model where the factors serve simultaneously as the assets whose returns they are trying to explain and the candidate factors that enter in the SDF. They suppose the number of factors large and propose a Bayesian estimator which has an interpretation in terms of penalization on the SDF coefficients. Our paper has also strong connections with the work of Korsaye, Quaini, and Trojani (2019). They propose a general method of finding a Smart SDF (S-SDF), a strictly positive SDF that tolerates pricing errors for dubious assets. Our method finds the distance between the empirical SDF of the researcher and the S-SDF, without the non-arbitrage constraint. Barillas and Shanken (2018) put forth a method to compare asset pricing models. They also show that returns of the test assets are irrelevant when comparing asset pricing models with just traded factors. However, the test assets become essential when one deals with non-traded factors. In this paper, we are dealing with both types of factors. Finally, as we evaluate models under a misspecified setting, our paper is related to Hall and Inoue (2003) who established the distribution of Generalized Method of Moments (GMM) estimators when moments are misspecified.

The paper is organized as follows. Section 2 presents the framework under which we evaluate models and the issues related to the weighting matrix. Section 3 introduces several regularization methods as well as their interpretations. The section also presents the asymptotic properties of the SDF parameter estimators. Section 4 treats model comparison using regularization, and section 5 contains the results of the simulations. Section 6 compares four empirical asset pricing models using a dataset of 252 portfolios. Finally, section 7 concludes. The proofs are collected in the appendix.

2 Asset pricing model under misspecification

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 must satisfy the fundamental equation of asset pricing: $E[r_t.y_t] = 0$.

Define $Y_t = \begin{bmatrix} f_t \\ r_t \end{bmatrix}$. Its mean and covariance matrix are given by $\mu = E[Y_t] = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and $V = V(Y_t) = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$. We define also $\tilde{r}_t = r_t - \mu_2$ 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 $e(\theta) = E[r_t.y_t(\theta)] = \mu_2 - V_{21}\theta$ via

$$Q_W = e(\theta)' W e(\theta), \tag{1}$$

where W is a positive-definite matrix.

The SDF prices correctly the returns, when one can find θ such that $Q_W(\theta) = 0$. Otherwise, the model is considered globally misspecified.

Remark 1. *The reason for demeaning the factors is the following. When models are misspecified, Proposition 1 of Kan and Robotti (2008) shows that the ranking of asset pricing models using Q_W with raw factors can be altered by performing an affine transformation of the factors. To impose invariance to affine transformations of the factors, one should demean the factors.*

In the particular case, where $W = V_{22}^{-1}$, the covariance of the returns, Q_W is a modified Hansen and Jagannathan (1997) distance, where the mean of the SDF is constrained to 1. Let

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

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

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

θ_{HJ} can also be written as $V_{11}^{-1}(\beta' V_{22}^{-1}\beta)^{-1}\beta' V_{22}^{-1}\mu_2 = V_{11}^{-1}\gamma$ where $\beta = V_{21}V_{11}^{-1}$ is the exposure of the returns to the factors f_t and $\gamma = (\beta' V_{22}^{-1}\beta)^{-1}\beta' V_{22}^{-1}\mu_2$ represents the risk premium. 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 been already established (see (Cochrane, 2005, p. 261), chapter 7 of Ferson (2019) or Goyal (2012)). In this setting, the asset pricing model is misspecified when $e = \mu_2 - V_{21}V_{11}^{-1}\gamma = \mu_2 - \beta\gamma \neq 0$.

We represent a misspecified linear asset-pricing model with SDF $y_t = 1 - \tilde{f}_t^{'}\theta$ by the following formulation

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

where β is a matrix $N \times K$, $e \in \mathbb{R}^N$, $\gamma \in \mathbb{R}^K$, the $N \times 1$ error terms ϵ_t are assumed uncorrelated with the factors. In addition, the errors have mean 0 and variance $V(\epsilon_t \mid f_t) = \Sigma_\epsilon = [\sigma_{i,j}]_{i,j=1,\cdots,N}$ of full rank where $\sigma_{i,j} = E\left[\epsilon_{it}\epsilon_{jt}\right]$. We note $\sigma_i^2 = \sigma_{i,i}$ and $\epsilon = [\epsilon_1, ..., \epsilon_T]'$. Remark that Equation (3) does not impose a factor structure on r_t because the error term ϵ_t is allowed to be serially correlated (see Assumption 2 below). Moreover, the intercept e_i may vary with the asset i.

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 $\bar{f}_t = f_t - \hat{\mu}_1$. $\hat{\mu}_1 = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} f_t$ and $\hat{\mu}_2 = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} r_t$ are respectively the estimators of μ_1 and μ_2.

The SDF parameter is estimated by

$$\hat{\theta}_{HJ} = \hat{V}_{11}^{-1}(\hat{\beta}^{'}\hat{V}_{22}^{-1}\hat{\beta})^{-1}\hat{\beta}^{'}\hat{V}_{22}^{-1}\hat{\mu}_2,$$

and

$$\begin{aligned} \hat{\delta}^2 &= \hat{\mu}_2^{'}\hat{V}_{22}^{-1}\hat{\mu}_2 - \hat{\mu}_2^{'}\hat{V}_{22}^{-1}\hat{V}_{21}(\hat{V}_{12}\hat{V}_{22}^{-1}\hat{V}_{21})^{-1}\hat{V}_{12}\hat{V}_{22}^{-1}\hat{\mu}_2 \\ &= \hat{\mu}_2^{'}\hat{V}_{22}^{-1}\hat{\mu}_2 - \hat{\mu}_2^{'}\hat{V}_{22}^{-1}\hat{\beta}(\hat{\beta}^{'}\hat{V}_{22}^{-1}\hat{\beta})^{-1}\hat{\beta}^{'}\hat{V}_{22}^{-1}\hat{\mu}_2. \end{aligned}$$

Using excess returns, Lemma 4 of Kan and Robotti (2008) gives the asymptotic distribution of $\hat{\theta}_{HJ}$ under a misspecified setting and for N fixed. Specifically,

$$\sqrt{T}(\hat{\theta}_{HJ} - \theta_{HJ}) \to N(0_K, V(\hat{\theta}_{HJ})),$$

where

$$V(\hat{\theta}_{HJ}) = \sum_{j=-\infty}^{\infty} E[q_t q_{t+j}^{'}], \tag{4}$$

$q_t = HV_{12}V_{22}^{-1}(r_t - \mu_2)y_t + H[(f_t - \mu_1) - V_{12}V_{22}^{-1}(r_t - \mu_2)]u_t + \theta_{HJ}$, $H = (V_{12}V_{22}^{-1}V_{21})^{-1}$ and $u_t = e^{'}V_{22}^{-1}(r_t - \mu_2)$.

2.2 Issues with the weighting matrix

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, $W = S^{-1}$, where

$$S = \sum_{j=-\infty}^{\infty} E[(r_t.y_t)\,, (r_{t-j}.y_{t-j})^{'}].$$

However, using this weighting matrix to compare asset pricing models may be misleading for several reasons.

First, in this case, the objective function (1) equates to the over-identification test of Hansen (1996). However, it has been shown that this diagnostic is model-dependent and tends to reward models with volatile SDF and pricing errors as their over-identification statistics tends to be lower (Ludvigson (2013), p.810).

Second, from a perspective of looking at the GMM estimator as a portfolio optimization with the inverse of the eigenvalues of S as weights, it tends to produce huge leverage portfolios as S is near singular with many assets (Cochrane (1996), p. 592).

Other matrices can be used. For example, the inverse of $V_{22} - V_{21}V_{11}^{-1}V_{12}$, the residuals of the regression of r on f, is used in Shanken (1985) and Shanken and Zhou (2007) to estimate the risk premium γ. 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 shown 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. However, V_{22}^{-1} is often near singular as securities are very correlated and N is often large. This singularity may be even higher than that of S. Therefore, it brings forth the same issues as pointed out by p.216 of Cochrane (2005). In addition, near singularity deteriorates the small sample properties of the SDF estimator or misspecification test.

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. 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 Types of regularization

Before introducing the regularization techniques, we introduce several objects to recast the problem as an inverse problem. $\Sigma = \frac{V_{22}}{N} = E\left[\frac{(r_t-\mu_2)(r_t-\mu_2)^{'}}{N}\right] = E\left[\frac{\tilde{r}_t\tilde{r}_t^{'}}{N}\right] = E\left[\frac{\tilde{R}^{'}\tilde{R}}{NT}\right]$ is a $N \times N$ matrix, where $\tilde{r}_t = r_t - \mu_2$ and $\tilde{R} = \left[(r_1-\mu_2)^{'} \quad \cdots \quad (r_T-\mu_2)^{'}\right]^{'}$ is $T \times N$ matrix. We endow $\mathbb{R}^N$ with the norm $\| \phi \|_N^2 = \frac{\phi_1^{'}\phi_2}{N}$ with associated inner product $< \phi_1, \phi_2 >_N = \frac{\phi_1^{'}\phi_2}{N}$, and $\mathbb{R}^T$ with norm $\| v \|_T^2 = \frac{v^{'}v}{T}$ induced by inner product $< v_1, v_2 >_T = \frac{v_1^{'}v_2}{T}$. Let H be the operator from $\mathbb{R}^N$ to $\mathbb{R}^T$ defined by $H\phi = \frac{\bar{R}\phi}{N}$ and H^*, the adjoint of H, i.e. $H^*v = \frac{\bar{R}^{'}v}{T}$, operator from $\mathbb{R}^T$ to $\mathbb{R}^N$. With that, we have the operator $H^*H\phi = \frac{\bar{R}^{'}\bar{R}}{NT}\phi = \hat{\Sigma}\phi$ 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 $H\phi_j = \sqrt{\hat{\lambda}_j}\hat{v}_j$ and $H^*v_j = \sqrt{\hat{\lambda}_j}\hat{\phi}_j$. Note that $\left\{\hat{\lambda}_j, \hat{\phi}_j\right\}$, $j = 1, 2, ..., min(N, T)$ are the non zero eigenvalues and eigenvectors of $\hat{\Sigma}$.

In addition, we define other norms that will be useful in the sequel.

Definition 1.

1. *For a vector* $v \in \mathbb{R}^N$, $\| v \|$ *is the euclidian norm.*
2. *For an arbitrary* $(K \times N)$ *matrix* V, *the operator norm of* V *is* $\| V \| = \sup_{\|\phi\|=1} \| V\phi \|$. *Therefore, for any vector* $u \in \mathbb{R}^N$, $\| Vu \| \leq \| V \| \| u \|$.
3. *Let* $\left\{\phi_j\right\}_{j=1,\cdots,N}$ *be a complete orthonormal basis in* $\mathbb{R}^N$. *For any* $\phi \in R^N$, $\| \phi \|_N^2 = \sum_{i=1}^{N} < \phi, \phi_j >_N^2$ *and if* V *is a* $(N \times N)$ *symmetric matrix, we define the following operator norm* $\| V \|_N = \sup_{\|\phi\|_N=1} < V\phi, \phi >_N$.

4. 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 $u \in \mathbb{R}^N$, $\| Vu \| \leq \| V \|_F \| u \|$.

5. If $\| v \|_N < \infty$ when $N \to \infty$, we note $\| v \|_\infty$ its limit value.

Assumption 1. *(i)*

$$r_t = e + \beta(\tilde{f}_t + \gamma) + \epsilon_t, \tag{5}$$

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

(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) $\| e \|_N = O(1)$.

Remark 2. *The first part of Assumption 1 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 $\| \beta_k \|_\infty < \infty, k = 1, \cdots, K$.*

(ii) imposes that $\| e \|_N$ is bounded, this is a mild condition which is satisfied as soon as each element e_i of e is bounded.

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*

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

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

(ii) V_{11} *is non singular.*

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

Remark 3. *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(iii) implies that $E[\| \epsilon_t \|_N^2] = O(1)$ as $N \to \infty$.*

Lemma 1. *Under Assumptions 1 and 2, for a linear asset pricing model, we have the following results as* $N \to \infty$,

1. $E[\| r_t \|_N^2] = O(1)$.
2. $tr(\Sigma) = O(1)$.

Remark 4. *Lemma 1 indicates that the expected norm of the returns is finite when* N *is large. In addition,* Σ *is trace class, i.e the sum of its eigenvalues is finite. This implies that* Σ *is in the family of Hilbert-Schmidt operators which are compact. The result has several implications. First, the set of eigenvalues is countable and its largest one is bounded (see Theorem 2.39 of Carrasco, Florens, and Renault (2007)). Second, as* $N \to \infty$, *its smallest eigenvalue decreases to* 0.

Let $\alpha > 0$ 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:

1. **Ridge regularization**

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

2. **Tikhonov regularization**

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

For α small, the regularized inverse will be close to the actual inverse while being much more stable. In practice, the tuning parameter α is chosen to go to zero with the sample size. Its choice is discussed later.

Definition 2.

1. *For an operator* $A : G \to E$ *that maps a Hilbert Space* G *(with norm* $\| \, . \, \|_G$*) into a Hilbert Space* E *(with norm* $\| \, . \, \|_E$*), the range,* $\mathcal{R}(A)$, *is the set* $\{\psi \in E : \psi = A\phi$ *for some* $\phi \in G$ *such that* $\| \phi \|_H < \infty\}$.

2. *For a positive self-adjoint compact operator with spectrum* $\left\{\lambda_j, \varphi_j, j = 1, \cdots\right\}$ $\Sigma : G \to G$ *that maps a Hilbert Space* G *(equipped with the inner product* $< . >_G$*) into itself, the* $\omega-$*regularity space of the operator* Σ *, for all* $\omega > 0$, *is*

$$\Phi_\omega = \left\{ \phi : \phi \in G \text{ and } \sum_{j=1}^{\infty} \frac{|< \phi, \varphi_j >_G|^2}{\lambda_j^{2\omega}} < \infty \right\}.$$

3. *The Reproducing Kernel Hilbert Space (RKHS)* $\mathcal{H}(\Sigma)$ *of the operator* Σ *corresponds to* Φ_ω *with* $\omega = \frac{1}{2}$.

Remark 5. Φ_ω *is a decreasing family of subspaces of* $\mathbb{R}^N$ *as* $\omega > 0$ *increases. The regularity space parameter* ω *qualifies the smoothness of* ϕ. *It also permits to characterize the regularization bias.*

Remark 6. *Notice that as* $\hat{\beta} = \frac{\bar{R}' P_F}{T}$, *where* $P_F = \bar{F}(\frac{\bar{F}'\bar{F}}{T})^{-1} = \begin{bmatrix} P_F^1 & \cdots & P_F^K \end{bmatrix}$. *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 of* $\hat{\Sigma}$.

We make a stronger assumption on the β_k and e.

Assumption 3. *(i)* $\beta_k, e \in \Phi_\omega$, *with* $\omega = 3$.

