Université de Montréal

Misspecified financial models in a data-rich environment

par

Cheikh Nokho

Département de sciences économiques
Faculté des arts et des sciences

Thèse présentée en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en sciences économiques

March 18, 2024

Université de Montréal

Faculté des arts et des sciences

Cette thèse intitulée

Misspecified financial models in a data-rich environment

présentée par

Cheikh Nokho

sera évaluée par un jury composé des personnes suivantes :

Benoit Perron
(président-rapporteur)

Marine Carrasco
(directrice de recherche)

René Garcia
(membre du jury)

Bertille Antoine
(examinatrice externe)

Louis Doray
(représentant du doyen de la FESP)

Remerciements

Je remercie infiniment ma directrice de recherche pour avoir accepté de m'encadrer durant ce projet de thèse. Elle a été très disponible à collaborer et a su partager ses connaissances avec moi. J'ai énormément appris auprès d'elle sur le processus de recherche. Un grand merci pour les conseils de rédaction, de méthodologie et le support financier.

Je voudrais aussi remercier René Garcia pour m'avoir offert un support financier durant ma dernière année et Benoit Perron pour ses remarques et conseils lors des séminaires Dagenais. Ils se sont impliqués comme membres de mon comité de thèse dès les premières années.

Mes remerciements sont également adressés au Centre Interuniversitaire de Recherche en Économie Quantitative (CIREQ) et au Fonds de Recherche du Québec – Société et Culture (FRQSC). Cette thèse a été réalisée avec leur appui financier.

Je remercie également mes collègues avec qui j'ai partagé cette étape importante de mon parcours académique et toutes les personnes dont j'ai fait la connaissance à Montréal et qui m'ont beaucoup soutenu et aidé.

Je voudrais particulièrement remercier mon père, Abou, et ma mère, Kédiagou, pour m'avoir continuellement soutenu dans mes entreprises scolaires. J'ai une profonde reconnaissance envers ma conjointe, Awa, pour sa patience durant la recherche, ses encouragements et son soutien.

Enfin, je remercie toutes les personnes qui m'ont aidé de près ou de loin durant ce processus : la liste est longue.

Résumé

En finance, les modèles d'évaluation des actifs tentent de comprendre les différences de rendements observées entre divers actifs. Hansen and Richard (1987) ont montré que ces modèles sont des représentations fonctionnelles du facteur d'actualisation stochastique que les investisseurs utilisent pour déterminer le prix des actifs sur le marché financier. La littérature compte de nombreuses études économétriques qui s'intéressent à leurs estimations et à la comparaison de leurs performances, c'est-à-dire de leur capacité à expliquer les différences de rendement observées. Cette thèse, composée de trois articles, contribue à cette littérature.

Le premier article examine l'estimation et la comparaison des modèles d'évaluation des actifs dans un environnement riche en données. Nous mettons en œuvre deux méthodes de régularisation interprétables de la distance de Hansen and Jagannathan (1997, HJ ci-après) dans un contexte où les actifs sont nombreux. Plus précisément, nous introduisons la régularisation de Tikhonov et de Ridge pour stabiliser l'inverse de la matrice de covariance de la distance de HJ. La nouvelle mesure, qui en résulte, peut être interprétée comme la distance entre le facteur d'actualisation d'un modèle et le facteur d'actualisation stochastique valide le plus proche qui évalue les actifs avec des erreurs contrôlées. Ainsi, ces méthodes de régularisation relâchent l'équation fondamentale de l'évaluation des actifs financiers. Aussi, elles incorporent un paramètre de régularisation régissant l'ampleur des erreurs d'évaluation. Par la suite, nous présentons une procédure pour estimer et faire des tests sur les paramètres d'un modèle d'évaluation des actifs financiers avec un facteur d'actualisation linéaire en minimisant la distance de HJ régularisée. De plus, nous obtenons la distribution asymptotique des estimateurs lorsque le nombre d'actifs devient grand. Enfin, nous déterminons la distribution de la distance régularisée pour comparer différents modèles d'évaluation des actifs. Empiriquement, nous estimons et comparons quatre modèles à l'aide d'un ensemble de données comportant 252 portefeuilles.

Le deuxième article estime et compare dix modèles d'évaluation des actifs, à la fois inconditionnels et conditionnels, en utilisant la distance de HJ régularisée et 3 198 portefeuilles s'étendant de juillet 1973 à juin 2018. Ces portefeuilles combinent les portefeuilles bien connus triés par caractéristiques avec des micro-portefeuilles. Les micro-portefeuilles sont formés à l'aide de variables financières mais contiennent peu d'actions (5 à 10), comme indiqué dans Barras (2019). Par conséquent, ils sont analogues aux actions individuelles, offrent une grande variabilité de rendements et améliorent le pouvoir discriminant des portefeuilles classiques triés par caractéristiques. Parmi les modèles considérés, quatre sont des modèles macroéconomiques ou théoriques, dont le modèle de CAPM avec consommation (CCAPM), le modèle de CAPM avec consommation durable (DCAPM) de Yogo (2006), le modèle de CAPM avec capital humain (HCAPM) de Jagannathan and Wang (1996), et le modèle d'évaluation des actifs avec intermédiaires financiers (IAPM) de He, Kelly, and Manela (2017). Cinq modèles basés sur les anomalies sont considérés, tels que les modèles à trois (FF3) et à cinq facteurs (FF5) proposés par Fama and French, 1993 et 2015, le modèle de Carhart (1997) intégrant le facteur Momentum dans FF3, le modèle de liquidité de Pástor and Stambaugh (2003) et le modèle q5 de Hou et al. (2021). Le modèle de consommation de Lettau and Ludvigson (2001) utilisant des données trimestrielles est également estimé. Cependant, il n'est pas inclus dans les comparaisons en raison de la puissance de test réduite. Par rapport aux modèles inconditionnels, les modèles conditionnels tiennent compte des cycles économiques et des fluctuations des marchés financiers en utilisant les indices d'incertitude macroéconomique et financière de Ludvigson, Ma, and Ng (2021). Ces modèles conditionnels ont des erreurs de spécification considérablement réduites. Les analyses comparatives des modèles inconditionnels indiquent que les modèles macroéconomiques présentent globalement les mêmes pouvoirs explicatifs. De plus, ils ont un pouvoir explicatif global inférieur à celui des modèles basés sur les anomalies, à l'exception de FF3. L'augmentation de FF3 avec le facteur Momentum et de liquidité améliore sa capacité explicative. Cependant ce nouveau modèle est inférieur à FF5 et q5. Pour les modèles conditionnels, les modèles macroéconomiques DCAPM et HCAPM surpassent CCAPM et IAPM. En outre, ils ont des erreurs de spécification similaires à celles des modèles conditionnels de Carhart et de liquidité, mais restent en deçà des modèles FF5 et q5. Ce dernier domine tous les autres modèles.

Le troisième article présente une nouvelle approche pour estimer les paramètres du facteur d'actualisation linéaire des modèles d'évaluation d'actifs linéaires mal spécifiés avec de nombreux actifs. Contrairement au premier article de Carrasco and Nokho

(2022), cette approche s'applique à la fois aux rendements bruts et excédentaires. La méthode proposée régularise toujours la distance HJ : l'inverse de la matrice de second moment est la matrice de pondération pour les rendements bruts, tandis que pour les rendements excédentaires, c'est l'inverse de la matrice de covariance. Plus précisément, nous dérivons la distribution asymptotique des estimateurs des paramètres du facteur d'actualisation stochastique lorsque le nombre d'actifs augmente. Nous discutons également des considérations pertinentes pour chaque type de rendements et documentons les propriétés d'échantillon fini des estimateurs. Nous constatons qu'à mesure que le nombre d'actifs augmente, l'estimation des paramètres par la régularisation de l'inverse de la matrice de covariance des rendements excédentaires présente un contrôle de taille supérieur par rapport à la régularisation de l'inverse de la matrice de second moment des rendements bruts. Cette supériorité découle de l'instabilité inhérente à la matrice de second moment des rendements bruts. De plus, le rendement brut de l'actif sans risque présente une variabilité minime, ce qui entraîne une colinéarité significative avec d'autres actifs que la régularisation ne parvient pas à atténuer.

Mots clés : Econométrie financière, modèles d'évaluation des actifs, modèles de grande dimension, distance de Hansen-Jagannathan, sélection de modèles, spécification erronée d'un modèle, méthodes de régularisation.

Abstract

In finance, asset pricing models try to understand the differences in expected returns observed among various assets. Hansen and Richard (1987) showed that these models are functional representations of the discount factor investors use to price assets in the financial market. The literature counts many econometric studies that deal with their estimation and the comparison of their performance, i.e., how well they explain the differences in expected returns. This thesis, divided into three chapters, contributes to this literature.

The first paper examines the estimation and comparison of asset pricing models in a data-rich environment. We implement two interpretable regularization schemes to extend the renowned Hansen and Jagannathan (1997, HJ hereafter) distance to a setting with many test assets. Specifically, we introduce Tikhonov and Ridge regularizations to stabilize the inverse of the covariance matrix in the HJ distance. The resulting misspecification measure can be interpreted as the distance between a proposed pricing kernel and the nearest valid stochastic discount factor (SDF) pricing the test assets with controlled errors, relaxing the Fundamental Equation of Asset Pricing. So, these methods incorporate a regularization parameter governing the extent of the pricing errors. Subsequently, we present a procedure to estimate the SDF parameters of a linear asset pricing model by minimizing the regularized distance. The SDF parameters completely define the asset pricing model and determine if a particular observed factor is a priced source of risk in the test assets. In addition, we derive the asymptotic distribution of the estimators when the number of assets and time periods increases. Finally, we derive the distribution of the regularized distance to compare comprehensively different asset pricing models. Empirically, we estimate and compare four empirical asset pricing models using a dataset of 252 portfolios.