(ii) As $N \to \infty$, $C_\beta = \frac{1}{N}\beta' \Sigma^{-1}\beta =< \Sigma^{-\frac{1}{2}}\beta, \Sigma^{-\frac{1}{2}}\beta >_N \to C$, *where* C *is positive-definite matrix.*

Assumption 3(i) implies that β_k and e belong to the range of Σ^ω so that objects $\Sigma^{-\omega}\beta_k$ and $\Sigma^{-\omega}e$ are well defined even when N goes to infinity.

3.2 Regularization as penalization

This section aims to provide two interpretations of the regularized HJ-distance. One in terms of a penalization on the norm of the pricing error, the other in terms of a penalization on the Lagrange multipliers.

First, recall that as pointed out by Kan and Robotti (2008), δ^2 (not regularized) gives the distance between the proposed SDF y_t and the set of correct SDFs of mean 1 in $\mathcal{M}$, the set of square integrable random variables.

$$\delta^2 = \underset{m_t \in \mathcal{M}, E[m_t]=1}{min} E\left(m_t - y_t\right)^2 \quad \text{subject to } E[m_t r_t] = 0. \tag{6}$$

Below, we are going to show that the regularized δ_α^2 measures how far y is to the closest valid SDF of mean 1 which prices returns with an error controlled by α. To prove it, we make the following assumption.

Assumption 4. $\exists m_0 \in L^2 : E\left[m_0\right] = 1$ *and* $\| E[m_0 r] \|_N^2 < \infty$.

Remark 7. *Assumption 4 guarantees the existence of at least one SDF with finite pricing error.*

Proposition 1. *Under assumption 4, we have the following results:*

1. For ridge,

$$\delta_R^2 = \underset{m \in \mathcal{M}, E[m]=1}{inf} E[(m-y)^2] + \frac{1}{\alpha} \| E[mr] \|_N^2, \tag{7}$$

2. For Tikhonov,

$$\delta_K^2 = \underset{m \in \mathcal{M}, E[m]=1}{inf} E[(m-y)^2] + \frac{1}{\alpha} \| E[mr] \|_{N,\Sigma}^2, \tag{8}$$

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

The previous proposition shows that regularization is equivalent to relaxing the constraint of problem (6). Low values of α put the emphasis on the fundamental equation of asset pricing, while high values allow for possible errors in the pricing of assets.

To get insights on the mecanism behind the penalization in (7), we consider the dual of the optimization problem (7) (see the proof of Proposition 1 in Appendix):

$$\delta_R^2 = \max_{\nu_1 \in \mathbb{R}^N, \nu_2 \in \mathbb{R}} E\left\{2y\nu_1^{'}\frac{r}{N} - \frac{\nu_1^{'}rr^{'}\nu_1}{N^2} - \nu_2^2 - 2\frac{\nu_1^{'}r\nu_2}{N} - \frac{\alpha}{N} \| \nu_1 \|^2\right\}, \tag{9}$$

where ν_1 and ν_2 can be interpreted as Lagrange multipliers where ν_1 is associated with the condition $\| E[mr] \|_N^2 < c$ for some constant c and ν_2 with the condition $E(m) = 1$. Equation (9) is the penalized version of the dual of (6) with a penalization applied to ν_1. The first order condition with respect to ν_1 gives

$$e - \frac{E(rr')}{N}\nu_1 - \nu_2 - \alpha\nu_1 = 0. \tag{10}$$

The first order condition with respect to ν_2 gives $\nu_2 = -\nu_1'\mu_2/N$. Replacing ν_2 in Equation (10) gives

$$\nu_1 = (\Sigma + \alpha I)^{-1}e = \Sigma_\alpha^{-1}e.$$

And the solution is the ridge regularized Hansen-Jagannathan distance

$$\delta_R^2 = \frac{e'\Sigma_\alpha^{-1}e}{N}.$$

For Tikhonov regularization, the dual is given by

$$\delta_K^2 = \max_{\nu_1 \in \mathbb{R}^N, \nu_2 \in \mathbb{R}} E\left\{2y\nu_1^{'}\frac{r}{N} - \frac{\nu_1^{'}rr^{'}\nu_1}{N^2} - \nu_2^2 - 2\frac{\nu_1^{'}r\nu_2}{N} - \frac{\alpha}{N}\nu_1'\Sigma^{-1}\nu_1\right\}.$$

Solving in ν_1 and ν_2 yields $\nu_1 = [\Sigma^2 + \alpha I_N]^{-1}\Sigma E[ry] = [\Sigma^2 + \alpha I_N]^{-1}\Sigma e$ and

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

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. In the extreme case where $\alpha \to \infty$, $\nu_1 = 0$ and no restriction on $E(mr)$ is imposed. In the other extreme where $\alpha = 0$, the condition $E(mr) = 0$ is strictly enforced at the risk of getting an unstable solution involving Σ^{-1}.

Now we investigate how Tikhonov regularization acts on the constraint. As assets with very low eigenvalues tend to have abnormally bigger weights in the HJ-distance, the Tikhonov regularization induces a rebalancing of the weights. Using the diagonalization of $\Sigma = P^{'}\Lambda P$, where P is the matrix of eigenvectors and Λ, the matrix of eigenvalues λ_j, we can rewrite the penalization as follows:

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

where $\omega_j = \dfrac{\lambda_j}{\alpha}$. $(Pr)_j$ can be interpreted as the principal component of r. The Tikhonov penalization entails the repackaging of the assets into N portfolios $(Pr)_j$ with weights given by ω_j. The lower the eigenvalues λ_j is, the lower the contribution of asset $(Pr)_j$ to the minimization, and vice-versa.

Korsaye, Quaini, and Trojani (2019) propose a Smart SDF (S-SDF), M. The latter is a non-negative random variable that tolerates pricing errors for $D \in \mathbb{N}$ dubious assets (R_d) while pricing correctly $S \in \mathbb{N}$ sure assets (R_s).

$$E[MR_s] - q_s = 0_N \text{ and } h(E[MR_d] - q_d) \leq \tau,$$

where $\tau > 0$ and $h : \mathbb{R}^D \rightarrow [0, +\infty]$ is a closed and convex pricing function. q_s and q_d are the prices of the sure and dubious assets. Such SDF always exists in an arbitrage-free economy with frictions. In the search of a minimum dispersion S-SDF, the latter materialized itself as a penalization of the portfolio weights of the dubious assets in the dual portfolio problem. This penalization represents transaction costs which equal to the minimum execution cost for buying the dubious assets.

Remark that it is equivalent to penalize the norm $\|E\,[mr]\|^2$ in (7) or to impose a constraint of the form $\|E\,[mr]\|^2 \leq \tau$ so our approach is very similar to that of Korsaye, Quaini, and Trojani (2019). However, we do not impose $M \in L^2_+$, i.e non-negative L^2 random variable and we allow for a double asymptotic where both N and T go to infinity.

3.3 Asymptotic distribution of the regularized SDF parameter of misspecified models

For any regularization schemes, the estimator of θ_{HJ} is given by

$$\hat{\theta}^{\alpha}_{HJ} = \underset{\theta}{argmin}(\hat{\mu}_2 - \hat{V}_{21}\theta)^{'}\hat{\Sigma}^{-1}_{\alpha}(\hat{\mu}_2 - \hat{V}_{21}\theta). \tag{11}$$

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

and the regularized HJ-distance is

$$\hat{\delta}_\alpha^2 = \frac{\hat{\mu}_2' \hat{\Sigma}_\alpha^{-1} \hat{\mu}_2}{N} - \frac{\hat{\mu}_2' \hat{\Sigma}_\alpha^{-1} \hat{V}_{21}}{N} \left(\frac{\hat{V}_{12} \hat{\Sigma}_\alpha^{-1} \hat{V}_{21}}{N} \right)^{-1} \frac{\hat{V}_{12} \hat{\Sigma}_\alpha^{-1} \hat{\mu}_2}{N}$$

$$= \frac{\hat{\mu}_2' \hat{\Sigma}_\alpha^{-1} \hat{\mu}_2}{N} - \frac{\hat{\mu}_2' \hat{\Sigma}_\alpha^{-1} \hat{\beta}}{N} \left(\frac{\hat{\beta}' \hat{\Sigma}_\alpha^{-1} \hat{\beta}}{N} \right)^{-1} \frac{\hat{\beta}' \hat{\Sigma}_\alpha^{-1} \hat{\mu}_2}{N}.$$

$\hat{\Sigma}_\alpha^{-1}$ 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

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

where $\bar{\epsilon} = \frac{1}{T} \sum_{t=1}^{T} \epsilon_t$ and $\hat{\theta}_{HJ}^\alpha$ can be decomposed as such

$$\begin{aligned} \hat{\theta}_{HJ}^\alpha - \theta_{HJ} = & (\hat{V}_{11}^{-1} - V_{11}^{-1})\gamma + \hat{V}_{11}^{-1}(\hat{\mu}_1 - \mu_1) \\ & + \hat{V}_{11}^{-1}(\hat{\beta}' \hat{\Sigma}_\alpha^{-1} \hat{\beta})^{-1} [\, \hat{\beta}' \hat{\Sigma}_\alpha^{-1} (\beta - \hat{\beta})(\gamma + \hat{\mu}_1 - \mu_1) \\ & + \hat{\beta}' \hat{\Sigma}_\alpha^{-1} e + \hat{\beta}' \hat{\Sigma}_\alpha^{-1} \bar{\epsilon} \,]. \end{aligned} \tag{12}$$

Equivalence between Ridge and Tikhonov.

Because $\hat{\beta}$ depends on r, it is possible to rewrite Ridge as Tikhonov regularization. Ridge regularization gives

$$\begin{aligned} \hat{\Sigma}_\alpha^{-1} \hat{\beta} &= \left(\frac{\overline{R}' \overline{R}}{NT} + \alpha I_N\right)^{-1} \frac{\overline{R}' \overline{F}}{T} \\ &= \sum_{j=1}^{min(N,T)} \frac{q(\alpha, \sqrt{\lambda_j})}{\sqrt{\hat{\lambda}_j s}} < \overline{F}, \phi_j >_N \phi_j, \end{aligned}$$

where $q(\mu) = \frac{\mu}{\alpha + \mu}$ and $\{\lambda_j, \phi_j\}$ are the eigenvalues and eigenvectors of Σ (see Appendix for more details). Tikhonov regularization gives the same formula but with $q(\alpha, \sqrt{\lambda_j})$ replaced by $q(\alpha, \lambda_j)$. So both regularizations give basically the same results (the only difference is that the optimal rate for α may be different). For this reason, we focus on Tikhonov regularization. From now on, $\hat{\theta}_{HJ}^\alpha$ and $\hat{\delta}_\alpha^2$ correspond to the estimators obtained by Tikhonov regularization.

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

Assumption 5. *For $\rho > 0$ defined in Assumption 2 (i), we assume:*

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

(ii) $\lim_{N\to\infty} E[\| \epsilon_t \|_N^{4+2\rho}] < \infty$.

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

(iv) $0 < \lim_{N,T\to\infty} Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} < r_t, \Sigma^{-1}e >_N) < \infty$.

(v) $0 < \lim_{N,T\to\infty} Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} < \Sigma^{-1}\beta, r_t >_N) < \infty$.

(vi) $0 < \lim_{N,T\to\infty} Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} < \epsilon_t, \Sigma^{-1}e >_N) < \infty$.

Proposition 2. *Suppose Assumptions 1-5 are satisfied.*

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

1. $\hat{\theta}_{HJ}^{\alpha} \xrightarrow{P} \theta_{HJ}$

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

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

$h_t = \widetilde{f}_t y_t + \gamma + C_\beta^{-1}\frac{\beta'\Sigma^{-1}}{N}(\epsilon_t y_t - \tilde{r}_t\tilde{u}_t + e) + C_\beta^{-1}V_{11}^{-1}\widetilde{f}_t\frac{\epsilon_t'\Sigma^{-1}e}{N}$, *and* $\tilde{u}_t = \frac{\tilde{r}_t'\Sigma^{-1}e}{N}$.

Proposition 2 can also be used when the model is correctly specified by setting $e = 0$. *Ω* can be estimated using the HAC estimator of Andrews (1991). The results of Proposition 2 are keys to compare competing asset pricing models.

The regularization parameter must be chosen in a way such that the bias vanishes as $T \to \infty$. In general, as T and N go to ∞, if $\alpha \sim \frac{1}{T^\kappa}$, $\kappa \in]\frac{1}{2}; 1[$, the rates of convergence of Proposition 2 are satisfied. In practice, we let the data choose α.

3.4 Choice of the regularization parameter

We rely on a data-driven approach to choose the regularization parameter α. For a given sample size T, we divide the historic data in two parts. We use the first part to estimate γ and employ it to predict returns in the second part. We choose a that maximizes the out-of-sample R-square, R_{oos}^2.

$$R_{oos}^2 = 1 - \frac{(\mu_2^o - \beta^o\hat{\gamma}_\alpha)'(\mu_2^o - \beta^o\hat{\gamma}_\alpha)}{\mu_2^{o'}\mu_2^o}, \tag{13}$$

where quantity with $.^o$ are estimated from the withheld sample.

4 Tests of equality of HJ distance of two asset pricing models

We compare two competing models (Models 1 and 2) using their regularized HJ distances. Their SDFs are defined as $y_{1t}(\eta) = 1 - (x_{1t} - E[x_{1t}])'\theta_1$ and $y_{2t}(\lambda) = 1 - (x_{2t} - E[x_{2t}])'\theta_2$. $x_{1t} = [f_{1t}', f_{2t}']'$ and $x_{2t} = [f_{1t}', f_{3t}']'$ are two sets of factors, that are used in Model 1 and Model 2, respectively. f_{it} is of dimension $K_i \times 1, : i = 1, 2, 3$. $\theta_1 = [\theta_{11}', \theta_{12}']'$ and $\theta_2 = [\theta_{21}', \theta_{22}']'$.

The two corresponding pricing models are respectively

$$r_t = e_1 + \beta_1(x_{1t} - E[x_{1t}] + \gamma_1) + \epsilon_{1t}, \tag{14}$$

and

$$r_t = e_2 + \beta_2(x_{2t} - E[x_{2t}] + \gamma_2) + \epsilon_{2t},$$

with $\theta_1 = \beta_1\gamma_1$ and $\theta_2 = \beta_2\gamma_2$. e_m represents the vector of pricing errors of model $m = 1, 2$. We note $\delta_m^2, m = 1, 2$ the HJ distances of the two models.

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

Model m is estimated using solely factors in x_m.

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. We will treat the nested and nonnested cases separately.