The second paper estimates and compares ten asset pricing models, both unconditional and conditional, utilizing the regularized HJ distance and 3198 portfolios spanning

July 1973 to June 2018. These portfolios combine the well-known characteristic-sorted portfolios with micro portfolios. The micro portfolios are formed using firms' observed financial characteristics (e.g. size and book-to-market) but contain few stocks (5 to 10), as discussed in Barras (2019). Consequently, they are analogous to individual stocks, offer significant return spread, and improve the discriminatory power of the characteristics-sorted portfolios. Among the models, four are macroeconomic or theoretical models, including the Consumption Capital Asset Pricing Model (CCAPM), Durable Consumption Capital Asset Pricing Model (DCAPM) by Yogo (2006), Human Capital Capital Asset Pricing Model (HCAPM) by Jagannathan and Wang (1996), and Intermediary Asset pricing model (IAPM) by He, Kelly, and Manela (2017). Five anomaly-driven models are considered, such as the three (FF3) and Five-factor (FF5) Models proposed by Fama and French, 1993 and 2015, the Carhart (1997) model incorporating momentum into FF3, the Liquidity Model by Pástor and Stambaugh (2003), and the Augmented q-Factor Model (q5) by Hou et al. (2021). The Consumption model of Lettau and Ludvigson (2001) using quarterly data is also estimated but not included in the comparisons due to the reduced power of the tests. Compared to the unconditional models, the conditional ones account for the economic business cycles and financial market fluctuations by utilizing the macroeconomic and financial uncertainty indices of Ludvigson, Ma, and Ng (2021). These conditional models show significantly reduced pricing errors. Comparative analyses of the unconditional models indicate that the macroeconomic models exhibit similar pricing performances of the returns. In addition, they display lower overall explanatory power than anomaly-driven models, except for FF3. Augmenting FF3 with momentum and liquidity factors enhances its explanatory capability. However, the new model is inferior to FF5 and q5. For the conditional models, the macroeconomic models DCAPM and HCAPM outperform CCAPM and IAPM. Furthermore, they have similar pricing errors as the conditional Carhart and liquidity models but still fall short of the FF5 and q5. The latter dominates all the other models.

This third paper introduces a novel approach for estimating the SDF parameters in misspecified linear asset pricing models with many assets. Unlike the first paper, Carrasco and Nokho (2022), this approach is applicable to both gross and excess returns as test assets. The proposed method still regularizes the HJ distance : the inverse of the second-moment matrix is the weighting matrix for the gross returns, while for excess returns, it is the inverse of the covariance matrix. Specifically, we derive the asymptotic distribution of the SDF estimators under a double asymptotic condition

where the number of test assets and time periods go to infinity. We also discuss relevant considerations for each type of return and document the finite sample properties of the SDF estimators with gross and excess returns. We find that as the number of test assets increases, the estimation of the SDF parameters through the regularization of the inverse of the excess returns covariance matrix exhibits superior size control compared to the regularization of the inverse of the gross returns second-moment matrix. This superiority arises from the inherent instability of the second-moment matrix of gross returns. Additionally, the gross return of the risk-free asset shows minimal variability, resulting in significant collinearity with other test assets that the regularization fails to mitigate.

Keywords : Financial Econometrics, Asset Pricing models, High-dimensional models, Hansen-Jagannathan distance, Model selection, Model misspecification, Regularization methods.

Table des matières

Liste des tableaux

Table des figures

Chapitre 1

Hansen-Jagannathan distance with many assets *

1.1 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 the null hypothesis of correct specification is often rejected (Hodrick and Zhang (2001) ; Ludvigson, 2013). Therefore, models are usually misspecified. 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 has been 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

*. This chapter is co-authored with Marine Carrasco.

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 is larger than the length of the time-series. Cochrane (2005) advanced that a number of assets larger than $1/10$ of the time periods 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 et al. (2010) discussed. Therefore, when the GMM optimal matrix is poorly estimated, using an identity matrix for the first-step GMM may offer greater robustness compared to using the estimated optimal matrix.

This paper examines the evaluation and comparison of asset pricing models with many test assets, therefore an unstable covariance matrix. Our main contributions can be summarized as follows. First, relying on the inverse problem literature (see Carrasco et al. (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 et al. (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 et al., 2019 ; Ledoit and Wolf, 2003 ; Ledoit and Wolf, 2020). This paper is also related to the work of Korsaye et al. (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 estimated parameters with a misspecified GMM.

The paper is organized as follows. Section 1.2 presents the framework under which we evaluate models and the issues related to the weighting matrix. Section 1.3 introduces several regularization methods as well as their interpretations. The section also presents the asymptotic properties of the SDF parameter estimator. Section 1.4 treats model comparison using regularization, and section 1.5 contains the results of the simulations. Section 1.6 compares four empirical asset pricing models using a dataset of 252 portfolios. Finally, section 1.7 concludes. The proofs are collected in the appendix.

1.2 Asset pricing model under misspecification

1.2.1 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

$$Q_W = e(\theta)' W e(\theta), \tag{1.1}$$

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

$$Q_{V_{22}} = \delta^2 = (\mu_2 - V_{21}\theta)^{'} V_{22}^{-1}(\mu_2 - V_{21}\theta). \tag{1.2}$$

We define θ_{HJ} as the solution to the minimization of (1.2).

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

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

$$r_t = e + \beta(\tilde{f}_t + \gamma) + \epsilon_t, \tag{1.3}$$

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 $\sigma_{i,j} = E[\epsilon_{it}\epsilon_{jt}]$. We note $\sigma_i^2 = \sigma_{i,i}$ and $\epsilon = [\epsilon_1, ..., \epsilon_T]^{'}$. Remark that Equation (1.3) 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 $\bar{f}_t = f_t - \hat{\mu}_1$. $\hat{\mu}_1 = \frac{1}{T}\sum_{t=1}^{T} f_t$ and $\hat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T} r_t$ 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,

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

where

$$V(\hat{\theta}_{HJ}) = \sum_{j=-\infty}^{\infty} E[q_t q_{t+j}'], \tag{1.4}$$

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

1.2.2 Issues with the weighting matrix

When models are misspecified, the SDF parameter, that minimizes (1.1), depends on the weighting matrix. Therefore, its choice is primordial.

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 setting to compare asset pricing models may be misleading for several reasons.

First, in this case, the objective function (1.1) 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 Cochrane (2005, p. 216). In addition, near singularity deteriorates the small sample properties of the SDF estimator or misspecification test.

1.3 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 et al. (2007)), we introduce two regularization methods to stabilize the weighting matrix and improve the estimation of asset pricing models.

1.3.1 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 $\| v \|_T^2 = \frac{v'v}{T}$ generated 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. $H^*v = \frac{\bar{R}'v}{T}$, 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.

(i) For a vector $v \in \mathbb{R}^N$, $\parallel v \parallel$ is the euclidian norm.

(ii) For an arbitrary $(K \times N)$ matrix V, the operator norm of V is $\parallel V \parallel = \underset{\parallel\phi\parallel=1}{sup} \parallel V\phi \parallel$. Therefore, for any vector $u \in \mathbb{R}^N$, $\parallel Vu \parallel \leq \parallel V \parallel \parallel u \parallel$.

(iii) Let $\{\phi_j\}_{j=1,\cdots,N}$ be a complete orthonormal basis in $\mathbb{R}^N$. For any $\phi \in R^N$, $\parallel \phi \parallel_N^2 = \sum_{i=1}^N < \phi, \phi_j >_N^2$ and if V is a $(N \times N)$ symmetric matrix, we define the following operator norm $\parallel V \parallel_N = \underset{\parallel\phi\parallel_N=1}{sup} < V\phi, \phi >_N$.

(iv) We define the Frobenius norm as $\parallel V \parallel_F = \left(tr(V'V)\right)^{\frac{1}{2}}$. We have $\parallel V \parallel \leq \parallel V \parallel_F$ and for any vector $u \in \mathbb{R}^N$, $\parallel Vu \parallel \leq \parallel V \parallel_F \parallel u \parallel$.

(v) If $\parallel v \parallel_N < \infty$ when $N \to \infty$, we note $\parallel v \parallel_\infty$ its limit value.

Assumption 1. *As $N \to \infty$,*

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

(ii) $\parallel e \parallel_N = O(1)$.

Remark 2. The first part of assumption 1 is the same as assumption 2 of Raponi et al. (2020). Positive-definite Σ_β excludes spurious factors and cross sectionally constant β_i. Also, this assumption implies that $\parallel \beta_k \parallel_\infty < \infty, k = 1, \cdots, K$.

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 $\rho > 0$. $\alpha_x(l) = \underset{i,k\geq 1}{sup}\,\underset{A,B}{sup}\left[\mid P(A \cap B) - P(A)P(B) \mid: A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_l^\infty\right]$, 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.*

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

Lemma 1. *Under assumption 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. **(i)** 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 $\{\psi \in E : \psi = A\phi \; for \; some \; \phi \in G \; such \; that \; \| \phi \|_H < \infty\}$.

(ii) For a positive self-adjoint compact operator with spectrum $\{\lambda_j, \varphi_j, j = 1, \cdots\}$ $\Sigma : G \to G$ 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

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

(iii) The reproducing Kernel Hilbert pace (RKHS) $\mathcal{H}(\Sigma)$ of the operator Σ corresponds to $\omega = \frac{1}{2}$.

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.