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 of two models as pointed out by Kan and Robotti (2009). We define $C_2 = \left(V_{12,2} V_{22}^{-1} V_{21,2}\right)^{-1}$ and partition it as below

$$C_2 = \begin{bmatrix} C_{2,11} & C_{2,12} \\ C_{2,21} & C_{2,22} \end{bmatrix}.$$

We assume $C_{2,22}^{-1}$ is a full rank matrix. Kan and Robotti (2009) shows that the difference of HJ distances $(\delta_1^2 - \delta_2^2)$ between the two models is equal to

$$\delta_1^2 - \delta_2^2 = \theta_{22}' C_{2,22}^{-1} \theta_{22}. \tag{15}$$

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 3. *Suppose Assumption 1-5 are satisfied. We have the following results:*

1. $\delta_1^2 = \delta_2^2$ *if and only if* $\theta_{22} = 0_{K_3}$

2. Under the hypothesis $\theta_{22} = 0_{K_3}$*, as* T*,* N *go to infinity and* α *goes to zero, if* α *is chosen such that* $\alpha T \to \infty$ *and* $\alpha^2 T \to 0$*,*

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

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

Remark 8. *Proposition 3 implies that we can perform two kinds of tests to compare Model 1 with factor* f_1 *and Model 2 with factors* f_1 *and* f_3*. On the one hand, we can focus on the SDF parameter* λ_2 *and test* $H_0 : \lambda_2 = 0_{K_3}$ *using Proposition 2 in a framework where returns are governed by (14). On the other hand, we can compute the HJ distance difference of the two models using the same level of penalization or (15) and use the statistics* $T(\hat{\delta}^2_{1,\alpha} - \hat{\delta}^2_{2,\alpha})$ *to compare them.*

The coefficients ξ_j *are all nonnegative, hence the test presented in Proposition 3 is a one-sided test.*

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 $y_1 \neq y_2$ but $\delta_1^2 = \delta_2^2$. Both cases need to be treated separately.

4.2.1 Test of SDFs equality

In this section, we test the equality of the SDFs, $y_1 = y_2$. Given the models are non-nested, the equality of SDFs can be achieved only if the both SDF depend on f_1 only. Consider $C_1 = \left(V_{12,1} V_{22}^{-1} V_{21,1}\right)^{-1}$, partition it as below

$$C_1 = \begin{bmatrix} C_{1,11} & C_{1,12} \\ C_{1,21} & C_{1,22} \end{bmatrix},$$

and assume $C_{1,22}$ is a full rank matrix. The difference between the HJ distances is

$$\delta_1^2 - \delta_2^2 = -\theta_{12}' C_{1,22}^{-1} \theta_{12} + \theta_{22}' C_{2,22}^{-1} \theta_{22}.$$

The following proposition outlines the main result.

Proposition 4. *Suppose Assumption 1-5 are satisfied. We have the following result:*

1. $y_1 = y_2$ *if and only if* $\theta_{12} = 0_{K_2}$ *and* $\theta_{22} = 0_{K_3}$ *and*

2. For Tikhonov, under the hypothesis $\theta_{12} = 0_{K_2}$ *and* $\theta_{22} = 0_{K_3}$ *, as* T*,* N *go to infinity and* α *goes to zero, if* α *is such that* $\alpha T \to \infty$ *and* $\alpha^2 T \to 0$*,*

$$T(\hat{\delta}^2_{1,\alpha} - \hat{\delta}^2_{2,\alpha}) \xrightarrow{d} \sum_{i=1}^{K_3} \xi_i \chi_i^2(1) \tag{16}$$

where ξ_i *are the eigenvalues of* $V\left(\begin{bmatrix}\hat{\theta}_{12}\\ \hat{\theta}_{22}\end{bmatrix}\right)^{\frac{1}{2}} \begin{bmatrix}-C_{1,22}^{-1} & 0_{K_2\times K_3}\\ 0_{K_3\times K_2} & C_{2,22}^{-1}\end{bmatrix} V\left(\begin{bmatrix}\hat{\theta}_{12}\\ \hat{\theta}_{22}\end{bmatrix}\right)^{\frac{1}{2}}$*, and* $\chi_i^2(1)$ *are independent* $\chi^2(1)$ *random variables.*

Remark 9. *Proposition 4.1. shows that to compare asset pricing models with overlapping factors, one can test the simultaneous nullity of the coefficients of the common factors (*θ_{12} *and* θ_{22}*). In our regularized setting, each parameter can be estimated separately. Their variances given in 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. We can also realize a test based on the HJ difference using the result (16). The* ξ_i *may be positive or negative, hence this test is a two-sided test. Moreover, it may lack of power against certain alternatives, contrarily to the Wald test which is a consistent test.*

4.2.2 Comparison of non-nested models with distinct SDFs

To compare two non-nested models with distinct SDFs ($y_1 \neq y_2$), one has to rely on the distribution of the aggregate pricing errors or δ^2 under misspecification. Hansen, Heaton, and Luttmer (1995) and Kan and Robotti (2008) have already given the distribution of the HJ distance and the modified HJ distance when models are misspecified.

Specifically Hansen, Heaton, and Luttmer (1995) showed, in the case of gross returns, that when $\delta \neq 0$

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

where $v_1 = var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} q_t)$, $q_t = y_t^2 - (y_t - \nu' r_t^g)^2 - 2\nu' 1_N - \delta^2$, and r_t^g is a $N \times 1$ vector of gross returns. The term ν is the Lagrange multiplier ($\nu = E[r_t^g r_t^{g'}]^{-1}(E[r_t^g y_t] - 1_N)$) of the unconstrained HJ distance saddle problem of Hansen and Jagannathan (1997).

Kan and Robotti (2008) adapted the results for the case of excess returns. They showed that the modified HJ distance has the following distribution

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

where $v_2 = var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} q_t^m)$ 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 (6).

It is worth noticing that the distribution of the distance does not need to take into account 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. To do so, we exploit the following expression of the penalized HJ distance (δ_p^2) derived from Section 3.2:

$$\delta_p^2 = \underset{\nu_1 \in \mathbb{R}^N}{max} E\left[q_t^P(\nu_1)\right],$$

where $q_t^P(\nu_1) = y_t^2 - (y_t - \nu_1^{'}(\frac{r_t}{N} - \frac{\mu_2}{N}))^2 + 2\nu_1^{'}\frac{\mu_2}{N} + \psi(\nu_1)$ and $\psi : \mathbb{R}^N \to \mathbb{R}$ is a concave function representing the penalty, namely $\psi(\nu_1) = -\alpha \parallel \nu_1 \parallel_N^2$ for Ridge and $\psi(\nu_1) = -\alpha \parallel \nu_1 \parallel_{N,\Sigma^{-1}}^2$ for Tikhonov.

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

The previous assumption ensures the interchangeability between integration and differentiation for any $0 < \alpha < \infty$.

Proposition 5. *Let $\hat{\delta}_\alpha^2$ be the regularized Hansen-Jagannathan distance with Ridge or Tikhonov regularization. Suppose Assumption 1-5 are satisfied and $\delta \neq 0$. As T, N go to infinity and α goes to zero, $\alpha T \to \infty$, and $\alpha^2 T \to 0$,*

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

where

$$v_4 = \underset{N,T\to\infty}{lim} \sum_{j=-\infty}^{\infty} E\left(l_t l_{t-j}\right),$$

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

Proposition 5 gives the distribution of the penalized HJ distance using the errors. The asymptotic variance v_4 can be estimated using a HAC estimator and replacing l_t by $\widehat{l_t}$:

$$\widehat{l_t} = 2y_t\widehat{u}_t - \widehat{u}_t^2 - \hat{\delta}^2 + 2\frac{\hat{\nu}_1^{'}\hat{\mu}_2}{N} + \psi\left(\hat{\nu}_1\right)$$

where $\widehat{u}_t = \widehat{\nu}_1^{\prime}\widetilde{r}_t/N$, $\hat{\nu}_1 = \Sigma_\alpha^{-1}\widehat{e}$, and $\widehat{e} = \sum_t r_t y_t/T$.

It can be used to compare two asset pricing models as presented in the following proposition.

Proposition 6. *Let $\hat{\delta}^2_\alpha$ be the regularized Hansen-Jagannathan distance with Ridge or Tikhonov regularization. Suppose Assumption 1-5 are satisfied, $y_1 \neq y_2$, and $\delta_1^2, \delta_2^2 \neq 0$. As T, N go to infinity and α goes to zero, if α is chosen such that $\alpha T \to \infty$ and $\alpha^2 T \to 0$,*

$$\sqrt{T}\left((\hat{\delta}^2_{1\alpha} - \hat{\delta}^2_{2\alpha}) - (\delta_1^2 - \delta_2^2)\right) \xrightarrow{d} \mathcal{N}(0, v_5),$$

where

$$v_5 = \underset{N,T\to\infty}{lim}\ var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\tilde{l}_{1t} - \tilde{l}_{2t}\right)\right],$$

where $\tilde{l}_{\mathcal{M},t} = 2y_{\mathcal{M},t}\nu_1^{'}\frac{\tilde{r}_t}{N} - \nu_{\mathcal{M},1}^{'}\frac{\tilde{r}_t\tilde{r}_t^{'}}{N^2}\nu_{\mathcal{M},1} - E\left[2y_{\mathcal{M},t}\nu_1^{'}\frac{\tilde{r}_t}{N} - \nu_{\mathcal{M},1}^{'}\frac{\tilde{r}_t\tilde{r}_t^{'}}{N^2}\nu_{\mathcal{M},1}\right]$ for $\mathcal{M} = 1, 2$. $y_{\mathcal{M},t}$ is the SDF of model $\mathcal{M}$ and $\nu_{\mathcal{M},1} = \Sigma^{-1}e_{\mathcal{M}}$,where $e_{\mathcal{M}}$ represents the pricing errors of model $\mathcal{M}$.

When using Proposition 6 to compare two asset pricing models, one should use the same value of the regularization parameter. We can use Proposition 6 to construct a Wald test of $H_0 : \delta_1^2 = \delta_2^2$, $W = T\left(\hat{\delta}^2_{1\alpha} - \hat{\delta}^2_{2\alpha}\right)^2 \widehat{v}_5^{-1}$, where $\widehat{v}_5$ is a HAC estimator of v_5.

4.3 Multiple comparison

In this section, we present a comparison test of multiple models. The test is based on the work of Wolak (1989), see also Gospodinov, Kan, and Robotti (2013). Suppose we have $p + 1$ models with HJ distance given by $\delta_i, \quad i = 1, \dots, p+1$. We are interested in testing whether a benchmark model (model 1) has an aggregate pricing errors as low as the other p models. Let $d_i = \delta_1^2 - \delta_i^2,\ i = 2, \dots, p+1$ be the difference between the HJ distance of the benchmark and the remaining models and $d = \begin{pmatrix} d_2 & \cdots & d_{p+1}\end{pmatrix}$. The null hypothesis of the test is $H_0 : d \leq 0_p$. To have the same framework as Wolak (1989), we rely on the fact that as $N, T \to \infty$ and $\alpha \to 0$,

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

using Proposition 6. The latter is valid only when the models are misspecified, that is $\delta_i > 0$, and the models have distinct SDFs. The test uses the sample counterpart of d, $\hat{d}_\alpha = \begin{pmatrix} d_{1\alpha} & \cdots & d_{p+1\alpha}\end{pmatrix}$ for a given value of α. Let $\tilde{d}_\alpha$ be the optimal solution of the following quadratic programming problem

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

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

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

The distribution of the previous statistics is obtained under the least favorable value, i.e.

$d = 0$. We have $LR_\alpha \xrightarrow{d} \sum_{i=0}^{p} w_{p-i}(\Omega_d)\chi^2(i)$, where the weights w_i sum up to one[1] and $\chi^2(i)$ are independent Chi-square random variables with i degrees of freedom.

5 Monte Carlo Simulations

In this section, we run several Monte Carlo simulations to showcase the value of the regularization schemes described previously. We describe the approach used to generate misspecified linear asset pricing model with parameters calibrated to data. We generate the excess returns and factors from a multivariate normal distribution with mean μ and covariance V, where $\mu = E\begin{bmatrix} f_t \\ r_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and $V = Var\begin{bmatrix} f_t \\ r_t \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$. Without loss of generality we set $\mu_1 = 0$. We use the framework of Gospodinov, Kan, and Robotti (2013) and choose μ_2 such that the model is misspecified. The pseudo-true SDF parameter θ_{HJ} associated with the SDF $y_t = 1 - f_t'\theta$ is given by

$$\theta_{HJ} = (V_{21}'V_{22}^{-1}V_{21})^{-1}V_{21}'V_{22}^{-1}\mu_2.$$

So, we have the following first-order condition $V_{21}'V_{22}^{-1}(V_{21}\theta_{HJ} - \mu_2) = 0$. We set $\mu_2 = V_{21}\theta_{HJ} + z$, where z is $N \times 1$ vector of constants. This implies that the first order condition is $V_{21}'V_{22}^{-1}z = 0$. A convenient choice of z is $\hat{e} = \hat{\mu}_2 - \hat{V}_{21}'(\hat{V}_{21}'\hat{V}_{22}^{-1}\hat{V}_{21})^{-1}\hat{V}_{21}'\hat{V}_{22}^{-1}\hat{\mu}_2$ because $\hat{V}_{21}'\hat{V}_{22}^{-1}\hat{e} = 0$.

Without loss of generality, assume that $f_t = \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix}$, where f_{1t} and f_{2t} are $K_1 \times 1$ and $K_2 \times 1$ vector with $K_1 + K_2 = K$. In order to verify the size of the test $H_0 : \theta_{HJ,1} = 0_{K_1}$, where $\theta_{HJ,1}$ is the SDF parameter of the first K_1 factors, we can choose

$$\theta_{HJ} = \begin{bmatrix} 0_{K_1} \\ (V_{21,c}'V_{22}^{-1}V_{21,c})^{-1}V_{21,c}'V_{22}^{-1}\mu_2 \end{bmatrix}.$$

In the previous expression, $V_{21,c} = E[r_t f_{2t}']$ is a $N \times K_2$ matrix.

The parameters of the generated returns μ_2 and V are calibrated using a monthly dataset of 252 combined portfolios going from 1964 to 2019 extracted from the Kenneth French's Website. We remove portfolios with missing values. The portfolios list is presented in Table 8 in Appendix.

5.1 SDF parameter estimates

In this section, we analyze the small sample properties of the SDF parameter test. In the latter, we are interested in testing whether a particular factor is priced in the returns (similar to a t-test).

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

This corresponds to testing whether a SDF parameter is null. We compare the small-samples size properties of our test with the one in Kan and Robotti (2008) using (4).

We simulate the three factor model of Fama and French (1993) (FF3), where the risk factors are the market excess return (r_{mkt}), the return difference between portfolios of small and large stocks (r_{SMB}), and the return difference between portfolios of high and low book-to-market ratios (r_{HML}). The SDF is written as below

$$y = 1 - \theta_{mkt}\left(r_{mkt} - E\left[r_{mkt}\right]\right) - \theta_{SMB}\left(r_{SMB} - E\left[r_{SMB}\right]\right) - \theta_{HML}\left(r_{HML} - E\left[r_{HML}\right]\right).$$

We also simulate the durable consumption CAPM (DCCAPM) of Yogo (2006) with the excess market return, the log consumption growth rate of non-durable goods (Δc_{ndur}) and the log consumption growth rate of the stock of durable goods (Δc_{dur}) as risk factors. The SDF of the model is

$$y = 1 - \theta_{mkt}\left(r_{mkt} - E\left[r_{mkt}\right]\right) - \theta_{ndur}\left(\Delta c_{ndur} - E\left[\Delta c_{ndur}\right]\right) - \theta_{dur}\left(\Delta c_{dur} - E\left[\Delta c_{dur}\right]\right).$$

Finally, we simulate a polynomial type of model used in Dittmar (2002). The SDF of the model is given by

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

For each model, we ran the following simulation: we generate data with expected return such that the model is misspecified, and one of the factors is not priced and estimate a full model with it. After running 10000 simulations, we compute the empirical level and power of the test. We set $N = 251$ and $T = 150$, 350, and 650. For all the models, the theoretical HJ distance is around 1.02.

Table 1 reports the empirical size of the SDF parameter test using the approach of Kan and Robotti (2008). We use the Moore-Penrose inverse of the covariance matrix when $N > T$. For the FF3 model, we noticed that the SDF parameter of the factors keep their theoretical size. For the durable consumption CAPM, the tests concerning the macroeconomic factors represented by the durable and nondurable consumption growth rate are oversized for all values of T. The same size distortion is observed for the polynomial model. The over-rejection of the macroeconomic variable is pervasive (see Gospodinov, Kan, and Robotti (2014)). Therefore, we can conclude that taking the generalized inverse does not guarantee appropriate test behavior when N is large.

Table 1: Empirical size of Kan and Robotti (2008) test with 252 assets

T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Fama-French three factors model									
	θ_{mkt}			θ_{SMB}			θ_{HML}		
150	0.101	0.048	0.010	0.102	0.050	0.010	0.104	0.051	0.010
350	0.094	0.048	0.009	0.098	0.048	0.009	0.098	0.048	0.009
650	0.095	0.049	0.009	0.097	0.048	0.009	0.097	0.048	0.009
Panel B: Linear durable consumption CAPM of Yogo (2006)									
	θ_{mkt}			θ_{ndur}			θ_{dur}		
150	0.133	0.072	0.017	0.478	0.395	0.256	0.474	0.394	0.253
350	0.126	0.067	0.017	0.269	0.185	0.082	0.267	0.182	0.078
650	0.109	0.056	0.012	0.131	0.070	0.016	0.134	0.072	0.020
Panel C: Nonlinear model of Dittmar (2002)									
	θ_{mkt}			$\theta_{mkt,2}$			$\theta_{mkt,3}$		
150	0.371	0.28	0.154	0.458	0.373	0.235	0.453	0.371	0.229
350	0.241	0.16	0.06	0.248	0.172	0.069	0.255	0.175	0.075
650	0.151	0.087	0.021	0.132	0.069	0.017	0.136	0.072	0.019

We use the Tikhonov regularization, through Proposition 2, to implement our t-test. We choose the value of alpha, between 0.001 and 0.1, which maximizes the out-of-sample R^2 : we use half of the sample as training data and the remaining as test data. Particularly, we choose the smallest value of α for $T = 650$. Table 1 presents the empirical size of the t-test for the factors in each simulated model. For FF3 (Panel A), we notice that the rejection rate is always close to their theoretical level. In addition, the Tikhonov regularization is able to correct the over-rejection of the t-test in the consumption (Panel B) and nonlinear model (Panel C).

We now turn our attention to the empirical power of our t-test. Table 2 presents the rejection rate of the factors when their SDF parameter is non null. For the FF3 (Panel A), the rejection rate of the market (r_{mkt}) and HML factor reach more than 50% when $T = 350$. The power is approaching 1 when $T = 650$. However, the SMB factor requires much more time series data to reach an acceptable power level, still lower than the level seen with the market and the value factor. For the durable consumption CAPM (Panel B), except for the market factor, power is lower compared to the FF3. The market factor has a higher rejection rate than the macroeconomic factors.

The low power can be attributed to the strength of the factor, i.e. the number of portfolios' returns significantly correlated with the factor. A low correlation between factor and returns induces a low β and a bigger variance through the inverse of $\beta' \Sigma^{-1} \beta$. Using an average of 442 individual securities and 145 factors, Bailey, Kapetanios, and Pesaran (2021) show that more than 60 percent of the factors are not significantly correlated to more than 55 percent of the securities. This aspect needs to be taken into account in future work.

For the cubic model (panel C), the rejection rate of the t-test is better than in the consumption model. The market has the highest power followed by its square and cubic counterpart.

Table 2: Empirical size of Tikhonov test under misspecification with 252 assets

T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Fama-French three factors model									
	θ_{mkt}			θ_{SMB}			θ_{HML}		
150	0.099	0.047	0.008	0.109	0.056	0.010	0.100	0.049	0.010
350	0.102	0.051	0.010	0.125	0.068	0.016	0.111	0.059	0.013
650	0.100	0.051	0.009	0.114	0.060	0.013	0.123	0.070	0.015
Panel B: Linear durable consumption CAPM of Yogo (2006)									
	θ_{mkt}			θ_{ndur}			θ_{dur}		
150	0.084	0.037	0.006	0.053	0.021	0.002	0.056	0.022	0.002
350	0.095	0.044	0.008	0.087	0.041	0.007	0.084	0.039	0.007
650	0.111	0.059	0.011	0.078	0.036	0.005	0.081	0.038	0.006
Panel C: Nonlinear model of Dittmar (2002)									
	θ_{mkt}			$\theta_{mkt,2}$			$\theta_{mkt,3}$		
150	0.079	0.036	0.006	0.062	0.025	0.002	0.066	0.026	0.003
350	0.111	0.055	0.010	0.107	0.052	0.010	0.110	0.057	0.012
650	0.091	0.047	0.009	0.118	0.060	0.012	0.085	0.037	0.006

Table 3: Empirical power of Tikhonov test under misspecification with 252 assets

T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Fama-French three factors model									
	θ_{mkt}			θ_{SMB}			θ_{HML}		
150	0.518	0.385	0.168	0.122	0.064	0.015	0.448	0.323	0.133
350	0.822	0.729	0.493	0.163	0.097	0.025	0.763	0.651	0.403
650	0.968	0.936	0.814	0.301	0.195	0.071	0.957	0.915	0.775
Panel B: Linear durable consumption CAPM of Yogo (2006)									
	θ_{mkt}			θ_{ndur}			θ_{dur}		
150	0.309	0.200	0.063	0.066	0.028	0.003	0.062	0.025	0.002
350	0.471	0.355	0.165	0.109	0.058	0.011	0.107	0.053	0.010
650	0.805	0.706	0.470	0.098	0.049	0.010	0.173	0.095	0.023
Panel C: Nonlinear model of Dittmar (2002)									
	θ_{mkt}			$\theta_{mkt,2}$			$\theta_{mkt,3}$		
150	0.441	0.351	0.190	0.091	0.044	0.006	0.086	0.039	0.007
350	0.482	0.406	0.259	0.185	0.105	0.027	0.188	0.108	0.026
650	0.896	0.829	0.636	0.484	0.347	0.147	0.224	0.138	0.042

5.2 Model comparison tests

In this section, we investigate the finite sample behavior of the pairwise and multiple comparison tests. Table 4 presents the results.

Panel A presents the tests developed in Proposition 4. The latter tests the equality of two non-nested SDFs. The simulated data are from FF3 and the nonlinear 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. Then, we estimate each model. The Wald test uses the estimated parameters as well as the variance from Proposition 2 to see whether the non-overlapping factors have null SDF parameter, while the Weighted χ^2 test uses (16). To analyze the power, we set the SDF parameters of the non-overlapping factors to non-null values and repeat the tests. The regularization parameter lies between 0.001 and 0.1. We choose α by running a single model with all the factors and using (13). The results show that the two tests exhibit perfect size control despite the squared and cubic market variable. This would not be the case if one uses the approach of Kan and Robotti (2008) as the test overrejects for the polynomial factors. In addition, the empirical power is high.

Panel B presents the test of equality of the HJ distances of two models when $y_1 \neq y_2$. The test uses the statistic of Proposition 6. To evaluate the size of the test, we simulate two misspecified models with three factors. The two models have r_{SMB} and r_{HML}. For each model, we add the market factor r_{mkt} plus a normally distributed error mean 0 and variance 20% of the market variance. This guarantees that the models have different SDFs and the same HJ distance of 1.026. To evaluate the power, we simulate a misspecified model with the durable consumption factor and a FF3 model. The durable consumption model has a HJ of 1.042. We observe that the test is very conservative. This is not the case when N is small as shown in Gospodinov, Kan, and Robotti (2013). On the other hand, it is able to detect the difference between the durable consumption and the FF3 model. One must keep in mind that when comparing models, it is essential to use the same penalization value. A small value of penalization provides maximum power without compromising size, while a larger value diminishes it. This comes from the fact that as the penalization increases the regularized HJ of the compared models decreases.

Panel C shows the finite sample behavior of the comparison test of multiple models. The test uses the statistic (17). To evaluate the size, we repeat the same process 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 the durable consumption factor (benchmark) and a FF3 for $p = 1$. For $p = 2$, we use the model with durable consumption factor (benchmark), the FF3 and the nonlinear model. The latter has a squared HJ distance of 1.029. We employ the α of the benchmark model to run the tests. The results show that the Wolak test is conservative and exhibits high empirical power. Particularly, the pairwise test ($p = 1$) has a higher empirical power than the Normal pairwise test of Panel B.

Table 4. Model comparison tests

T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
	Panel A: Pairwise tests of equality of two SDFs											
	Wald test						Weighted χ^2 test					
	Size			Power			Size			Power		
150	0.039	0.012	0.001	0.207	0.115	0.026	0.067	0.031	0.004	0.261	0.169	0.055
350	0.082	0.036	0.005	0.567	0.425	0.191	0.093	0.045	0.008	0.504	0.396	0.207
650	0.106	0.055	0.010	0.901	0.830	0.625	0.104	0.052	0.010	0.774	0.702	0.522
	Panel B: Normal pairwise test of equality of two HJ distances											
	Size			Power								
150	0.002	0.000	0.000	0.146	0.042	0.002						
350	0.009	0.001	0.000	0.473	0.251	0.029						
650	0.032	0.008	0.000	0.828	0.675	0.260						
	Panel C: Multiple comparison test (Wolak test)											
	Wolak test (p=1)						Wolak test (p=2)					
	Size			Power			Size			Power		
150	0.012	0.002	0.000	0.358	0.137	0.007	0.000	0.000	0.000	0.206	0.061	0.002
350	0.036	0.007	0.000	0.706	0.433	0.076	0.028	0.004	0.000	0.563	0.296	0.031
650	0.061	0.015	0.000	0.925	0.825	0.434	0.034	0.01	0.000	0.869	0.724	0.292

6 Empirical application

For the empirical application, we consider the Fama and French (1993) model (FF3), the durable consumption CAPM of Yogo (2006), and the nonlinear model of Dittmar (2002) as before, plus the Fama and French (2015) model (FF5). The latter add two new factors to the FF3: the profitability and investment factors. These two factors are built similarly to the value factor in FF3. The profitability factor (Robust Minus Weak) is the difference between the return on the robust operating profitability portfolios minus the return on the weak operating profitability portfolios. The investment factor (Conservative Minus Aggressive) is the difference between the return on the low investment portfolios (conservative) minus the return on the high investment portfolios (aggressive).

For this analysis, we combined 252 portfolios formed on the firm characteristics such as size, book-to-market, market beta, size, operational profitability, investment, earning/price ratio, cash-flow/price ratio, dividend yield, and industries. These portfolios are from Kenneth French's website. Table 8 of the appendix presents the details of these portfolios. We estimate the SDF parameters of the four models and then compare their pricing performances. It is essential to keep the same level of penalization to compare the models. We use a penalization level of 0.001 as the simulations show it is adequate.

Table 5 presents the estimation of the SDF parameters of the four models. For the FF3, we note that the market and the value factors are the only priced variables. Their SDF parameters are non-null with a confidence of 5%. For the consumption model of Yogo, the consumption variables are not priced in the SDF. This model has an aggregate pricing error higher than FF3. For the nonlinear model, no factor exhibits significant SDF parameters, and the level of pricing errors is similar to that of Yogo. Finally, in the FF5 model, the size factor is significant. In addition, the profitability and investment patterns are priced. However, 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 the four factors without HML. The model exhibits the lowest pricing errors.

Table 5: SDF parameter estimates under a misspecified setting

	FF3			YOGO		
Factors	θ_{mkt}	θ_{SMB}	θ_{HML}	θ_{mkt}	θ_{ndur}	θ_{dur}
SDF	0.034***	0.016	0.051**	0.025***	0.348	0.507
t-ratio	2.807	1.063	2.541	2.226	0.575	0.646
HJ	0.123			0.139		

	Nonlinear model		
Factors	θ_{mkt}	$\theta_{mkt,2}$	$\theta_{mkt,3}$
SDF	-0.033	0.066	0.273
t-ratio	-0.263	0.243	1.063
HJ	0.138		

	FF5				
Factors	θ_{mkt}	θ_{SMB}	θ_{HML}	θ_{RMW}	θ_{CMA}
SDF	0.046***	0.037**	0.009	0.086***	0.077**
t-ratio	3.806	2.118	0.321	2.888	2.138
HJ	0.100				

***,**,* indicate that the null hypothesis of unpriced source of risk is rejected at the 1%, 5%, and 10% levels.

We also examine whether the models exhibit different explanatory power, assessed through the HJ distance. To achieve this, we initially perform pairwise comparison tests utilizing the distribution of the squared HJ distance when N is large. Table 6 presents the results of the tests. The results can be summarized as follows: FF3, YOGO, and the nonlinear model show no statistically significant differences in pricing performance, as indicated by the high p-values for the differences in squared HJ distance. Meanwhile, FF5 outperforms all other models. We also augment the basic cubic model with the return on human capital (r_t^l) as in Dittmar (2002). The latter is a two-month moving average of the growth rate in labor income:

$$r_t^l = \frac{L_{t-1} + L_{t-2}}{L_{t-2} + L_{t-3}} - 1,$$

where L_t is the per capita labor income (difference between total personal income and dividend payments divided by the total population). Specifically, we include cubic polynomial expressions of r^l_t. This model does not outperform the others in pricing. FF5 dominates it, though the evidence is now weaker, with a p-value of 0.07.