1.3.2 Regularization as penalization

As pointed out by Kan and Robotti (2008), δ^2 gives the distance between the proposed SDF y_t and the set of correct SDFs of mean 1 in $\mathcal{M}$, set of square integrable random variables.

$$\delta^2 = \underset{m_t \in \mathcal{M}, E[m_t]=1}{min} E\left(m_t - y_t\right)^2 \quad subject\, to\, E[m_t r_t] = 0. \tag{1.5}$$

To account for increasing N, we normalize r_t by N, omit the subscript t to simplify the presentation, and get the problem below

$$\underset{m \in \mathcal{M}, E[m]=1}{min} E\left(m - y\right)^2 \quad subject : to : E[\frac{r}{N} m] = 0.$$

Using the saddle problem, and the fact that $E[y] = 1$, Kan and Robotti (2008) rewrite the problem as below

$$\begin{aligned}\delta^2 &= \underset{m}{min} : \underset{\nu_1 \in \mathbb{R}^N, \nu_2 \in \mathbb{R}}{max} E\left\{(y-m)^2 + 2\nu_1^{'}\frac{r}{N}m + 2\nu_2(E[m]-1)\right\} \\ &= \underset{m}{min} : \underset{\nu_1 \in \mathbb{R}^N, \nu_2 \in \mathbb{R}}{max} E\left\{(y - \nu_1^{'}\frac{r}{N} - \nu_2 - m)^2 + 2\nu_1^{'}\frac{r}{N}y + 2\nu_2 y \right. \\ &\quad \left. -\nu_1^{'}\frac{rr^{'}}{N^2}\nu_1 - 2\nu_1^{'}\frac{r}{N}\nu_2 - \nu_2^2 - 2\nu_2\right\}.\end{aligned}$$

Fixing the Lagrange multipliers and solving the minimization problem with respect to m yields [2] the following dual problem

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

The first order condition related to ν_2 is given by $E\left\{\frac{r}{N}^{'}\nu_1 + \nu_2\right\} = 0$. We can eliminate ν_2 from the previous problem and obtain

$$\delta^2 = \underset{\nu_1 \in \mathbb{R}^N}{max} E\left\{y^2 - (y - \nu_1^{'}(\frac{r}{N} - \frac{\mu_2}{N}))^2 + 2\nu_1^{'}\frac{\mu_2}{N}\right\}. \tag{1.6}$$

The resulting ν_1 is $\nu_1^* = \Sigma^{-1}e$, where $e = E[ry]$ and

$$\delta^2 = \frac{e^{'}\Sigma^{-1}e}{N}.$$

To take into account the large number of assets (big N), we penalize the Lagrange multiplier ν_1 in (1.6) as follows :

$$\delta_R^2 = \underset{\nu_1 \in \mathbb{R}^N}{max} E\left\{y^2 - (y - \nu_1^{'}(\frac{r}{N} - \frac{\mu_2}{N}))^2 + 2\nu_1^{'}\frac{\mu_2}{N}\right\} - \frac{\alpha}{\mathbf{N}}\nu_{\mathbf{1}}^{'}\nu_{\mathbf{1}}.$$

This yields $\nu_{1\alpha}^* = [\Sigma + \alpha I_N]^{-1}E[ry] = [\Sigma + \alpha I_N]^{-1}e$ and

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

As a consequence, this penalization leads to the Ridge regularization where the weighting matrix is $(\Sigma + \alpha I_N)^{-1}$.

2. The minimum is attained by setting $m = y - \nu_1^{'}\frac{r}{N} - \nu_2$

In the same vein, we obtain the Tikhonov regularization through the problem below where the penalization depends on Σ^{-1}.

$$\begin{aligned}
\delta_K^2 &= \max_{\nu_1 \in \mathbb{R}^N} : E\left\{y^2 - (y - \nu_1'(\frac{r}{N} - \frac{\mu_2}{N}))^2 + 2\nu_1'\frac{\mu_2}{N}\right\} - \alpha \nu_{\mathbf{1}}' \mathbf{V}_{\mathbf{22}}^{-\mathbf{1}} \nu_{\mathbf{1}} \\
&= \max_{\nu_1 \in \mathbb{R}^N} : E\left\{y^2 - (y - \nu_1'(\frac{r}{N} - \frac{\mu_2}{N}))^2 + 2\nu_1'\frac{\mu_2}{N}\right\} - \frac{\alpha}{\mathbf{N}} \nu_{\mathbf{1}}' \mathbf{\Sigma}^{-\mathbf{1}} \nu_{\mathbf{1}}.
\end{aligned}$$

It yields $\nu_{1\alpha}^* = [\Sigma^2 + \alpha I_N]^{-1}\Sigma E[ry] = [\Sigma^2 + \alpha I_N]^{-1}\Sigma e$.

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

1.3.3 Regularization as Minimum Discrepancy SDF solution

There is another interpretation of the regularized HJ-distance. δ_α^2 measures how far y is to the closest valid SDF of mean 1 which prices returns with errors controlled by α. To prove it, we make the following assumption.

Assumption 4. *$\exists\, m_0 \in L^2 : E\,[m_0] = 1$, and $\parallel E[m_0 r] \parallel_N^2 < \infty$.*

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,

$$\delta_R^2 = \inf_{m \in \mathcal{M}, E[m]=1} E[(m-y)^2] + \frac{1}{\alpha} \parallel E[mr] \parallel_N^2, \tag{1.7}$$

2. For Tikhonov,

$$\delta_K^2 = \inf_{m \in \mathcal{M}, E[m]=1} E[(m-y)^2] + \frac{1}{\alpha} \parallel E[mr] \parallel_{N,\Sigma}^2, \tag{1.8}$$

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

Remark 8. The previous proposition shows that regularization is equivalent to relaxing the constraint of problem (1.5). 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.

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 following :

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

where $\omega_j = \frac{\lambda_j}{\alpha}$. $(Pr)_j$ 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.

Remark 9. 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).

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

where $\tau > 0$ and $h : \mathbb{R}^D \to [0, +\infty]$ 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 (1.7) 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. In Proposition 1, we have one sure asset $R_s = 1$ with price $q_s = 1$ and N dubious assets.

1.3.4 Asymptotic distribution of the regularized SDF parameter of misspecified models

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

$$\hat{\theta}^{\alpha}_{HJ} = \underset{\theta}{argmin}\,(\hat{\mu}_2 - \hat{V}_{21}\theta)'\hat{\Sigma}^{-1}_{\alpha}(\hat{\mu}_2 - \hat{V}_{21}\theta). \tag{1.9}$$

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

and the regularized HJ-distance is

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

$\hat{\Sigma}^{-1}_{\alpha}$ 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 $\hat{\theta}^{\alpha}_{HJ}$ can be decomposed as such

$$\begin{aligned}\hat{\theta}^{\alpha}_{HJ} - \theta_{HJ} = {}& (\hat{V}^{-1}_{11} - V^{-1}_{11})\gamma + \hat{V}^{-1}_{11}(\hat{\mu}_1 - \mu_1) \\ &+ \hat{V}^{-1}_{11}(\hat{\beta}'\hat{\Sigma}^{-1}_{\alpha}\hat{\beta})^{-1}\,[\,\hat{\beta}'\hat{\Sigma}^{-1}_{\alpha}(\beta - \hat{\beta})(\gamma + \hat{\mu}_1 - \mu_1) \\ &+ \hat{\beta}'\hat{\Sigma}^{-1}_{\alpha}e + \hat{\beta}'\hat{\Sigma}^{-1}_{\alpha}\bar{\epsilon}\,].\end{aligned} \tag{1.10}$$

Equivalence between Ridge and Tikhonov.

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

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

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 which 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) $E[f_{kt}^{4+2\rho}] < \infty$ *for* $k=1,\dots,K$.

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

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

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

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

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

Proposition 2. *Suppose Assumption 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 $\Omega = \underset{N,T\to\infty}{lim} var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_t\right]$. *$h_t$ 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 a nonparametric heteroskedasticity and autocorrelation consistent (HAC) estimator of Newey and West (1987) using the sample analog of h_t . The results of Proposition 2 are key 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.

1.3.5 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 α that maximizes the out-of-sample R-square, R^2_{oos}.

$$R^2_{oos} = 1 - \frac{(\mu_2^o - \beta^o \hat{\gamma}_\alpha)'(\mu_2^o - \beta^o \hat{\gamma}_\alpha)}{\mu_2^{o'} \mu_2^o}, \tag{1.11}$$

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

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

The analysis of this section is similar to section 2 of Kan and Robotti (2009). 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 asset pricing models are respectively

$$r_t = e_1 + \beta_1(x_{1t} - E[x_{1t}] + \gamma_1) + \epsilon_{1t}, \tag{1.12}$$

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.

Two asset-pricing models have equal HJ distance under two circumstances. They can have the same SDF or aggregate pricing errors when their SDFs are distinct. As a result, we can compare asset pricing models by looking at the SDF parameters when the models are nested or non-nested with overlapping factors. When the models are non-nested with distinct SDFs, we will rely on the expression of their aggregate pricing errors, the regularized HJ distance.

1.4.1 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

$$\delta_1^2 - \delta_2^2 = \theta_{22}' C_{2,22}^{-1} \theta_{22}. \tag{1.13}$$

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. *$\delta_1^2 = \delta_2^2$ $(y_1 = y_2)$ 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 $\alpha^2 T \to 0$,

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

where $\chi^2_j(1)$ are independent chi-square random variables and ξ_j are eigenvalues defined in Appendix.

Remark 10. 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 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 (1.12). On the other hand, we can compute the HJ distance difference of the two models using the same level of penalization or (1.13) and use the statistics $T(\hat{\delta}^2_{1,\alpha} - \hat{\delta}^2_{2,\alpha})$ to compare them.

1.4.2 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 $\delta^2_1 = \delta^2_2$. Both cases need to be treated separately.

1.4.2.1 Test of SDFs equality

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

$$\delta^2_1 - \delta^2_2 = -\theta'_{12} C^{-1}_{1,22} \theta_{12} + \theta'_{22} C^{-1}_{2,22} \theta_{22}.$$

The following proposition outlines the principal 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 $\alpha^2 T \to 0$,

$$T(\hat{\delta}^2_{1,\alpha} - \hat{\delta}^2_{2,\alpha}) \xrightarrow{d} \sum_{i=1}^{K_3} \xi_i \chi^2(1) \tag{1.14}$$

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

Remark 11. Proposition 4.1. shows that to compare asset pricing models with overlapping factors, one can test the simultaneous nullity of the common factors of the two SDFs (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. We can also realize the test by estimating the asset pricing model with factors f_1, f_2, and f_3 and test the nullity of the parameter associated with f_2 and f_3. These options do not directly test the nullity of the difference in HJ distance, but the equality of the SDF of the two models.

1.4.2.2 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 (1.5).

It is noteworthy to see 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.

Using Assumptions 1-4, we give the distribution of the penalized HJ distance when models are misspecified. To do so, we use the following expression of the penalized HJ distance (δ_p^2)

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

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

The pricing errors are estimated using $\hat{e} = \frac{1}{T}\sum_{t=1}^{T} r_t y_t$.

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

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

$$v_4 = \underset{N,T\to\infty}{lim} var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{l}_t\right],$$

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

Proposition 6 gives the distribution of Penalized HJ distance using the errors. It can be used to compare two asset pricing models as presented in 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 $\delta^2_1, \delta^2_2 \neq 0$. As T, N go to infinity and α goes to zero, if α is chosen such that $\alpha T \to \infty$ and $\alpha^2 T \to 0$,*

$$\sqrt{T}\left((\hat{\delta}^2_{1\alpha} - \hat{\delta}^2_{2\alpha}) - (\delta^2_1 - \delta^2_2)\right) \xrightarrow{d} \mathcal{N}(0, v_5),$$

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 penalization.

Summary. The various asymptotic distributions employed in comparing asset pricing models suggest the following sequential procedure. When evaluating nested or non-nested models with overlapping factors, the initial step involves testing the hypothesis $H_0 : y_1 = y_2$ using Proposition 3 and 4. If this hypothesis is rejected, then we proceed to test $H_0 : \delta^2_1 = \delta^2_2 \neq 0$ by using Proposition 6. Denote ζ_1 and ζ_2 the significance level of the first and second tests, the global significance level of this sequential procedure is $max\{\zeta_1, \zeta_2\}$. Alternatively, we can directly apply Proposition 6 by assuming that models with differing factors inherently possess different Stochastic Discount Factors (SDFs), as suggested by Kan and Robotti (2009).

1.4.3 Multiple comparison

When several models are being evaluated, pairwise comparisons might not clearly identify the best-performing model. In this section, we establish a formal test for comparing multiple models, applicable to both non-nested and nested models. The test is based on the work of Wolak (1989).

Suppose we have $p + 1$ models with HJ distance given by $\delta_i, \quad i = 1, \ldots, p + 1$. We are interested in testing whether a benchmark model (model 1) has an aggregate pricing errors as low as the other p models. Let $d_i = \delta^2_1 - \delta^2_i,\ i = 2, \ldots, p + 1$ 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 $H_0 : d \leq 0_p$while the alternative is $H_1 : d \in \mathbb{R}^p$. 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 $\delta_i > 0$ and the models have distinct SDFs. The test uses the sample counterpart of d, ie $\hat{d}_\alpha = \begin{pmatrix} d_{1\alpha} & \cdots & d_{p+1\alpha} \end{pmatrix}$ for a fixed value of α of the benchmark model. Let $\tilde{d}_\alpha$ be the optimal solution in the following quadratic programming problem

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

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

$$LR_\alpha = T(\hat{d}_\alpha - \tilde{d}_\alpha)' \hat{\Omega}_{d,\alpha}^{-1} (\hat{d}_\alpha - \tilde{d}_\alpha). \tag{1.15}$$

The distribution of the previous statistics is obtained under the least favorable value, ie $d = 0$. 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 [3] and $\chi^2(i)$ are independent Chi-square random variables with i degrees of freedom.

1.5 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 et al. (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

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

3. Appendix C of Gospodinov et al. (2013) gives the procedure to compute $w_i(\Omega_d)$ and the p-value of the test

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 2.10 in the Appendix.

1.5.1 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). This corresponds to testing whether a SDF parameter is null. We compare the small size properties of our test with the one in Kan and Robotti (2008) using (1.4).

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 as $N > T$. 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 et al. (2014)). Therefore, we can conclude that taking the generalized inverse does not guarantee appropriate test behavior when N is large.

Table 1.1 – Empirical size of the Kan and Robotti (2008) test with 252 assets

T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A : Three factors model of Fama and French (1993)									
	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 : Non-linear 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 select the value of alpha by evaluating a grid of 15 values ranging from 0.001 to 0.1 aiming to maximize the out-of-sample R^2 : we use half of the sample as training data and the remaining as a test. Particularly, we choose the smallest value of α for $T = 650$. Table 1.2 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 non-linear model (Panel C).

We now turn our attention to the empirical power of our t-test. Table 1.3 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 et al. (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 non-linear 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 1.2 – Empirical size of the Tikhonov test under misspecification with 252 assets.

T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A : Three factors model of Fama and French (1993)									
	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 : Non-linear 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 1.3 – Empirical power of the Tikhonov test under misspecification with 252 assets .

T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A : Three factors model of Fama and French (1993)									
	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 : Non-linear 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

1.5.2 Model comparison tests

In this section, we present the finite sample behavior of the pairwise and multiple comparison tests. Table 1.4 presents the results.

Panel A presents the tests developed in proposition 4. The latter verifies the equality of two non-nested SDFs. The simulated data are from FF3 and the non linear 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 estimated 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 (1.14). 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 (1.11). 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 over rejects 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 et al. (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. Therefore, it is advisable to use the least amount of penalization when comparing the model using the distribution o δ^2.

Panel C shows the finite sample behavior of the comparison test of multiple models. The test uses the statistic (1.15). 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 non-linear model. The latter has a misspecification 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 1.4 – 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

1.6 Empirical application

We employ the earlier presented asset pricing models to illustrate the regularized HJ distance. We consider the Fama and French (1993) model (FF3), the durable consumption CAPM of Yogo (2006), and the non-linear model of Dittmar (2002). We also add 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, cashflow price ratio, dividend yield, and industries. These portfolios are from Kenneth French's website and range from 1964 to 2019. Table 2.10 of the appendix presents the details of these portfolios. We estimate the SDF parameters of the four models and then compare their pricing performance. It is essential to keep the same level of penalization to compare the models. We use a data-driven approach to estimate the SDF parameters with values of α ranging from 0.001 to 0.1 and a penalization level of 0.001 for the comparison of the models.

Table 1.5 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 non-linear model, no factor exhibits significant SDF parameters. 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 lined the results of with Fama and French (2015), who argues that the value factor is redundant as the model with the five factors does not improve the model with just the four factors without HML. The model exhibits the lowest pricing errors.

Table 1.5 – 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
α	0.001				0.001		
HJ	0.123				0.139		
	Non-linear model						
Factors	θ_{mkt}	$\theta_{mkt,2}$	$\theta_{mkt,3}$				
SDF	-0.174	-0.010	0.476				
t-ratio	-0.676	-0.021	0.872				
α	0.015						
HJ	0.0428						
	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		
α	0.001						
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 1.6 presents the results of the tests. The results can be summarized as follows : FF3, YOGO, and the non-linear 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 non-linear 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_t^l. 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 1.6 – Pairwise HJ distance comparison tests

	YOGO	Non-linear	Non-linear 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)
Non-linear			0.003	0.038**
			(0.780)	(0.040)
Non-linear with human capital				-0.035*
				(0.073)

P-values are in brackets. ***,**,* indicate that the null hypothesis of unpriced source of risk is rejected at the 1%, 5%, and 10% levels.

Finally, we implement the multiple model comparison of Wolak (1989). The test compares the square HJ distance of a benchmark model against the square HJ distance of more than two models. In our context, 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 again the models, we remove FF5 from the alternative.

Table 2.9 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 non-linear model). For YOGO and the non-linear 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, the consumption and non-linear models.

Table 1.7 – Multiple model comparison tests

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

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

1.7 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. Therefore, 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 proposed 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 the 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 non-linear models represents an interesting extension.

1.8 Appendix A : Short review on regularization

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

1. $\mid \frac{q(\alpha,\mu)}{\mu} \mid < c(\alpha)$ 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 sums up in changing Σ^{-1} by $\Sigma_\alpha^{-1} = (\Sigma^2 + \alpha I_N)^{-1}\Sigma$.

1.9 Appendix B : Proofs

1.9.1 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 (1.3), 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,

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

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

From Cauchy-Schwarz inequality, we have $\mid tr(\frac{e\gamma^{'}\beta^{'}}{N}) \mid \leq tr(\frac{ee^{'}}{N}).tr(\frac{(\gamma^{'}\beta^{'})\beta\gamma}{N})$. Also $tr\left((\gamma^{'}\beta^{'})\beta\gamma\right) = tr(\beta\gamma\gamma^{'}\beta^{'})$ and $tr(\frac{\beta\gamma\gamma^{'}\beta^{'}}{N}) \leq \sqrt{tr(\frac{(\beta\beta^{'})^2}{N^2}).tr((\gamma\gamma^{'})^2)} \leq tr(\frac{\beta\beta^{'}}{N})tr(\gamma\gamma^{'})$.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.

1.9.2 Proof of Proposition 1

We transform the primal problem to be able to use the Fenchel-Rockafellar Duality (See Bauschke and Combettes, 2017, chapter 15) 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 $x = (x_1^{'}, x_{-1})^{'}\mathbb{R}^N \times \mathbb{R}$, $\chi_{\{2\}}$ 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 (1.7) can be rewritten as below

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

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 $(ri\, dom(g)) \bigcap (ri\, A(dom(f_y))) \neq \emptyset$. [4]

4. For convex set $S \subseteq \mathbb{R}^N$, $ri\, S$ is its relative interior. The latter is the interior with respect

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 $E\left[2m_0\right] = 2$, and $\| \ E[m_0 r] \ \|_N^2 < \infty$, $g(A(m_0)) = \frac{N}{4\alpha} \ \| \ E[m \frac{2r}{N}] \ \|^2 = \frac{1}{\alpha} \ \| \ E[m_0 r] \ \|_N^2 < \infty$, 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.