Table 6: Tests-of-RHJ-Pairwise HJ distance comparison tests

	YOGO	Nonlinear	Nonlinear with human capital	FF5
FF3	-0.020	-0.015	-0.011	0.023
	(0.280)	(0.278)	(0.537)	(0.069)
YOGO		0.0010	0.0039	0.039
		(0.920)	(0.820)	(0.019)
Nonlinear			0.003	0.038
			(0.780)	(0.040)
Nonlinear with human capital				-0.035
				(0.073)

P-values are in brackets.

Finally, we implement the multiple model comparison of Wolak (1989). The test compares the squared HJ distance of a benchmark model against the squared HJ distance of two or more 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 7 presents the results of these comparisons. Each line represents the benchmark model. For FF3, the p-value of 0.6 suggests that its pricing performance is not significantly different from the alternatives (YOGO and the nonlinear model). For YOGO and the nonlinear model, the low p-values indicates that these models are dominated by one of the alternatives. Finally, the null hypothesis cannot be rejected for FF5. In conclusion, the FF5 dominates FF3, YOGO and the nonlinear models.

Table 7: Multiple model comparison tests

Benchmark	$p-1$	$\hat{\delta}^2_\alpha$	LR	p-value
FF3	2	0.123	0.399	0.601
YOGO	2	0.139	2.744	0.025
Nonlinear	2	0.138	2.105	0.042
Nonlinear with human capital	2	0.135	1.608	0.077
FF5	2	0.100	0.4166	0.5834

7 Conclusion

In this paper, we develop a measure of model misspecification when many assets are involved. Specifically, we use Tikhonov and Ridge regularizations to extend the HJ distance. Our approach consists of finding the distance between the empirical SDF proposed by the researcher and the closest valid SDF that prices the returns with errors. The latter depends on a regularization parameter that we choose using a data-driven technique through the out-of-sample R^2. The regularization permits to stabilize the inverse of the covariance matrix. Consequently, the SDF parameter can always be estimated as the minimum of the regularized Hansen-Jagannathan distance even if N is greater than T.

We also propose several comparison tests that used the regularized distance. These tests compare the explanatory power of asset pricing models. As the paper focused on linear asset pricing models, we have analytical formulas that can be simply implemented. We run extensive Monte Carlo simulations to gauge the finite sample behavior of the various tests. They show that our regularization method corrects the oversized nature of the classical tests proposed in the literature when the number of assets is large.

There is room for improvement. There is a need to develop tests adapted to factors that tend to have a low correlation with the returns. In addition, the methods proposed here are only applicable to linear asset pricing models. So, inference on nonlinear models represents an interesting extension. HJ distance is based on the second moment, it would be interesting to consider other discrepancy measures based on higher moments as in Almeida and Garcia (2012).

8 Appendix A: Short review on regularization

A regularization method replaces the explosive eigenvalues of Σ^{-1}, $\dfrac{1}{\lambda_j}, j = 1, \cdots, N$ by $\dfrac{q(\alpha, \mu_j)}{\lambda_j}$, where $q : (0, +\infty) \times (0, \underset{j}{max} \mu_j) \to \mathbb{R}_+$ is a bounded damping function such that

1. $\mid \dfrac{q(\alpha, \mu)}{\mu} \mid < c(\alpha)$ for all μ,

2. $\underset{\alpha \to 0}{lim} : q(\alpha, \mu) \to 1$ for any given μ.

α is the regularization parameter and the expression of $q(\alpha, \mu_j)$ depends on the regularization scheme considered. Taking into account the damping function, the general expression of the regularized weighting matrix noted (Σ_α^{-1}) is given by

$$\Sigma_\alpha^{-1} Y = \sum_{j=1}^{N} \frac{q(\alpha, \mu_j)}{\lambda_j} < Y, \phi_j >_N \phi_j$$

where Y is a conformable vector.

We consider two types of function $q(\alpha, \mu_j)$.

1. **Ridge regularization**

   In this regularization, $\mu_j = \lambda_j$ and the damping function is given by the following expression:

   $$q(\alpha, \lambda_j) = \frac{\lambda_j}{\lambda_j + \alpha}.$$

   This is the same as replacing the matrix Σ^{-1} by $\Sigma_\alpha^{-1} = (\Sigma + \alpha I_N)^{-1}$.

2. **Tikhonov regularization**

   It consists of replacing μ_j by λ_j^2. In addition, the damping function is

   $$q(\alpha, \lambda_j^2) = \frac{\lambda_j^2}{\lambda_j^2 + \alpha}.$$

   The method consists in replacing Σ^{-1} by $\Sigma_\alpha^{-1} = (\Sigma^2 + \alpha I_N)^{-1}\Sigma$.

9 Appendix B: Proofs

9.1 Proof of lemma 1

$tr(\Sigma) = tr\left(\frac{E(r_t r_t^{'})}{N}\right) - tr(\frac{\mu_2\mu_2^{'}}{N})$. By Equation (3), we have
$E(r_t r_t^{'}) = ee^{'} + e\gamma^{'}\beta^{'} + \beta\gamma e^{'} + E[\beta\tilde{f}_t\tilde{f}_t^{'}\beta^{'}] + \beta\gamma\gamma^{'}\beta^{'} + E(\epsilon_t\epsilon_t^{'})$. Therefore,

$$tr(E(r_t r_t^{'})) = tr(ee^{'}) + tr(e\gamma^{'}\beta^{'}) + tr(\beta\gamma e^{'}) + tr(E[\beta\tilde{f}_t\tilde{f}_t^{'}\beta^{'}]) + tr(\beta\gamma\gamma^{'}\beta^{'}) + tr(\Sigma_\epsilon).$$

By Assumption 1, $tr(\frac{ee^{'}}{N}) = \| e \|_N^2 = O(1)$.

From Cauchy-Schwarz inequality, we have $\mid tr(\frac{e\gamma^{'}\beta^{'}}{N}) \mid \leq \sqrt{tr(\frac{ee^{'}}{N}).tr(\frac{(\gamma^{'}\beta^{'})\beta\gamma}{N})}$. Also $tr(\beta\gamma\gamma^{'}\beta^{'}) = tr\left(\gamma\gamma^{'}\beta^{'}\beta\right)$ and $tr\left(\frac{\gamma\gamma^{'}\beta^{'}\beta}{N}\right) \leq \sqrt{tr(\frac{(\beta'\beta)^2}{N^2}).tr((\gamma'\gamma)^2)} \leq tr(\frac{\beta'\beta}{N})tr(\gamma'\gamma)$.The last inequality comes from the fact $\mid tr(AB) \mid \leq trAtrB$ when A and B are positive semi-definite matrices (see Bernstein (2009)). As a result, $tr(\frac{\beta\gamma\gamma^{'}\beta^{'}}{N}) = O(1)$.

Moreover, $tr(\frac{E[\beta\tilde{f}_t\tilde{f}_t^{'}\beta^{'}]}{N}) = tr(\frac{\beta^{'}\beta}{N}E(\tilde{f}_t\tilde{f}_t^{'})) = O(1)$.

We can conclude that

$$tr(E[\frac{r_t r_t^{'}}{N}]) = E[\| r_t \|_N] = O(1).$$

For the mean of the returns , $\mu_2 = e + \beta\gamma$. Therefore,

$$tr(\mu_2 \mu_2^{'}) = tr(ee^{'}) + tr(\beta\gamma e^{'}) + tr(e\gamma^{'}\beta^{'}) + tr((\beta\gamma)(\beta\gamma)^{'}).$$

Using the same arguments as before, we have $tr(\frac{\mu_2\mu_2^{'}}{N}) = O(1)$.

Therefore $tr(\Sigma) = O(1)$. Hence the result.

9.2 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 (2019).

Define $X = \begin{bmatrix} \frac{2r}{N} \\ 2 \end{bmatrix}$.

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 the characteristic function of the set $\{2\}$, i.e

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

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

Problem (7) can be rewritten as

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

It is straightforward to see that g is a convex function. Moreover, f_y is convex as $x \mapsto x^2$ is convex and A is bounded. From Theorem 4.2 of Borwein and Lewis (1992), strong duality holds if $(ri\, dom(g)) \bigcap (ri\, A(dom(f_y))) \neq \emptyset$.[2]

The previous condition is met when Assumption 4 is satisfied. As $\exists m_0 \in L^2$, $E(m_0 - y)^2 < \infty$, $m_0 \in dom(f_y)$, and $A(m_0) \in ri\, A(dom(f_y))$. In addition, because $E[2m_0] = 2$ and $\| E[m_0 r] \|_N^2 <$

[2]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, $ri\, S = \left\{ x \in S : B_\epsilon(x) \bigcap aff\, S \subseteq S \right\}$, 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 $B_\epsilon(x) = \{ y \in \mathbb{R}^N : \| y - x \| < \epsilon \}$.

∞, $g(A(m_0)) = \frac{N}{4\alpha} \parallel E[m\frac{2r}{N}] \parallel^2 = \frac{1}{\alpha} \parallel E[m_0 r] \parallel_N^2 < \infty$ and $A(m_0) \in ri\, dom(g)$. Finally, $(ri\, dom(g)) \bigcap (ri\, A(dom(f_y))) \neq \emptyset$.

The previous result implies that

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

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.

$f_y^*(z) = E\left\{ \sup_{w \in L^2} : zw - (w-y)^2 \right\} = E\left\{ zy + \frac{1}{4}z^2 \right\}$[3], $A^* : R^{N+1} \rightarrow L^2$ and $A^*(\theta) = X'\theta$. Finally $g^*(\nu) = h^*(\nu_1) + \chi_{\{2\}}^*(\nu_2)$ as χ and h are two independent functions. Their conjugates are given by $\chi_{\{2\}}^*(\nu_2) = \sup_{x \in \{2\}} x\nu_2 = 2\nu_2$ and $h^*(\nu_1) = \frac{\alpha}{N} \parallel \nu_1 \parallel^2$.[4]

So, $g^*(\nu) = 2\nu_2 + \frac{\alpha}{N} \parallel \nu_1 \parallel^2$.

Therefore

$$\begin{aligned}
\delta_R^2 &= - \min_{\nu_1 \in \mathbb{R}^N, \nu_2 \in \mathbb{R}} E\left\{ -2y\nu_1' \frac{r}{N} - 2y\nu_2 - \frac{\nu_1 r r' \nu_1}{N^2} - \nu_2^2 - 2\frac{\nu_1' r \nu_2}{N} + 2\nu_2 + \frac{\alpha}{N} \parallel \nu_1 \parallel^2 \right\} \\
&= \max_{\nu_1 \in \mathbb{R}^N, \nu_2 \in \mathbb{R}} E\left\{ 2y\nu_1' \frac{r}{N} + 2y\nu_2 - \frac{\nu_1 r r' \nu_1}{N^2} - \nu_2^2 - 2\frac{\nu_1' r \nu_2}{N} - 2\nu_2 - \frac{\tau}{N} \parallel \nu_1 \parallel^2 \right\}
\end{aligned}$$

Now, we use the fact that $E[y] = 1$. As a result,

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

which is the penalized version of (9). The resulting ν_1 is given by

$$\nu_1 = (\Sigma + \alpha I)^{-1} e = \Sigma_\alpha^{-1} e$$

and $\delta_R^2 = e'\Sigma_\alpha^{-1} e$.

We can do the same for Tikhonov by setting $h(x) = \frac{N}{4\alpha} \parallel x \parallel_\Sigma^2$. The latter can be rewritten as $h(x) = \frac{N}{2\alpha} n(x)$, where $n(x) = \frac{1}{2} \parallel x \parallel_\Sigma^2$. Therefore the convex conjugate of h is $h^*(z) = \frac{N}{2\alpha} n^*(\frac{z}{\frac{N}{2\alpha}})$.

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

[3]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.

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

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

9.3 Proof of Proposition 2

The proof of Proposition 2 uses the following lemmas.

Lemma 2. *Suppose Assumptions 2 and 5 are satisfied. Then,* $\parallel \epsilon'\bar{F} \parallel^2_F = O_p(NT)$, *where* ϵ is a $T \times N$ *matrix with (t,i) element* ϵ_{ti}.

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

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

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, $| E[\bar{f}_t'\bar{f}_t\epsilon_t'\epsilon_t] | \leq \sqrt{E[\parallel \bar{f}_t \parallel^4]E[\parallel \epsilon_t \parallel^4]} = O(N)$. 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)) [5] (with $q = r = 2+\rho$),

[5] 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: $cov(X,Y) \leq 12\alpha(\sigma(X),\sigma(Y))^{\frac{1}{p}}E[| X |^q]^{\frac{1}{q}}E[| Y |^r]^{\frac{1}{r}}$, where $\sigma(X)$ is the sigma algebra generated by X and α is the strong mixing coefficient.

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

As a result,

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

From Assumption 22, and Cauchy-Schwarz, $\mid E[(\bar{f}_{kt}\epsilon_{it})^{2+\rho}] \mid\leq E[\bar{f}_{kt}^{4+2\rho}]^{\frac{1}{2}}E[\epsilon_{it}^{4+2\rho}]^{\frac{1}{2}} \leq c^{\frac{1}{2}}E[\bar{f}_{kt}^{4+2\rho}]^{\frac{1}{2}}$. 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, $E[\| \frac{\epsilon'\bar{F}}{T} \|_F^2] = O(\frac{N}{T})$. In conclusion,

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

Lemma 3. *Suppose Assumptions 2 and 5 are satisfied. Then,* $\| \hat{\beta}-\beta \|_F^2 = O_p(\frac{N}{T})$.

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

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

Therefore,

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

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

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

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

As a result,

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

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

Lemma 5. *Under Assumption 3, we have the following result:*

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

Proof of Lemma 5. We follow the proof of Lemma 3 of Carrasco (2012). We have the following decomposition.

$$\begin{aligned} \| \quad \hat{\Sigma}_\alpha^{-1}\hat{\beta}_k - \Sigma^{-1}\beta_k \|_N^2 \quad &\leq \; 3 \| \hat{\Sigma}_\alpha^{-1} \left(\hat{\beta}_k - \beta_k\right) \|_N^2 && (18) \\ + \quad 3 \| \left(\hat{\Sigma}_\alpha^{-1} - \Sigma_\alpha^{-1}\right) \beta_k \|_N^2 \\ + \quad 3 \| \left(\Sigma_\alpha^{-1} - \Sigma^{-1}\right) \beta_k \|_N^2 &&& (19) \end{aligned}$$

So,

$$\begin{aligned} \| \Sigma_\alpha^{-1}\phi \|_N^2 &= \sum_{j=1}^{\infty} \frac{q(\alpha, \lambda_j^2)^2}{\lambda_j^2} (\phi_j, \phi)_N^2 \\ &\leq \sup_j \frac{q(\alpha, \lambda_j^2)^2}{\lambda_j^2} \| \phi \|_N^2 \\ &\leq \sup_j \frac{q(\alpha, \lambda_j^2)}{\lambda_j^2} \leq \frac{1}{\alpha}. \end{aligned}$$

Therefore, (18) is $O_P(\frac{1}{\alpha T})$.

Let $\phi = \Sigma_\alpha^{-1}\beta_k$. For (19), we have

$$\begin{aligned} \| \left(\hat{\Sigma}_\alpha^{-1} - \Sigma_\alpha^{-1}\right)\beta_k \|_N^2 &= \| \hat{\Sigma}_\alpha^{-1}\left(\hat{\Sigma}_\alpha - \Sigma_\alpha\right)\phi \|_N^2 \\ &\leq \| \hat{\Sigma}_\alpha^{-1} \|_{op}^2 \| \left(\hat{\Sigma}_\alpha - \Sigma_\alpha\right)\phi \|_N^2, \end{aligned}$$

where $\| \Sigma_\alpha^{-1} \|_{op} = \underset{\|\phi\|_N \leq 1}{sup} \| \Sigma_\alpha^{-1}\phi \|_N$ and

$$\Sigma_\alpha Y = \sum_{j/q\neq 0} \frac{\lambda_j}{q(\alpha, \lambda_j^2)} \left\langle \phi_j, Y \right\rangle_N \phi_j$$

is the generalized inverse of Σ_α^{-1}.

We rewrite $\left(\hat{\Sigma}_\alpha - \Sigma_\alpha\right)\phi$ as follows

$$\begin{aligned} (\hat{\Sigma}_\alpha - \Sigma_\alpha)\phi &= (\hat{\Sigma} - \Sigma)\phi + (\hat{\Sigma}_\alpha - \hat{\Sigma})\phi + (\Sigma - \Sigma_\alpha)\phi \\ &= (\hat{\Sigma} - \Sigma)\phi + \sum_{j/q\neq 0} \hat{\lambda}_j \left(\frac{1 - q(\alpha, \hat{\lambda}_j^2)}{q(\alpha, \hat{\lambda}_j^2)}\right) \left\langle \hat{\phi}_j, \phi \right\rangle_N \hat{\phi}_j \\ &\quad + \sum_{j/q\neq 0} \lambda_j \left(\frac{q(\alpha, \lambda_j^2) - 1}{q(\alpha, \lambda_j^2)}\right) \left\langle \phi_j, \phi \right\rangle_N \phi_j. \end{aligned}$$

Therefore,

$$\begin{aligned} \| (\hat{\Sigma}_\alpha - \Sigma_\alpha)\phi \|_N^2 \leq 3 &\ \| (\hat{\Sigma} - \Sigma)\phi \|_N^2 \\ &+3 \sum_{j/q\neq 0} \hat{\lambda}_j^2 \left(\frac{1 - q(\alpha, \hat{\lambda}_j^2)}{q(\alpha, \hat{\lambda}_j^2)}\right)^2 \left\langle \hat{\phi}_j, \phi \right\rangle_N^2 \\ &+3 \sum_{j/q\neq 0} \lambda_j^2 \left(\frac{q(\alpha, \lambda_j^2) - 1}{q(\alpha, \lambda_j^2)}\right)^2 \left\langle \phi_j, \phi \right\rangle_N^2. \end{aligned}$$

We have

$$\begin{aligned} \sum_{j/q\neq 0} \lambda_j^2 \left(\frac{q(\alpha, \lambda_j^2) - 1}{q(\alpha, \lambda_j^2)}\right)^2 \left\langle \phi_j, \phi \right\rangle_N^2 &= \alpha^2 \sum_{j/q\neq 0} \frac{1}{\lambda_j^2} \left\langle \phi_j, \phi \right\rangle_N^2. \\ &= O(\alpha^2) \end{aligned}$$

as $\beta_k \in \Phi_3$. As a result, $\| (\hat{\Sigma}_\alpha - \Sigma_\alpha)\phi \|_N^2 = O_P(\frac{1}{T}) + O_P(\alpha^2)$ and $\| \left(\hat{\Sigma}_\alpha^{-1} - \Sigma_\alpha^{-1}\right)\beta_k \|_N^2 = O_p(\frac{1}{\alpha T}) + O(\alpha)$

Finally, the term (19) satisfies

$$\begin{aligned}\| \left(\Sigma_\alpha^{-1} - \Sigma^{-1}\right) \beta_k \|_N^2 &= \sum_j \left(\frac{q(\alpha, \lambda_j^2) - 1}{\lambda_j}\right)^2 \left\langle \phi_j, \beta_k \right\rangle_N^2 \\ &= \alpha^2 \sum_j \frac{1}{\lambda_j^2(\lambda_j^2 + \alpha)^2} \left\langle \phi_j, \beta_k \right\rangle_N^2 \\ &\leq \alpha^2 \sum_j \frac{1}{\lambda_j^6} \left\langle \phi_j, \beta_k \right\rangle_N^2 = O(\alpha^2)\end{aligned}$$

as $\beta_k \in \Phi_3$.

Lemma 6. *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 $u \in \mathbb{R}^N$ *isnot random and* $\| u \|_N = O(1)$. *If Assumptions 2(i) and 5(iii) hold, then*

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

when T, N *go simultaneously to* ∞.

Proof of Lemma 6.

First, consider the case when $\{r_t\}_{t=1,\cdots,T}$ are independent. From Assumptions 5 5, $E\left[\| r_t \|_N^{2+\rho}\right] = O(1)$ 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 $\sigma_{T,N}^2 = T.var(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Y_{t,T,N}) = T.\frac{u' \Sigma u}{N} \equiv T\sigma_N^2 > 0$.

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

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

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

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

as $\sigma_N=O(1)$, $E\left[|<\tilde{r}_t,u>_N|^{2+\rho}\right]\leq E\left[\parallel\tilde{r}_t\parallel_N^{2+\rho}\right]\parallel u\parallel_N^{2+\rho}=O(1)$by Assumption 5(iii).

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

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

As a result, $0<\underset{N,T\to\infty}{lim}var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_{t,T,N})<\infty$. 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) for a similar application of this result. Hence, $X_{T,N}$ asymptotically converges to a normal distribution when T,N go to ∞.

Lemma 7. *Suppose Assumption 5 is satisfied. For any $u,v\in\mathbb{R}^N$ with $\parallel u\parallel_\infty<\infty$ and $\parallel v\parallel_\infty<\infty$, as N and T go to ∞, if*

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

then $\sqrt{T}<\left(\hat{\Sigma}-\Sigma\right)v,u>_N$ converges to a gaussian distribution of mean 0 and variance $\sigma_{u,v}^2$.

Proof of Lemma 7. 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,

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

Following Lemma 6, $<v,\left(\mu_2-\hat{\mu}_2\right)>_N\frac{1}{\sqrt{T}}\sum_{t=1}^{T}<\tilde{r}_t,u>_N=O_p(\frac{1}{\sqrt{T}})$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}<v,\tilde{r}_t>_N<\left(\mu_2-\hat{\mu}_2\right),u>_N=O_p(\frac{1}{\sqrt{T}})$, and $\sqrt{T}<v,\left(\mu_2-\hat{\mu}_2\right)>_N<\left(\mu_2-\hat{\mu}_2\right),u>_N=O_p(\frac{1}{\sqrt{T}})$. As a result,

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

From here on, the proof is similar to that of Lemma 6. To apply the central limit theorem of Francq and Zakoïan (2005), we need

$$
\underset{N\rightarrow\infty}{lim}:\underset{t}{sup}E[(<v,\tilde{r}_t>_N<\tilde{r}_t,u>_N)^{2+\rho}]<\infty
$$

for some $\rho>0$. This condition is met because of Assumption 5 5.

Proof of Proposition 2

Consistency:

Recall that by Equation (3), we have

$$\hat{\mu}_2 = \frac{1}{T}\sum_t r_t = e + \beta(\gamma + \hat{\mu}_1 - \mu_1) + \bar{\epsilon},$$
$$\mu_2 = e + \beta\gamma.$$

This yields the following decomposition of the $\hat{\theta}_{HJ}$:

$$\begin{aligned}\hat{\theta}^{\alpha}_{HJ} - \theta_{HJ} = {} & (\hat{V}_{11}^{-1} - V_{11}^{-1})\gamma \\ & + \hat{V}_{11}^{-1}(\hat{\mu}_1 - \mu_1) \\ & + \hat{V}_{11}^{-1}(\frac{\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}\hat{\beta}}{N})^{-1}\left[\frac{\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}(\beta - \hat{\beta})}{N}(\gamma + \hat{\mu}_1 - \mu_1)\right. \\ & \left. + \frac{\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}e}{N} + \frac{\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}\bar{\epsilon}}{N}\right].\end{aligned} \tag{20}$$

For the first two rows, $(\hat{V}_{11}^{-1} - V_{11}^{-1})\gamma$ 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).

$$\frac{\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}\hat{\beta}}{N} = \left[\frac{\hat{\beta}'_{k_1}\hat{\Sigma}_{\alpha}^{-1}\hat{\beta}_{k_2}}{N}\right]_{k_1,k_2=1,\cdots,K}$$
$$= < \hat{\beta}_{k_1}, \hat{\Sigma}_{\alpha}^{-1}\hat{\beta}_{k_2} >_{N;k_1,k_2=1,\cdots,K} .$$

We have

$$< \hat{\beta}_{k_1}, \hat{\Sigma}_{\alpha}^{-1}\hat{\beta}_{k_2} >_N = < \hat{\beta}_{k_1} - \beta_{k_1}, \hat{\Sigma}_{\alpha}^{-1}\hat{\beta}_{k_2} - \Sigma^{-1}\beta_{k_2} >_N + \tag{21}$$
$$< \hat{\beta}_{k_1} - \beta_{k_1}, \Sigma^{-1}\beta_{k_2} >_N + \tag{22}$$
$$< \beta_{k_1}, \hat{\Sigma}_{\alpha}^{-1}\hat{\beta}_{k_2} - \Sigma^{-1}\beta_{k_2} >_N + \tag{23}$$
$$< \beta_{k_1}, \Sigma^{-1}\beta_{k_2} >_N - C \tag{24}$$
$$+ C. \tag{25}$$

where C was defined in Assumption 3. $|(21)| \leq \| \hat{\beta}_{k_1} - \beta_{k_1} \|_N \| \hat{\Sigma}_{\alpha}^{-1}\hat{\beta}_{k_1} - \Sigma^{-1}\beta_{k_1} \|_N \to 0$ as $N, T \to \infty$, and $\alpha T \to \infty$ using Lemma 5.

For (22), we have

$$|< \hat{\beta}_{k_1} - \beta_{k_1}, \Sigma^{-1}\beta_{k_2} >_N| \leq \| \hat{\beta}_{k_1} - \beta_{k_1} \|_N \| \Sigma^{-\frac{1}{2}}\beta_{k_2} \|_N \to 0$$

as $N, T \to \infty$, using Lemma 4.

The same is true for (23).

Finally, using assumption 3, (24) goes to 0 as N goes to ∞.

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

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

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

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

Distribution:

We detail the proof of the asymptotic normality proof for Tikhonov estimator. The result for ridge could be shown similarly. We analyze the decomposition (20) using the following results:

- $\hat{\beta} - \beta = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \epsilon_t \bar{f}_t' \hat{V}_{11}^{-1}$
- Note that $\hat{C}_\beta = \dfrac{\hat{\beta}'\hat{\Sigma}_\alpha^{-1}\hat{\beta}}{N}$ and we already have shown that $\hat{C}_\beta - C_\beta \xrightarrow{P} 0_{K,K}$.
- $\hat{\beta}'\hat{\Sigma}_\alpha^{-1}e = (\hat{\beta} - \beta)'\hat{\Sigma}_\alpha^{-1}e + \beta'(\hat{\Sigma}_\alpha^{-1} - \Sigma^{-1})e + \beta'\Sigma^{-1}e$. The last term is $0_{K,1}$ as the population first order condition of (2).
- We have

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

Therefore,

$$\begin{aligned}
&\sqrt{T}\left(\hat{\theta}^{\alpha}_{HJ}-\theta_{HJ}\right)+\sqrt{T}\hat{V}_{11}^{-1}\hat{C}_{\beta}^{-1}\beta'\hat{\Sigma}_{\alpha}^{-1}(\hat{\Sigma}_{\alpha}-\hat{\Sigma})\Sigma^{-1}\frac{e}{N} \qquad (26)\\
&=\hat{V}_{11}^{-1}\Bigg\{-\sqrt{T}(\hat{V}_{11}-V_{11})V_{11}^{-1}\gamma\\
&+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(f_t-\mu_1)\\
&+\hat{C}_{\beta}^{-1}\Bigg[-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}\epsilon_t\bar{f}_t'}{N}\hat{V}_{11}^{-1}(\gamma+\hat{\mu}_1-\mu_1)\\
&-\sqrt{T}\beta'\hat{\Sigma}_{\alpha}^{-1}(\hat{\Sigma}-\Sigma)\Sigma^{-1}\frac{e}{N}\\
&+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{V}_{11}^{-1}\frac{\bar{f}_t\epsilon_t'\hat{\Sigma}_{\alpha}^{-1}e}{N}\\
&+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}\epsilon_t}{N}\Bigg]\Bigg\}.
\end{aligned}$$

We need prove the asymptotic normality of each component of (26) to get the result of Proposition 2.

- Note that

$$\begin{aligned}
\sqrt{T}(\hat{V}_{11}-V_{11})V_{11}^{-1}\gamma &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\tilde{f}_t\tilde{f}_t'\theta_{HJ}-\gamma\right)+(\mu_1-\hat{\mu}_1)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(r_t-\mu_2)'\theta_{HJ}\\
&+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(r_t-\mu_2)(\mu_2-\hat{\mu}_2)'\theta_{HJ}\\
&+\sqrt{T}(\mu_2-\hat{\mu}_2)(\mu_2-\hat{\mu}_2)'\theta_{HJ}\\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\tilde{f}_t\tilde{f}_t'\theta_{HJ}-\gamma\right)+o_p(1).
\end{aligned}$$

- Normality of the second row comes from assumption 2.

- For the third row, $\hat{\mu}_1 \xrightarrow{P} \mu_1$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\bar{f}_t'\hat{V}_{11}^{-1}(\gamma+\hat{\mu}_1-\mu_1)$ has a gaussian distribution.