$f_y^*(z) = E \left\{ \sup_{w \in L^2} zw - (w-y)^2 \right\} = E\left\{ zy + \frac{1}{4} z^2 \right\}$ [5], $A^* : R^{N+1} \rightarrow L^2$ 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 $h^*(\nu_1) = \frac{\alpha}{N} \ \| \ \nu_1 \ \|^2$. [6]

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

Therefore

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

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

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

which is the penalized version (1.6). The resulting ν_1 is given by

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

We can do the same for Tikhonov by setting $h(x) = \frac{N}{4\alpha} \ \| \ x \ \|_\Sigma^2$. The latter can be

to the affine hull of S, $aff\, S$. Specifically, $ri\, S = \{x \in S : B_\epsilon(x) \bigcap aff\, S \subseteq S\}$,where $aff\, S = \{\theta_1 x_1 + \cdots + \theta_k x_k : x_1, \ldots, x_k \in S, \theta_1 + \cdots + \theta_k = 1\}$ and $B_\epsilon(x) = \left\{ y \in \mathbb{R}^N : \| \ y - x \ \| < \epsilon \right\}$.

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

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

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 $h^*(z) = \frac{N}{2\alpha}n^*(\frac{z}{\frac{N}{2\alpha}})$.

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

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

1.9.3 Proof of Proposition 2

The proof of proposition 2 uses the following lemmas.

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

Proof : We note $Y_{\epsilon f,t} = \epsilon_t \bar{f}_t^{'}$.

First,

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

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, $\mid E[\bar{f}_t^{'}\bar{f}_t\epsilon_t^{'}\epsilon_t] \mid \leq \sqrt{E[\parallel \bar{f}_t \parallel^4]E[\parallel \epsilon_t \parallel^4]} = O(N)$. 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)) [7] (with $q = r = 2+\rho$),

7. 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

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

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 2**(iii)**, 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, $E[\parallel \frac{\epsilon'\bar{F}}{T} \parallel_F^2] = O(\frac{N}{T})$. In conclusion,

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

Lemma 3. *Suppose Assumption 2 is satisfied.* $\parallel \hat{\beta}-\beta \parallel_F^2 = O_p(\frac{N}{T})$.

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

Proof : 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 $k = 1, \cdots, K$, $||\hat{\beta}_k - \beta_k||_N = O_p(\frac{1}{\sqrt{T}})$.*

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 : We follow the proof of Lemma 3 of Carrasco (2012).

We have the following decomposition.

$$\begin{aligned}\| \hat{\Sigma}_\alpha^{-1}\hat{\beta}_k - \Sigma^{-1}\beta_k \|_N^2 &\leq 3 \| \hat{\Sigma}_\alpha^{-1}\left(\hat{\beta}_k - \beta_k\right) \|_N^2 & (1.16)\\ &+ 3 \| \left(\hat{\Sigma}_\alpha^{-1} - \Sigma_\alpha^{-1}\right)\beta_k \|_N^2 & (1.17)\\ &+ 3 \| \left(\Sigma_\alpha^{-1} - \Sigma^{-1}\right)\beta_k \|_N^2 & (1.18)\end{aligned}$$

We have $\| \Sigma_\alpha^{-1} \|_N = \underset{\|\phi\|_N \leq 1}{sup} \| \Sigma_\alpha^{-1}\phi \|_N$. So,

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

Therefore, (1.16) is $O_P(\frac{1}{\alpha T})$.

Let $\phi = \Sigma_\alpha^{-1}\beta_k$. For the second line of the decomposition, (1.17), we have

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

where

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

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

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

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

Therefore,

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

We have

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

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 (1.18) satisfies

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

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 : **Consider the case when $\{r_t\}_{t=1,\cdots,T}$ are independent.** From Assumptions 5 **(iii)**, $E\left[\| r_t \|_N^{2+\rho}\right] = O(1)$ when N goes to infinity for $\rho > 0$. To establish the central limit theorem, we need to verify the Lindeberg condition for a double indexed process of Phillips and Moon (1999)(see their Theorem 2). In our setting, this condition can be rewritten as

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

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 $\mid \frac{Y_{t,T,N}}{\sigma_{T,N}\varepsilon} \mid > 1$,

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

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

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

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

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 as the Lindeberg condition (iii) of Page 1168 still applies when N goes to infinity. 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 $\| u \|_\infty < \infty$ and $\| v \|_\infty < \infty$, as N and T go to ∞, if*

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

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

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

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

Therefore,

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

Following Lemma 6, $< v, (\mu_2 - \hat{\mu}_2) >_N \frac{1}{\sqrt{T}} \sum_{t=1}^T < \tilde{r}_t, u >_N = O_p(\frac{1}{\sqrt{T}})$, $\frac{1}{\sqrt{T}} \sum_{t=1}^T < v, \tilde{r}_t >_N < (\mu_2 - \hat{\mu}_2), u >_N = O_p(\frac{1}{\sqrt{T}})$, and $\sqrt{T} < v, (\mu_2 - \hat{\mu}_2) >_N < (\mu_2 - \hat{\mu}_2), u >_N = O_p(\frac{1}{\sqrt{T}})$. 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, the proof is similar to that of Lemma 6. To apply the central limit theorem of Francq and Zakoïan (2005), we need

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

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

Proof of Proposition 2 Consistency :
Recall that by Equation (1.3), we have

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

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

$$\begin{aligned} \hat{\theta}^{\alpha}_{HJ} - \theta_{HJ} = & (\hat{V}_{11}^{-1} - V_{11}^{-1}) \gamma \\ & + \hat{V}_{11}^{-1} (\hat{\mu}_1 - \mu_1) \\ & + \hat{V}_{11}^{-1} (\frac{\hat{\beta}' \hat{\Sigma}_{\alpha}^{-1} \hat{\beta}}{N})^{-1} \left[\frac{\hat{\beta}' \hat{\Sigma}_{\alpha}^{-1} (\beta - \hat{\beta})}{N} (\gamma + \hat{\mu}_1 - \mu_1) \right. \\ & \left. + \frac{\hat{\beta}' \hat{\Sigma}_{\alpha}^{-1} e}{N} + \frac{\hat{\beta}' \hat{\Sigma}_{\alpha}^{-1} \bar{\epsilon}}{N} \right]. \end{aligned} \tag{1.19}$$

For the first two rows, $(\hat{V}_{11}^{-1} - V_{11}^{-1}) \gamma$ et $\hat{V}_{11}^{-1} (\hat{\mu}_1 - \mu_1)$ 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

$$< \hat{\beta}_{k_1}, \hat{\Sigma}_{\alpha}^{-1}\hat{\beta}_{k_2} >_N \; = \; < \hat{\beta}_{k_1} - \beta_{k_1}, \hat{\Sigma}_{\alpha}^{-1}\hat{\beta}_{k_2} - \Sigma^{-1}\beta_{k_2} >_N + \tag{1.20}$$
$$< \hat{\beta}_{k_1} - \beta_{k_1}, \Sigma^{-1}\beta_{k_2} >_N + \tag{1.21}$$
$$< \beta_{k_1}, \hat{\Sigma}_{\alpha}^{-1}\hat{\beta}_{k_2} - \Sigma^{-1}\beta_{k_2} >_N + \tag{1.22}$$
$$< \beta_{k_1}, \Sigma^{-1}\beta_{k_2} >_N - C_{k_1,k_2} + \tag{1.23}$$
$$C_{k_1,k_2}. \tag{1.24}$$

where C was defined in Assumption 3. $|(3.12)| \leq \| \hat{\beta}_{k_1} - \beta_{k_1} \|_N \| \hat{\Sigma}_{\alpha}^{-1}\hat{\beta}_{k_1} - \Sigma^{-1}\beta_{k_1} \|_N \to 0$ as $N, T \to \infty$, and $\alpha T \to \infty$ using Lemma 5.

For (3.13), 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 (3.14).

Finally, using assumption 3, (3.15) goes to 0 as N goes to ∞.

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

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

For $\frac{\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}e}{N}$, 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 (1.2). Therefore when as $N, T \to \infty$, and $\alpha T \to \infty$, $\frac{\hat{\beta}'\hat{\Sigma}_{\alpha}^{-1}e}{N} \xrightarrow{P} 0$.

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 (1.19) using the following results :

— $\hat{\beta} - \beta = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t \bar{f}_t{}' \hat{V}_{11}^{-1}$

— Note that $\hat{C}_\beta = \frac{\hat{\beta}'\hat{\Sigma}_\alpha^{-1}\hat{\beta}}{N}$ 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 (1.2).

— 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,

$$\begin{aligned}
&\sqrt{T}\left(\hat{\theta}_{HJ}^{\alpha} - \theta_{HJ}\right) + \sqrt{T}\hat{V}_{11}^{-1}\hat{C}_\beta^{-1}\beta'\hat{\Sigma}_\alpha^{-1}(\hat{\Sigma}_\alpha - \hat{\Sigma})\Sigma^{-1}\frac{e}{N} \\
&= \hat{V}_{11}^{-1}\Bigg\{-\sqrt{T}(\hat{V}_{11} - V_{11})V_{11}^{-1}\gamma \\
&+ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(f_t - \mu_1) \\
&+ \hat{C}_\beta^{-1}\Bigg[-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\hat{\beta}'\hat{\Sigma}_\alpha^{-1}\epsilon_t\bar{f}_t'}{N}\hat{V}_{11}^{-1}(\gamma + \hat{\mu}_1 - \mu_1) \\
&- \sqrt{T}\beta'\hat{\Sigma}_\alpha^{-1}(\hat{\Sigma} - \Sigma)\Sigma^{-1}\frac{e}{N} \\
&+ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{V}_{11}^{-1}\frac{\bar{f}_t\epsilon_t'\hat{\Sigma}_\alpha^{-1}e}{N} \\
&+ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\hat{\beta}'\hat{\Sigma}_\alpha^{-1}\epsilon_t}{N}\Bigg]\Bigg\}.
\end{aligned} \tag{1.25}$$

We prove the asymptotic normality of each component of (1.25) 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

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\hat{\beta}'\hat{\Sigma}_\alpha^{-1}\epsilon_t\bar{f}_t'}{N}\hat{V}_{11}^{-1}(\gamma+\hat{\mu}_1-\mu_1) &= \frac{\hat{\beta}'\hat{\Sigma}_\alpha^{-1}}{N}[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\tilde{f}_t'V_{11}^{-1}\gamma & (1.26) \\ &+ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\bar{f}_t'(\hat{V}_{11}^{-1}-V_{11}^{-1})\gamma & (1.27) \\ &- \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t(\hat{\mu}_1-\mu_1)'V_{11}^{-1}\gamma & (1.28) \\ &+ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\bar{f}_t'\hat{V}_{11}^{-1}(\hat{\mu}_1-\mu_1)] & (1.29)\end{aligned}$$

The term (1.27) 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, (1.27) is $o_p(1)$ when $N, T \to \infty, and\, \alpha T \to \infty$.