To see this, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}\epsilon_t\bar{f}_t'}{N}\hat{V}_{11}^{-1}(\gamma+\hat{\mu}_1-\mu_1) = \frac{\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}}{N}[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\tilde{f}_t'V_{11}^{-1}\gamma \quad (27)$$

$$+ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\bar{f}_t'(\hat{V}_{11}^{-1}-V_{11}^{-1})\gamma \quad (28)$$

$$- \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t(\hat{\mu}_1-\mu_1)'V_{11}^{-1}\gamma \quad (29)$$

$$+ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\bar{f}_t'\hat{V}_{11}^{-1}(\hat{\mu}_1-\mu_1)] \quad (30)$$

The term (28) can be rewritten as $T^{-\frac{1}{2}}\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}\frac{\epsilon'\bar{F}}{N}(\hat{V}_{11}^{-1}-V_{11}^{-1})\gamma$. From Lemma 2, $\| \epsilon'\bar{F} \|_F^2 = O_p(NT)$. So $\| \epsilon'\bar{F} \|^2 = O_p(NT)$. From Lemma 5, $\| \hat{\Sigma}_{\alpha}^{-1}\hat{\beta}_k - \Sigma^{-1}\beta_k \|_N \to 0$, as : $N, T \to \infty$, and : $\alpha T \to \infty$. In addition, $\| (\hat{V}_{11}^{-1}-V_{11}^{-1})\gamma \|^2 = O_p(\frac{1}{T})$. Therefore, (28) is $o_p(1)$ when $N, T \to \infty$ and $\alpha T \to \infty$.

Using the fact that $\theta_{HJ} = V_{11}^{-1}\gamma$, we can rewrite (29) as

$$(\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}-\beta'\Sigma^{-1})\frac{1}{N\sqrt{T}}\epsilon'\Theta(\hat{\mu}_1-\mu_1)+\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}<\Sigma^{-1}\beta,\epsilon_t>_N\right)(\hat{\mu}_1-\mu_1)'V_{11}^{-1}\gamma,$$

with $\Theta = \begin{bmatrix}\theta_{HJ}' \\ \vdots \\ \theta_{HJ}'\end{bmatrix}$ is a $T \times K$ matrix. We have $\| \epsilon'\Theta \|^2 = O_P(N)$, $\| \hat{\mu}_1 - \mu_1 \|^2 = O_P(\frac{1}{T})$. As a result, $(\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}-\beta'\Sigma^{-1})\frac{1}{N\sqrt{T}}\epsilon'\Theta(\hat{\mu}_1-\mu_1)$ is $o_p(1)$ when $N, T \to \infty$, and : $\alpha T \to \infty$.

Using Lemma 6, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} < \Sigma^{-1}\beta, \epsilon_t >_N$ is $O_p(1)$ as $\| \Sigma^{-1}\beta \|_N < \infty$ 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, (29) is $o_p(1)$.

For (30), we can rewrite it as

$$\frac{T^{-\frac{1}{2}}}{N}(\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}-\beta'\Sigma^{-1})(\beta-\hat{\beta})(\hat{\mu}_1-\mu_1).$$

From Lemma 3 and 5, the expression is $o_p(1)$ as $N, T \to \infty$, and $\alpha T \to \infty$.

Finally, (27) is equal to

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

The first part is $o_P(1)$ when $N,T\to\infty$, and $\alpha T\to\infty$. For the second part, $\forall m\in\mathbb{R}^K$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}<\Sigma^{-1}\beta m,\epsilon_t\tilde{f}_t'V_{11}^{-1}\gamma>_N$$

converges to a normal distribution by virtue of Lemma 6. Therefore, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}<\Sigma^{-1}\beta,\epsilon_t\tilde{f}_t'V_{11}^{-1}\gamma>_N$ has a gaussian distribution when $N,T\to\infty$ by the Cramer Wold device.

- $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{V}_{11}^{-1}\frac{\bar{f}_t\epsilon_t'\hat{\Sigma}_\alpha^{-1}e}{N}$ has a gaussian distribution by using the following decomposition

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{V}_{11}^{-1}\frac{\bar{f}_t\epsilon_t'\hat{\Sigma}_\alpha^{-1}e}{N} &= V_{11}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\tilde{f}_t\epsilon_t'\hat{\Sigma}_\alpha^{-1}e}{N} && (31)\\ &+\frac{1}{N\sqrt{T}}\left(\hat{V}_{11}^{-1}-V_{11}^{-1}\right)\bar{F}'\epsilon\hat{\Sigma}_\alpha^{-1}e && (32)\\ &+\frac{1}{N\sqrt{T}}V_{11}^{-1}(\mu_1-\hat{\mu}_1)\sum_{t=1}^{T}\epsilon_t'\hat{\Sigma}_\alpha^{-1}e. && (33)\end{aligned}$$

By Lemma 6 and the Cramer Wold device, (31) converges to a gaussian distribution. Using Lemmas 2 and 5, (32) and (33) are $o_p(1)$ when $N,T\to\infty$, and $\alpha T\to\infty$.

- As as N, $T\to\infty$, $\alpha\to 0$ and $\alpha T\to\infty$, the normality of $\sqrt{T}\beta'\hat{\Sigma}_\alpha^{-1}(\hat{\Sigma}-\Sigma)\Sigma^{-1}\frac{e}{N}$ comes from Lemma 7. Indeed,

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

  Using the proof of Lemma 7 , we can rewrite $<\Sigma^{-1}\beta,\sqrt{T}(\hat{\Sigma}-\Sigma)\Sigma^{-1}e>_N$ as

$$\begin{aligned}<\Sigma^{-1}\beta,\sqrt{T}\left[\hat{\Sigma}-\Sigma\right]\Sigma^{-1}e\ \ >_N &=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[<\tilde{r}_t,\Sigma^{-1}e>_N<\tilde{r}_t,\Sigma^{-1}\beta>_N\\ &-E[<\tilde{r}_t,\Sigma^{-1}e>_N<\tilde{r}_t,\Sigma^{-1}\beta>]]+o_p(1).\end{aligned}$$

This term is asymptotically gaussian.

- For the term $\sqrt{T}\beta'\hat{\Sigma}_\alpha^{-1}(\hat{\Sigma}_\alpha - \hat{\Sigma})\Sigma^{-1}\frac{e}{N}$, notice that

$$\beta_k'\hat{\Sigma}_\alpha^{-1}(\hat{\Sigma}_\alpha - \hat{\Sigma})\Sigma^{-1}\frac{e}{N} = \sum_{j/q\neq 0}\hat{\lambda}_j\left(\frac{1-q(\alpha,\hat{\lambda}_j^2)}{q(\alpha,\hat{\lambda}_j^2)}\right) < \hat{\phi}_j, \Sigma^{-1}e >_N < \hat{\phi}_j, \hat{\Sigma}_\alpha^{-1}\beta_k >_N .$$

So $\mid \beta_k'\hat{\Sigma}_\alpha^{-1}(\hat{\Sigma}_\alpha - \hat{\Sigma})\Sigma^{-1}\frac{e}{N} \mid = \alpha \mid \sum_{j/q\neq 0}\frac{1}{\hat{\lambda}_j} < \hat{\phi}_j, \Sigma^{-1}e >_N < \hat{\phi}_j, \hat{\Sigma}_\alpha^{-1}\beta_k >_N \mid$. We have

$$\begin{aligned} \mid \sum_{j/q\neq 0}\frac{1}{\hat{\lambda}_j} < \hat{\phi}_j, \Sigma^{-1}e >_N < \hat{\phi}_j, \hat{\Sigma}_\alpha^{-1}\beta_k >_N \mid &\leq \left(\sum_{j/q\neq 0}\frac{1}{\hat{\lambda}_j} < \hat{\phi}_j, \Sigma^{-1}e >_N^2\right)^{\frac{1}{2}}\left(\sum_{j/q\neq 0}\frac{1}{\hat{\lambda}_j} < \hat{\phi}_j, \Sigma^{-1}e >_N^2\right. \\ &< \infty \end{aligned}$$

as $\beta_k, e \in \Phi_3$. So, $\| \sqrt{T}\beta'\hat{\Sigma}_\alpha^{-1}(\hat{\Sigma}_\alpha - \hat{\Sigma})\Sigma^{-1}\frac{e}{N} \|^2 = O_p(\alpha^2 T)$.

In conclusion, the term $\sqrt{T}\beta'\hat{\Sigma}_\alpha^{-1}(\hat{\Sigma}_\alpha - \hat{\Sigma})\Sigma^{-1}\frac{e}{N}$ is $o_p(1)$ as $N, T, \alpha T \to \infty$ and $\alpha^2 T \to 0$. Using the previous results, we have

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

where

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

and

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

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

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

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

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

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

9.4 Proof of Proposition 3

The proof follows closely that of Proposition 3 of Kan and Robotti (2009).

$$\begin{aligned}\delta_1^2 - \delta_2^2 \;=\;& \frac{\mu_2'\Sigma^{-1}\mu_2}{N} - \frac{\mu_2'\Sigma^{-1}V_{21,1}}{N}\left(\frac{V_{12,1}\Sigma^{-1}V_{21,1}}{N}\right)^{-1}\frac{V_{12,1}\Sigma^{-1}\mu_2}{N} - \\ & -\frac{\mu_2'\Sigma^{-1}\mu_2}{N} + \frac{\mu_2'\Sigma^{-1}V_{21,2}}{N}\left(\frac{V_{12,2}\Sigma^{-1}V_{21,2}}{N}\right)^{-1}\frac{V_{12,2}\Sigma^{-1}\mu_2}{N}.\end{aligned}$$

The population SDF parameter of models 1 and 2 are respectively

$$\theta_1 = (V_{12,1}\Sigma^{-1}V_{21,1})^{-1}V_{12}\Sigma^{-1}\mu_2$$

and

$$\theta_2 = (V_{12,2}\Sigma^{-1}V_{21,2})^{-1}V_{12,2}\Sigma^{-1}\mu_2.$$

Therefore

$$\frac{(V_{12,2}\Sigma^{-1}V_{21,2})}{N}\theta_2 = \frac{V_{12,2}\Sigma^{-1}\mu_2}{N}.$$

Noting that $V_{21,2} = \begin{bmatrix} V_{21,1} & V_{21,r}\end{bmatrix}$ where $V_{21,r}$ is the remaining of the matrix $V_{21,2}$ and

$$V_{21,1} = \begin{bmatrix} V_{21,1} & V_{21,r}\end{bmatrix}\begin{bmatrix} I_{K_1} \\ 0_{K_3,K_1}\end{bmatrix} = V_{21,2}\begin{bmatrix} I_{K_1} \\ 0_{K_3,K_1}\end{bmatrix},$$

we have

$$
\begin{aligned}
\delta_1^2-\delta_2^2 &= \theta_2'\left(\frac{V_{12,2}\Sigma^{-1}V_{21,2}}{N}\right)\theta_2 \\
&\quad -\frac{\mu_2'\Sigma^{-1}}{N}V_{21,2}\begin{bmatrix}\left(\frac{V_{12,1}\Sigma^{-1}V_{21,1}}{N}\right)^{-1} & 0_{K_1,K_3}\\ 0_{K_3,K_1} & 0_{K_3,K_3}\end{bmatrix}V_{12,2}\frac{\Sigma^{-1}\mu_2}{N} \\
&= \theta_2'\left(\frac{V_{12,2}\Sigma^{-1}V_{21,2}}{N}\right)\theta_2 \\
&\quad -\theta_2'\left(\frac{V_{12,2}\Sigma^{-1}V_{21,2}}{N}\right)\begin{bmatrix}\left(\frac{V_{12,1}\Sigma^{-1}V_{21,1}}{N}\right)^{-1} & 0_{K_1,K_3}\\ 0_{K_3,K_1} & 0_{K_3,K_3}\end{bmatrix}\left(\frac{V_{12,2}\Sigma^{-1}V_{21,2}}{N}\right)\theta_2 \\
&= \theta_2'\left(\frac{V_{12,2}\Sigma^{-1}V_{21,2}}{N}\right)\theta_2 \\
&\quad -\theta_2'\begin{bmatrix}\left(\frac{V_{12,1}\Sigma^{-1}V_{21,1}}{N}\right) & \left(\frac{V_{12,1}\Sigma^{-1}V_{21,r}}{N}\right)\\ \left(\frac{V_{12,1}\Sigma^{-1}V_{21,r}}{N}\right)' & \left(\frac{V_{12,1}\Sigma^{-1}V_{21,r}}{N}\right)'\left(\frac{V_{12,1}\Sigma^{-1}V_{21,1}}{N}\right)^{-1}\left(\frac{V_{12,1}\Sigma^{-1}V_{21,r}}{N}\right)\end{bmatrix}\theta_2 \\
&= \theta_{22}'\left[\frac{V_{12,r}\Sigma^{-1}V_{21,r}}{N}-\left(\frac{V_{12,1}\Sigma^{-1}V_{21,r}}{N}\right)'\left(\frac{V_{12,1}\Sigma^{-1}V_{21,1}}{N}\right)^{-1}\left(\frac{V_{12,1}\Sigma^{-1}V_{21,r}}{N}\right)\right]\theta_{22} \\
&= \theta_{22}'C_{2,22}^{-1}\theta_{22}.
\end{aligned}
$$

If $C_{2,22}^{-1}$ is full rank, $\delta_1^2-\delta_2^2=0$ if and only if $\theta_{22}=0$. This is the first result of Proposition 3.

For Tikhonov, under the hypothesis $\theta_{22}=0$, $z=\sqrt{T}V(\hat{\theta}_{22}^{\alpha})^{-\frac{1}{2}}\hat{\theta}_{22}^{\alpha}\xrightarrow{d}\mathcal{N}(0,I_{K_3})$ as T, N and αT go to infinity and α^2T goes to zero.

$T(\hat{\delta}_{1,\alpha}^2-\hat{\delta}_{2,\alpha}^2)=T\hat{\theta}_{22}^{\alpha'}\hat{C}_{2,22,\alpha}^{-1}\hat{\theta}_{22}^{\alpha}=z'V(\hat{\theta}_{22}^{\alpha})^{\frac{1}{2}}\hat{C}_{2,22,\alpha}^{-1}V(\hat{\theta}_{22}^{\alpha})^{\frac{1}{2}}z$. $\hat{C}_{2,22,\alpha}^{-1}$ converges in probability to $C_{2,22}^{-1}$ as $\hat{C}_{2,22}^{-1}$, a function of $\hat{C}_2$ which converges to C_2 when N,T and $\alpha T\to\infty$ and $\alpha\to 0$ (see the proof of the consistency of $\hat{\theta}_{HJ}^{\alpha}$). Therefore

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

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

9.5 Proof of Proposition 4

The proof of Point 1 follows from Lemma 3 in Kan and Robotti (2009).