For (1.28), using the fact that $\theta_{HJ} = V_{11}^{-1}\gamma$, we can rewrite it 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 $\parallel \epsilon'\Theta \parallel^2 = O_P(N)$, $\parallel \hat{\mu}_1 - \mu_1 \parallel^2 = O_P(\frac{1}{T})$. 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 $\parallel \Sigma^{-1}\beta \parallel_N < \infty$ 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, (1.28) is $o_p(1)$.

For (1.29), 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 $o_p(1)$ $N, T \to \infty$, and $\alpha T \to \infty$.

Finally, (1.26) 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

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^T \hat{V}_{11}^{-1}\frac{\bar{f}_t\epsilon_t'\hat{\Sigma}_\alpha^{-1}e}{N} &= V_{11}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T\frac{\tilde{f}_t\epsilon_t'\hat{\Sigma}_\alpha^{-1}e}{N} && (1.30)\\ &+ \frac{1}{N\sqrt{T}}\left(\hat{V}_{11}^{-1} - V_{11}^{-1}\right)\bar{F}'\epsilon\hat{\Sigma}_\alpha^{-1}e && (1.31)\\ &+ \frac{1}{N\sqrt{T}}V_{11}^{-1}(\mu_1 - \hat{\mu}_1)\sum_{t=1}^T\epsilon_t'\hat{\Sigma}_\alpha^{-1}e. && (1.32)\end{aligned}$$

By Lemma 6 and the Cramer Wold device, (1.30) converges to a gaussian distribution.

Using Lemmas 2 and 5, (1.31) and (1.32) are $o_p(1)$ when $N, T \to \infty$, and $\alpha T \to \infty$.

— As as N , $T \to \infty$, $\alpha \to 0$ and $\alpha T \to \infty$, normality of $\sqrt{T}\beta'\hat{\Sigma}_\alpha^{-1}(\hat{\Sigma}-\Sigma)\Sigma^{-1}\frac{e}{N}$ emanates 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 below

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

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

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

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

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

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

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 component 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

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

1.9.4 Proof of Proposition 3

We follow the proof 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 model 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 $\alpha^2 T$ goes to zero.

$T(\hat{\delta}_{1,\alpha}^2 - \hat{\delta}_{2,\alpha}^2) = T\hat{\theta}_{22}^{\alpha\prime}\hat{C}_{2,22,\alpha}^{-1}\hat{\theta}_{22}^{\alpha} = z'V(\hat{\theta}_{22}^{\alpha})^{\frac{1}{2}}\hat{C}_{2,22,\alpha}^{-1}V(\hat{\theta}_{22}^{\alpha})^{\frac{1}{2}}z$. $\hat{C}_{2,22,\alpha}^{-1}$ 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 $V(\hat{\theta}_{22}^{\alpha})^{-\frac{1}{2}} C_{2,22}^{-1} V(\hat{\theta}_{22}^{\alpha})^{-\frac{1}{2}}$.

1.9.5 Proof of Proposition 4

The proof is similar to the one of Proposition 3. First, note that

$$\delta_1^2 - \delta_2^2 = (\delta_1^2 - \delta_0^2) - (\delta_2^2 - \delta_0^2),$$

where δ_0^2 represents the population HJ distance of the model with factor f_1.

From Proposition 3, we have

$$\delta_1^2 - \delta_0^2 = -\theta_{12}' C_{1,22}^{-1} \theta_{12},$$

and

$$\delta_2^2 - \delta_0^2 = -\theta_{22}' C_{2,22}^{-1} \theta_{22}.$$

As a result,

$$\begin{aligned} \delta_1^2 - \delta_2^2 &= -\theta_{12}' C_{1,22}^{-1} \theta_{12} + \theta_{22}' C_{2,22}^{-1} \theta_{22} \\ &= \begin{bmatrix} \theta_{12}' & \theta_{22}' \end{bmatrix} \begin{bmatrix} -C_{1,22}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & C_{2,22}^{-1} \end{bmatrix} \begin{bmatrix} \theta_{12} \\ \theta_{22} \end{bmatrix}. \end{aligned}$$

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 αT go to infinity and α goes to zero.

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

The results follows from the singular value decomposition of

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

1.9.6 Proof of Proposition 5

For Ridge, $q_t^P(\nu_{1\alpha}) = 2y_t\frac{\nu_{1\alpha}'\tilde{r}_t}{N} - \frac{\nu_{1\alpha}'\tilde{r}_t\tilde{r}_t'\nu_{1\alpha}}{N^2} + 2\frac{\nu_{1\alpha}'\mu_2}{N} - \alpha\frac{\nu_{1\alpha}'\nu_{1\alpha}}{N}$.
For Tikhonov, $q_t^P(\nu_{1\alpha}) = 2y_t\frac{\nu_{1\alpha}'\tilde{r}_t}{N} - \frac{\nu_{1\alpha}'\tilde{r}_t\tilde{r}_t'\nu_{1\alpha}}{N^2} + 2\frac{\nu_{1\alpha}'\mu_2}{N} - \alpha\frac{\nu_{1\alpha}'\Sigma^{-1}\nu_{1\alpha}}{N}$.
We use the following decomposition of $\sqrt{T}\left(\hat{\delta}_\alpha^2 - \delta^2\right)$:

$$\begin{aligned}
\sqrt{T}\left(\hat{\delta}_\alpha^2 - \delta^2\right) &= \sqrt{T}\left(\hat{\delta}_\alpha^2 - \delta_\alpha^2\right) + \sqrt{T}\left(\delta_\alpha^2 - \delta^2\right)\\
&= \sqrt{T}\left(\hat{E}\left[q_t^P(\hat{\nu}_{1\alpha})\right] - \hat{E}\left[q_t^P(\nu_{1\alpha})\right]\right) \qquad (1.33)\\
&+ \sqrt{T}\left(\hat{E}\left[q_t^P(\nu_{1\alpha})\right] - E\left[q_t^P(\nu_{1\alpha})\right]\right) \qquad (1.34)\\
&+ \sqrt{T}\left(\delta_\alpha^2 - \delta^2\right) \qquad (1.35)
\end{aligned}$$

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

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

As $\nabla E\left[q_t^P(\nu_{1\alpha})\right] = 0$, 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,

$$\begin{aligned}
& \sqrt{T}\nabla\hat{E}\left[q_t^P(\nu_{1\alpha}^*)\right](\hat{\nu}_{1\alpha}-\nu_{1\alpha}) \\
= \; & \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left[y_t\frac{\tilde{r}_t}{N}-E[y_t\frac{\tilde{r}_t}{N}']\right]'(\hat{\nu}_{1\alpha}-\nu_{1\alpha}) && (1.36)\\
& -\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left[\nu_{1\alpha}'\frac{\tilde{r}_t\tilde{r}_t'}{N^2}-E[\nu_{1\alpha}'\frac{\tilde{r}_t\tilde{r}_t'}{N^2}]\right](\hat{\nu}_{1\alpha}-\nu_{1\alpha}). && (1.37)
\end{aligned}$$

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, equation (1.36) is equal to

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

while Equation (1.37) 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 (1.33)$= o_p(1)$.

Equation (1.34) can be rewritten as below

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

As $N, T\to\infty$ and $\alpha\to 0$, $<\sqrt{T}(\hat{\Sigma}-\Sigma)(\nu_{1\alpha}-\nu_1),\nu_{1\alpha}-\nu_1>_N\leq\|\sqrt{T}(\hat{\Sigma}-\Sigma)\|_N\|\nu_{1\alpha}-\nu_1\|_N^2\to 0$ and $<\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_1,\nu_{1\alpha}-\nu_1>_N\leq\|\sqrt{T}(\hat{\Sigma}-\Sigma)\nu_1\|_N\|\nu_{1\alpha}-\nu_1\|_N\to 0$. So (1.34) 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 (1.34) converges to a normal distribution with variance

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

Finally, we have

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

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

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

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

1.9.7 Proof of Proposition 6

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

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

TABLE 1.8 – 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

Chapitre 2

Comparison of Misspecified Asset Pricing Models Using Many Characteristic-sorted and Micro Portfolios *

2.1 Introduction

In the asset pricing literature, financial economists have proposed a combination of more than 400 risk factors, often termed anomalies, to explain the difference in expected returns in the financial market. These factors primarily relate to a company's profitability, investment strategies, value, and market trading frictions such as liquidity and volatility, among others. In contrast to these proposed anomaly-driven models, macroeconomic or theoretical models represent a class using a general equilibrium framework to understand holistically the expected return differences in the final markets. Even though testing whether a particular factor carries a non-null risk premium is essential, it is equally important to compare the performance of the different asset pricing models proposed in the literature.

Past comparisons of these asset pricing models reveal a similar performance, i.e., they equally explain the difference in expected returns among assets. Part of the results emanates from the low number of test assets or portfolios to be explained by the models.

*. I thank Marine Carrasco for her invaluable guidance and the participants of the 2023 Dagenais seminar.