Now, consider Point 2. For Tikhonov, under the hypothesis $\begin{bmatrix}\theta_{12}\\ \theta_{22}\end{bmatrix}=0$, $z=\sqrt{T}V\left(\begin{bmatrix}\hat{\theta}_{12}^{\alpha}\\ \hat{\theta}_{22}^{\alpha}\end{bmatrix}\right)^{-\frac{1}{2}}\begin{bmatrix}\hat{\theta}_{12}^{\alpha}\\ \hat{\theta}_{22}^{\alpha}\end{bmatrix}\xrightarrow{d}$ $\mathcal{N}(0,I_{K_3})$ as T, N and $\alpha T\to\infty$, $\alpha^2T\to 0$, and α goes to zero. Moreover,

$$
\begin{aligned}
T(\hat{\delta}^2_{1,\alpha} - \hat{\delta}^2_{2,\alpha}) &= T \begin{bmatrix} \hat{\theta}^{\alpha}_{12} \\ \hat{\theta}^{\alpha}_{22} \end{bmatrix}' \begin{bmatrix} -\hat{C}^{-1}_{1,22,\alpha} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & \hat{C}^{-1}_{2,22,\alpha} \end{bmatrix} \begin{bmatrix} \hat{\theta}^{\alpha}_{12} \\ \hat{\theta}^{\alpha}_{22} \end{bmatrix} \\
&= z' V \left(\begin{bmatrix} \hat{\theta}^{\alpha}_{12} \\ \hat{\theta}^{\alpha}_{22} \end{bmatrix} \right)^{\frac{1}{2}} \begin{bmatrix} -\hat{C}^{-1}_{1,22,\alpha} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & \hat{C}^{-1}_{2,22,\alpha} \end{bmatrix} V \left(\begin{bmatrix} \hat{\theta}^{\alpha}_{12} \\ \hat{\theta}^{\alpha}_{22} \end{bmatrix} \right)^{\frac{1}{2}} z.
\end{aligned}
$$

The results follows from the singular value decomposition of

$$
V \left(\begin{bmatrix} \hat{\theta}^{\alpha}_{12} \\ \hat{\theta}^{\alpha}_{22} \end{bmatrix} \right)^{\frac{1}{2}} \begin{bmatrix} -\hat{C}^{-1}_{1,22,\alpha} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & \hat{C}^{-1}_{2,22,\alpha} \end{bmatrix} V \left(\begin{bmatrix} \hat{\theta}^{\alpha}_{12} \\ \hat{\theta}^{\alpha}_{22} \end{bmatrix} \right)^{\frac{1}{2}}.
$$

9.6 Proof of Proposition 5

For Ridge, $q^P_t(\nu_{1\alpha}) = 2y_t \frac{\nu'_{1\alpha}\tilde{r}_t}{N} - \frac{\nu'_{1\alpha}\tilde{r}_t\tilde{r}'_t\nu_{1\alpha}}{N^2} + 2\frac{\nu'_{1\alpha}\mu_2}{N} - \alpha\frac{\nu'_{1\alpha}\nu_{1\alpha}}{N}$.

For Tikhonov, $q^P_t(\nu_{1\alpha}) = 2y_t \frac{\nu'_{1\alpha}\tilde{r}_t}{N} - \frac{\nu'_{1\alpha}\tilde{r}_t\tilde{r}'_t\nu_{1\alpha}}{N^2} + 2\frac{\nu'_{1\alpha}\mu_2}{N} - \alpha\frac{\nu'^{-1}_{1\alpha}\nu_{1\alpha}}{N}$.

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

$$
\begin{aligned}
\sqrt{T}\left(\hat{\delta}^2_{\alpha} - \delta^2\right) &= \sqrt{T}\left(\hat{\delta}^2_{\alpha} - \delta^2_{\alpha}\right) + \sqrt{T}\left(\delta^2_{\alpha} - \delta^2\right) \\
&= \sqrt{T}\left(\hat{E}\left[q^P_t(\hat{\nu}_{1\alpha})\right] - \hat{E}\left[q^P_t(\nu_{1\alpha})\right]\right) && (34) \\
&+ \sqrt{T}\left(\hat{E}\left[q^P_t(\nu_{1\alpha})\right] - E\left[q^P_t(\nu_{1\alpha})\right]\right) && (35) \\
&+ \sqrt{T}\left(\delta^2_{\alpha} - \delta^2\right) && (36)
\end{aligned}
$$

As $\hat{E}\left[q^P_t(\hat{\nu}_{1\alpha})\right]$ is concave, we have $\sqrt{T}\left(\hat{E}\left[q^P_t(\hat{\nu}_{1\alpha})\right] - \hat{E}\left[q^P_t(\nu_{1\alpha})\right]\right) \leq \sqrt{T}\nabla\hat{E}\left[q^P_t(\nu_{1\alpha})\right](\hat{\nu}_{1\alpha} - \nu_{1\alpha})$. The term $\nabla\hat{E}\left[q^P_t(\nu_{1\alpha})\right]$ is the Fr chet derivative of $\hat{E}\left[q^P_t(\nu_{1\alpha})\right]$ at $\nu_{1\alpha}$ and is an operator from $\mathbb{R}^N$ to $\mathbb{R}$ defined by

$$
\nabla\hat{E}\left[q^P_t(\nu_{1\alpha})\right] h = \frac{1}{T}\sum_{t=1}^{T}\left[2y_t < \tilde{r}_t, h >_N - 2 < \frac{\tilde{r}_t\tilde{r}'_t}{N}\nu_{1\alpha}, h >_N\right]
$$
$$
+ 2 < \mu_2, h >_N - 2\alpha < \nu_{1\alpha}, h >_N .
$$

As $\nabla E\left[q^P_t(\nu_{1\alpha})\right] = 0$, Assumption 6 implies $E\left[2y_t\frac{\tilde{r}_t}{N} - 2\frac{\tilde{r}_t\tilde{r}'_t\nu_{1\alpha}}{N^2} + 2\frac{\mu_2}{N} - 2\frac{\alpha}{N}\nu_{1\alpha}\right] = 0$ and $\frac{\mu_2}{N} - \frac{\alpha}{N}\nu_{1\alpha} = -E\left[y_t\frac{\tilde{r}_t}{N} - \frac{\tilde{r}_t\tilde{r}'_t\nu_{1\alpha}}{N^2}\right]$ for Ridge, while a similar formula holds for Tikhonov

Therefore, for Ridge and Tikhonov,

$$
\begin{aligned}
& \sqrt{T}\nabla \hat{E}\left[q_t^P(\nu_{1\alpha}^*)\right](\hat{\nu}_{1\alpha}-\nu_{1\alpha}) \\
= \; & \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left[y_t\frac{\tilde{r}_t}{N}-E[y_t\frac{\tilde{r}_t}{N}{}^{\prime}]\right]^{\prime}(\hat{\nu}_{1\alpha}-\nu_{1\alpha}) && (37)\\
& -\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left[\nu_{1\alpha}^{\prime}\frac{\tilde{r}_t\tilde{r}_t^{\prime}}{N^2}-E[\nu_{1\alpha}^{\prime}\frac{\tilde{r}_t\tilde{r}_t^{\prime}}{N^2}]\right](\hat{\nu}_{1\alpha}-\nu_{1\alpha}). && (38)
\end{aligned}
$$

We have

$$
\begin{aligned}
\| \hat{\nu}_{1\alpha}-\nu_{1\alpha} \|_N &= \| \hat{\Sigma}_\alpha^{-1}\hat{e}-\Sigma_\alpha^{-1}e \|_N \\
&= \| \hat{\Sigma}_\alpha^{-1}\hat{e}-\Sigma^{-1}e+\Sigma^{-1}e-\Sigma_\alpha^{-1}e \|_N \\
&\leq \| \hat{\Sigma}_\alpha^{-1}\hat{e}-\Sigma^{-1}e \|_N \\
&\quad + \| \Sigma^{-1}e-\Sigma_\alpha^{-1}e \|_N = o_p(1)
\end{aligned}
$$

when N, T, $\alpha T\rightarrow\infty$ and $\alpha\rightarrow 0$. So, (37) is equal to

$$
\begin{aligned}
\left| \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left(y_t\frac{\tilde{r}_t}{N}-E\left[y_t\frac{\tilde{r}_t}{N}\right]\right)^{\prime}(\hat{\nu}_{1\alpha}-\nu_{1\alpha}) \right| &\leq \left\| \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left(y_t\frac{\tilde{r}_t}{N}-E\left[y_t\frac{\tilde{r}_t}{N}\right]\right)^{\prime} \right\|_N \| \hat{\nu}_{1\alpha}-\nu_{1\alpha} \|_N \\
&= o_p(1).
\end{aligned}
$$

while (38) is bounded by

$$
\begin{aligned}
|< \sqrt{T}(\hat{\Sigma}-\Sigma)\nu_{1\alpha}, \hat{\nu}_{1\alpha}-\nu_{1\alpha} >_N| &\leq \| \sqrt{T}(\hat{\Sigma}-\Sigma) \|_N \| \nu_{1\alpha} \|_N \| \hat{\nu}_{1\alpha}-\nu_{1\alpha} \|_N \\
&\leq \| \sqrt{T}(\hat{\Sigma}-\Sigma) \|_N \| \nu_1 \|_N \| \hat{\nu}_{1\alpha}-\nu_{1\alpha} \|_N \\
&= o_p(1),
\end{aligned}
$$

when N, T, $\alpha T\rightarrow\infty$ and $\alpha\rightarrow 0$.

Therefore, as N, T, $\alpha T\rightarrow\infty$ and $\alpha\rightarrow 0$, $\sqrt{T}\nabla \hat{E}\left[q_t^P(\nu_{1\alpha})\right](\hat{\nu}_{1\alpha}-\nu_{1\alpha}) = o_p(1)$ and hence (34)$= o_p\left(1\right)$.

Term (35) can be rewritten as

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[2y_t\nu_{1\alpha}^{'}\frac{\tilde{r}_t}{N}-\nu_{1\alpha}^{'}\frac{\tilde{r}_t\tilde{r}_t^{'}}{N^2}\nu_{1\alpha}-E\left[2y_t\nu_{1\alpha}^{'}\frac{\tilde{r}_t}{N}-\nu_{1\alpha}^{'}\frac{\tilde{r}_t\tilde{r}_t^{'}}{N^2}\nu_{1\alpha}\right]\right] =$$

$$\begin{array}{rr} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}2y_t\nu_1^{'}\frac{\tilde{r}_t}{N}+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}2y_t\frac{\tilde{r}_t^{'}}{N}(\nu_{1\alpha}-\nu_1) & - \\ <\sqrt{T}(\hat{\Sigma}-\Sigma)(\nu_{1\alpha}-\nu_1),\nu_{1\alpha}-\nu_1>_N & - \\ 2<\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_1,\nu_{1\alpha}-\nu_1>_N & - \\ <\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_1,\nu_1>_N & . \end{array}$$

As $N,\ T\rightarrow\infty$ and $\alpha\rightarrow 0$, $<\sqrt{T}(\hat{\Sigma}-\Sigma)(\nu_{1\alpha}-\nu_1),\nu_{1\alpha}-\nu_1>_N\leq\|\sqrt{T}(\hat{\Sigma}-\Sigma)\|_N\|\nu_{1\alpha}-\nu_1\|_N^2\rightarrow 0$ and $<\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_1,\nu_{1\alpha}-\nu_1>_N\leq\|\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_1\|_N\|\nu_{1\alpha}-\nu_1\|_N\rightarrow 0$. So (35) is equivalent to $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}2y_t\nu_1^{'}\frac{\tilde{r}_t}{N}+<\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_1,\nu_1>_N$ which converges to a normal distribution using Lemma 6 and 7.

Therefore, Equation (35) converges to a normal distribution with variance

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

Finally, we have

$$\begin{array}{rcl} |\sqrt{T}\left(\delta_\alpha^2-\delta^2\right)|^2 & = & T<e,\left(\Sigma_\alpha^{-1}-\Sigma^{-1}\right)e>_N^2 . \\ & \leq T & <\Sigma_\alpha^{-1}e,\left(\Sigma-\Sigma_\alpha\right)\Sigma^{-1}e>_N^2 . \end{array}$$

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

For Tikhonov, we have

$$\begin{array}{rcl} <\Sigma_\alpha^{-1}e,\left(\Sigma-\Sigma_\alpha\right)\Sigma^{-1}e>_N & = & \sum_j\left(\lambda_j-\frac{\lambda_j^2+\alpha}{\lambda_j}\right)(\phi_j,\Sigma_\alpha^{-1}e)_N(\phi_j,\Sigma^{-1}e)_N \\ & = & -\alpha\sum_j\frac{1}{\lambda_j}(\phi_j,\Sigma_\alpha^{-1}e)_N(\phi_j,\Sigma^{-1}e)_N \\ |<\Sigma_\alpha^{-1}e,\left(\Sigma-\Sigma_\alpha\right)\Sigma^{-1}e>_N|^2 & \leq & \alpha^2\sum_j\frac{1}{\lambda_j^2}(\phi_j,\Sigma^{-1}e)_N^2\sum_j(\phi_j,\Sigma_\alpha^{-1}e)_N^2=O(\alpha^2) \end{array}$$

as $e\in\Phi_3$. Therefore, $|\sqrt{T}\left(\delta_\alpha^2-\delta^2\right)|^2=O(\alpha^2T)$.

9.7 Proof of Proposition 6

The distribution of the difference of HJ distances follows from Proposition 5, which gives the distribution for each model.

10 Appendix C: List of the Portfolios used in the simulations

Table 8: List of portfolios

25 Portfolios Formed on Size and Book-to-Market
49 Portfolios Formed Industry
25 Portfolios Formed on Size and market beta
10 Portfolios formed on Industry
Portfolios Formed on Operating Profitability
Portfolios Formed on Investment
Portfolios Formed on Size
Portfolios Formed on market beta
Portfolios Formed on Book-to-Market
Portfolios Formed on Earnings/Price
Portfolios Formed on Cashflow/Price
Portfolios Formed on Dividend Yield

References

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

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.

Bailey, N., G. Kapetanios, and M. H. Pesaran (2021): "Measurement of factor strength: Theory and practice," *Journal of Applied Econometrics*, 36(5), 587–613.

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

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

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.

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.

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.

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.

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.

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.

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.

——— (2014): "Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors," *The Review of Financial Studies*, 27(7), 2139–2170.

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

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.

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

Jagannathan, R., G. Skoulakis, and Z. Wang (2010): "CHAPTER 14 - The Analysis of the Cross-Section of Security Returns," in *Handbook of Financial Econometrics: Applications*, ed. by Y. Aït-Sahalia, and L. P. Hansen, vol. 2, pp. 73–134. Elsevier, San Diego.

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.

Korsaye, S. A., A. Quaini, and F. Trojani (2019): "Smart SDFs," in *December 2019 Finance Meeting EUROFIDAI - ESSEC*, Paris.

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.

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.

——— (2020): "The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation," *Journal of Financial Econometrics*.

Lewellen, J., S. Nagel, and J. Shanken (2010): "A skeptical appraisal of asset pricing tests," *Journal of Financial Economics*, 96(2), 175–194.

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.

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

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.

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

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.

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

Yogo, M. (2006): "A Consumption-Based Explanation of Expected Stock Returns," *The Journal of Finance*, 61(2), 539–580.