Researchers typically use the 25 size- and BM-sorted portfolios of Fama and French (1993), sometimes supplemented by the 49 industry portfolios from the suggestion of Lewellen et al. (2010). In recent studies, the number of test assets used to assess the models has expanded. For instance, He et al. (2017) utilized 124 portfolios covering a range of asset classes (such as US bonds, sovereign bonds, options, credit default swaps, commodities, and foreign exchange) to evaluate their model.

Another difficulty stems from the portfolios' construction method as pointed out by Barras (2019). Researchers do not use individual stocks to compare models. They regroup them into portfolios using some of their characteristics, e.g., size, book-to-market, and industry. Blume (1970) pioneered this approach. It alleviates the error-in-variables issue when the beta between test assets and factors is used as regressors to estimate asset pricing models. These so-called characteristics-sorted portfolios contains a considerable number of stocks and might weaken idiosyncratic information. In addition, as argued by Barras (2019) they lead to correlated beta across risk factors and a lack of discrimination power of the test assets despite the large spread of expected return they produce.

This paper estimates and compares several asset pricing models with more than 3000 test assets that combine the well-known characteristic-sorted portfolios with micro portfolios Barras (2019) suggested. These micro portfolios are also formed using observed financial characteristics but contain 5 to 10 stocks. Consequently, they are analogous to individual stocks, offer significant return spread, and might improve the lack of discriminatory power of the characteristics-sorted portfolios.

Estimating asset pricing models relies on the well-known Hansen and Jagannathan (1997, HJ hereafter) distance. It is a misspecification measure that indicates the distance between the stochastic discount factor (SDF) based on the risk factors of the proposed models and the true SDF that prices exactly the test assets. When a risk-free asset is available, Kan and Robotti (2008) show that it corresponds to a GMM distance with the inverse covariance matrix of excess returns as the weighting matrix. Interestingly, in this case, the distance is also the difference between two squared Sharpe Ratios : the squared Sharpe Ratio of the test assets and the squared Sharpe Ratio of the mimicking portfolio of the factors proposed by the model. There is also a closed connection to the cross-sectional R^2, as pointed out by Kan, Robotti, and Shanken (2013). Using the distance with 3000 test assets is empirically impossible as the covariance matrix is not invertible.

Carrasco and Nokho (2022) introduce a new version of the distance where a regularized version replaces the covariance matrix. This extension is equivalent to allowing for the presence of pricing errors for the true SDF and permits to account for measurement errors or market frictions as discussed in Korsaye, Quaini, and Trojani (2019). The distance can also be used to estimate SDF parameters.

Using the regularized HJ distance to compare models is identical to comparing their squared Sharpe ratios. Barillas and Shanken (2017) address such comparisons with models made of traded factors. They show that test assets are irrelevant in this case, and the extent to which each model can price the factors in the other model is the most important. This observation does not apply here as we also compare models without traded factors, i.e., macroeconomic models. In addition, we do not impose the traded factor restriction in the models as their factor-mimicking portfolios will replicate them very well.

Our statistical comparison uses the pairwise framework of Gospodinov et al. (2013) and depends on the relation between the models. When the factors of a model are entirely in another model (nested case), the sample HJ of the smaller model will always be higher. The comparison, then, determines if the inequality is strict or equality holds. Most of the time, models have only a set of common factors. In this case, they are non-nested, and the comparison determines whether the non-common factors have non-null SDF or whether the sample difference of their Sharpe ratios is significant. Finally, our method directly looks at the HJ differences when the models do not have common factors. We also use the multiple comparison method, where more than two models are simultaneously compared.

Our main contributions can be summarized as follows. We estimate and compare ten unconditional and conditional asset pricing models using the regularized distance and a combination of 3198 portfolios ranging from July 1973 to June 2018 ($T = 540$). Four are macroeconomic or theoretical models : the Consumption Capital Asset Pricing Model (CCAPM), the Durable Consumption Capital Asset Pricing Model (DCAPM) of Yogo (2006), the Human capital Capital Asset Pricing Model (HCAPM) of Jagannathan and Wang (1996), and the Intermediary Asset pricing model (IAPM) of He, Kelly, and Manela (2017). Five are anomaly-driven : the three (FF3) and Five-factor (FF5) Models of Fama and French, 1993 and 2015, the Carhart (1997) model, which adds momentum to FF3, the Liquidity Model of Pástor and Stambaugh (2003), and the Augmented q-Factor Model (q5) of Hou et al. (2021). We also estimate the Consumption model of Lettau and Ludvigson (2001) using quarterly data. We did not include the Consumption model of Lettau and Ludvigson (2001) in the comparisons as the lower amount of data reduces the tests' power. In contrast to unconditional models, conditional models consider the economic business cycles and fluctuations of financial market conditions. We use the macroeconomic and financial uncertainty indices of Ludvigson, Ma, and Ng (2021) to build conditional models. We find that these two variables significantly reduce the models' pricing errors. Comparisons of unconditional models reveal that macroeconomic models display similar performance in explaining the cross-section of returns. However, they have lower explanatory power than the anomaly-driven models. FF3 is the sole exception. Furthermore, augmenting the FF3 with momentum and liquidity factors enhances the model's explanatory capability. Nevertheless, this augmented model is inferior to the FF5 and q5. For conditional models, the DCAPM and HCAPM improve and dominate CCAPM and IAPM. These models also have similar pricing errors as the conditional Carhart and liquidity models but are still inferior to FF5 and q5. Overall, FF5 and q5 are indistinguishable in the unconditional and conditional setting and dominate all the other models.

This work is mainly related to the literature on the evaluation and comparison of asset pricing models. Notably, it is closest to the paper of Barillas et al. (2020). They use the same distance to compare models with fewer test assets. In addition, they emphasize on the anomaly-driven models and incorporate only the model of He et al. (2017) as a macroeconomic model, which we also include in this paper. In addition, they did not consider conditional models. This paper is also close to Kan and Robotti (2009). They develop a pairwise comparison framework in the context of gross returns and compare several models. They find the data too noisy to uncover differences in conditional and unconditional models. Finally, our work is related to the paper of Barras (2019). He compares asset pricing models using a new metric different from the HJ distance. This metric computes the proportion of mispriced micro portfolios for each asset pricing model.

The remainder of the paper is organized as follows : Section 2.2 introduces the regularized Hansen Jagannathan distance, our model comparison method, and the estimation of conditional models. Section 2.3 describes the construction of the micro-portfolios, the Characteristic-sorted portfolios used, and the conditioning variables we employ for the conditional models. Section 2.4 presents in detail the competing models. Finally, Section 2.5 exposes the estimation and comparison results of the models.

2.2 Asset pricing models with many assets and conditional models

In this section, we elucidate the strategy employed to estimate and compare asset pricing models in a large financial market. These models primarily aim to explain the difference in expected returns observed among companies. We extend the prominent approach of Hansen and Jagannathan (1997) to take into account many number of companies.

2.2.1 Dealing with many test assets

For a vector of N random excess returns (companies) r_t [2], if the law of one price holds, there exists a square-integrable random variable m_t that satisfies the fundamental Equation below

$$E\left[r_t m_t\right] = 0_N. \tag{2.1}$$

The random variable m_t is the marginal rate of substitution or the stochastic discount factor (SDF). It generalizes the idea of discounting. Linear asset pricing models posit a particular empirical SDF, $y_t(\theta, f_t) = \theta' f_t$, function of a vector of K factors, f_t, and unknown parameters $\theta \in \Theta$. In the context of excess returns, the proposed SDF should be specified such that $\theta = 0$ is not a solution to the equation 2.1. In this paper, we fix the mean of y as in Kan and Robotti (2008) [3]and consider SDF of the form

$$y_t(\theta, f_t) = 1 - \theta'(f_t - E[f_t]).$$

For a small number of test assets N, Kan and Robotti (2008) propose to evaluate models using the following distance

$$\delta^2 = \underset{m \in \mathcal{M}, E[m]=c}{inf} \; E\left[(m_t - y_t)^2\right] \quad s.t \quad E\left[m_t r_t\right] = 0_N.$$

This is a modified version of the Hansen and Jagannathan (1997) distance as the mean of the true SDF has a fixed value. This measure of misspecification leads to the following

2. Excess return equals the return earned by a company minus the risk free rate (r_f).

3. Fixing the mean of the empirical SDF y_t is similar to requiring the model to correctly price the risk free return r_f. Particularly, the risk free return verifies $E\left[r_f y_t\right] = 1$. So, $E[y_t] = \frac{1}{r_f} = c$.

GMM distance

$$\delta^2 = e(\theta)' V_r^{-1} e(\theta),$$

where $e(\theta) = E[r_t y_t(\theta)]$ and $V_r = cov(r_t)$. Suppose we have T observations on r_t and f_t at our disposal. The SDF parameter θ can be estimated by minimizing the empirical counterpart of δ^2. However, when $N > T$, i.e., the number of test assets is large, the empirical inverse of V_r does not exist and δ^2 cannot be used.

When N is large, Carrasco and Nokho (2022) propose to evaluate the models by measuring how far the proposed y is to the closest valid SDF that prices test portfolios with errors up to τ as below

$$\delta_\tau^2 = \underset{m \in \mathcal{M}, E[m]=c}{inf} E\left[(m_t - y_t)^2\right] \quad s.t \parallel E[m_t r_t] \parallel_{N,\Sigma}^2 \leq \tau. \tag{2.2}$$

In the previous problem, $\Sigma = \frac{V_r}{N}$ and$\parallel x \parallel_{N,\Sigma}^2 = \frac{x' \Sigma x}{N}$ for any $x \in \mathbb{R}^N$. As a result, the constraint $\parallel E[m_t r_t] \parallel_{N,\Sigma}^2 \leq \tau$ imposes only a bound on the Fundamental Equation of asset pricing instead of its exact nullity. Allowing for the presence of pricing errors for the true SDF permits to account for measurement errors or market frictions as discussed in Korsaye, Quaini, and Trojani (2019). This is quite plausible when we use a large number of test assets. We did not impose the no arbitrage constraint in the evaluation of the models ($m_t > 0$) as problem (2.2) yields closed-form analytical expressions. So, the models are just approximations and are assumed to be misspecified.

A similar and more tractable manner to see problem (2.2) is to rewrite it as a penalization

$$\delta_\alpha^2 = \underset{m \in \mathcal{M}, E[m]=1}{inf} E\left[(m_t - y_t)^2\right] + \frac{1}{\alpha} \parallel E[m_t r_t] \parallel_{N,\Sigma}^2,$$

where α controls the level of pricing errors allowed to the true SDF. This leads to the following regularized HJ distance

$$\delta_\alpha^2 = \frac{e(\theta)' \Sigma_\alpha^{-1} e(\theta)'}{N},$$

where $\Sigma_\alpha^{-1} = \left(\Sigma^2 + \alpha I_N\right)^{-1} \Sigma$. Using the linear SDF $y_t(\theta, f_t) = 1 - \theta' \tilde{f}_t$ where $\tilde{f}_t = f_t - E[f_t]$, we have

$$\delta_\alpha^2 = \frac{1}{N} (\mu_r - V_{rf}\theta)' \Sigma_\alpha^{-1} (\mu_r - V_{rf}\theta), \tag{2.3}$$

where $\mu_r = E[r_t]$, $V_{rf} = E[r_t \tilde{f}_t']$.

Carrasco and Nokho (2022) study the estimation of θ, the minimizer of δ^2, using the empirical counterpart of (2.3) when N and T goes to infinity and α goes to zero. They also take into account the fact that $e \neq 0$ for most models, i.e., they are globally misspecified. The parameter θ is estimated by

$$\hat{\theta}_\alpha = (\hat{V}_{rf}^{'}\hat{\Sigma}_\alpha^{-1}\hat{V}_{rf})^{-1}\hat{V}_{rf}^{'}\hat{\Sigma}_\alpha^{-1}\hat{\mu}_r,$$

where $\hat{V}_{rf} = \frac{1}{T}\sum_{t=1}^{T} r_t\bar{f}_t^{'}$, $\tilde{f}_t = f_t - \frac{1}{T}\sum_{t=1}^{T} f_t$ $\mu_r = \frac{1}{T}\sum_{t=1}^{T} r_t$, and $\hat{\Sigma}_\alpha^{-1} = \left(\hat{\Sigma}^2 + \alpha I_N\right)^{-1}\hat{\Sigma}$ with $\hat{\Sigma}$ the estimated covariance matrix of r_t. They show that when $N, T, \alpha T \to \infty$, and $\alpha, \alpha^2 T \to 0$

$$\sqrt{T}(\hat{\theta}_\alpha - \theta) \longrightarrow \mathcal{N}(0_K, V_f^{-1}\Omega V_f^{-1}), \tag{2.4}$$

where $\Omega = \underset{N,T\to\infty}{lim}\, var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_t\right]$, $h_t = \widetilde{f}_t y_t + \gamma + C_\beta^{-1}\frac{\beta^{'}\Sigma^{-1}}{N}(\epsilon_t y_t - \tilde{r}_t\tilde{u}_t + e) + C_\beta^{-1}V_f^{-1}\widetilde{f}_t\frac{\epsilon_t^{'}\Sigma^{-1}e}{N}$, $\tilde{r}_t = r_t - E[r_t]$, $V_f = cov(f_t)$, $\tilde{u}_t = \frac{\tilde{r}_t\Sigma^{-1}e}{N}$, ϵ_t is the residuals from projecting r_t on f_t and a constant, $\beta = V_{rf}V_f^{-1}$, $C_\beta = \frac{1}{N}\beta'\Sigma^{-1}\beta$, and $\gamma = V_f^{-1}\theta$ is the risk premium. Also, we can express the distance as

$$\delta_\alpha^2 = \mu_r^{'}\Sigma_\alpha^{-1}\mu_r - \mu_r^{'}\Sigma_\alpha^{-1}\beta(\beta^{'}\Sigma_\alpha^{-1}\beta)^{-1}\beta^{'}\Sigma_\alpha^{-1}\mu_r. \tag{2.5}$$

$\hat{\delta}_\alpha^2$ has an interesting economic interpretation. In particular, the regularized HJ coincides with the difference between the squared Sharpe Ratio of the tangency portfolio formed by test assets and the mimicking portfolio of the K factors. The mimicking portfolio of the K factors is an investment position designed to closely follow f_t using only r_t. The returns of this portfolio is given by $f_t^* = Ar_t,\ t = 1, \ldots, T$, where A is a $K \times N$ matrix that represents the weights. Its unconditional mean is $\mu^* = E[f_t^*] = A\mu_r$ and its variance is $V^* = Var(f_t^*) = A^{'}V_r^{-1}A$. We estimate the mimicking portfolio weights, A, via the following regression

$$f_t = a + Ar_t + u_t,\ t = 1, \ldots, T$$

where A is a $K \times N$ matrix and a is a vector of size K. Because we have a large number of test assets (explanatory variables) in the regression, we employ the following penalized minimization for each factor k

$$\underset{A_k}{min}\, E\left[(\tilde{f}_{kt} - A_k^{'}\tilde{r}_t)^2\right] + \alpha A_k^{'}\Sigma^{-1}A_k,\ k = 1, \ldots, K. \tag{2.6}$$

In the previous expression, we consider $A = [A_1, \cdots A_K]'$, where $A_k,\ k = 1, \ldots K$, are vectors of dimension N. So $A = \Sigma_\alpha^{-1} V_{rf}$ and its estimation is $\hat{A}_\alpha = \hat{\Sigma}_\alpha^{-1} \hat{V}_{rf}$. In addition, the estimated mean and variance of the mimicking portfolio are $\hat{\mu}_\alpha^* = \hat{V}_{fr} \hat{\Sigma}_\alpha^{-1} \hat{\mu}_r$ and $\hat{V}_\alpha^* = \hat{V}_{fr} \hat{\Sigma}_\alpha^{-1} \hat{V}_{rf}$ if we estimate Σ by $\hat{\Sigma}_\alpha$. Therefore $\hat{\theta}_\alpha = \hat{V}_\alpha^{*-1} \hat{\mu}_\alpha^*$, which is the normalized mean of the mimicking portfolio. Furthermore, the squared Sharpe Ratio of the tangency portfolio constructed from the K factor mimicking portfolios is given by

$$\begin{aligned} \hat{\mu}_\alpha^{*\prime} \hat{V}_\alpha^{*-1} \hat{\mu}_\alpha^* &= \left(\hat{\mu}_r \hat{\Sigma}_\alpha^{-1} \hat{V}_{rf}\right) \left(\hat{V}_{rf}' \hat{\Sigma}_\alpha^{-1} \hat{V}_{rf}\right)^{-1} \left(\hat{V}_{rf}' \hat{\Sigma}_\alpha^{-1} \hat{\mu}_r\right) \\ &\equiv \left(\hat{\mu}_r' \hat{\Sigma}_\alpha^{-1} \hat{\beta}\right) \left(\hat{\beta}' \hat{\Sigma}_\alpha^{-1} \hat{\beta}\right)^{-1} \left(\hat{\beta}' \hat{\Sigma}_\alpha^{-1} \hat{\mu}_r\right). \end{aligned}$$

The estimated squared Sharpe Ratio of the test asset is $\hat{\mu}_r' \hat{\Sigma}_\alpha^{-1} \hat{\mu}_r$. So, the HJ distance measures how close the squared Sharpe Ratio of the factor mimicking portfolio is to the squared Sharpe Ratio of the test assets.

We choose α by maximizing the out-of-sample R^2. We divide the sample into two parts, and then choose the best α between 0.0001 and 0.2 that maximizes

$$R^2_{oos,\,\alpha} = 1 - \frac{(\hat{\mu}_r^o - \hat{V}_{rf}^o \hat{\theta}_\alpha)'(\hat{\mu}_r^o - V_{rf}^o \hat{\theta}_\alpha)}{\hat{\mu}_r^{o\prime} \hat{\mu}_r^o}.$$

In the previous equation, $\hat{\theta}_\alpha$ is estimated from the first part of the sample and quantities with $.^o$ are estimated from the second part sample.

Aside from the estimation of the SDF parameter, we compare the ability of the models to correctly explain the difference in expected returns of the test assets. The regularized HJ distance offers a very good framework for that. Let $\mathcal{F}_1 = \{y_{1t}(\theta_1, .); \theta_1 \in \Theta_1\}$ and $\mathcal{F}_2 = \{y_{2t}(\theta_2, .); \theta_2 \in \Theta_2\}$ be the proposed linear SDF families of two models. In addition, δ_1^2 and δ_2^2 are the aggregate pricing errors of the models. The comparison of the two models investigates whether $\delta_1^2 = \delta_2^2$. From (2.5), it is noteworthy to see that this exercise amounts to comparing the Sharpe ratio of the mimicking portfolios of the two models. This test can be conducted using the fact when $N, T, \alpha T \to \infty$, and $\alpha, \alpha^2 T \to 0$ and $\delta_1, \delta_2 \neq 0$,

$$\sqrt{T}\left(\hat{\delta}_{1\alpha}^2 - \hat{\delta}_{2\alpha}^2 - (\delta_1^2 - \delta_2^2)\right) \xrightarrow{d} \mathcal{N}(0, v), \tag{2.7}$$

where $v = \underset{N,T\to\infty}{lim} var\left[\frac{1}{\sqrt{T}} \sum_{t=1}^T l_{1t} - l_{2t}\right]$. $l_{mt} = 2 y_{mt} \tilde{u}_{mt} - \tilde{u}_{mt}^2 - \delta_m^2 + 2\frac{e_m' \Sigma^{-1} \mu_R}{N}$,

$u_{mt} = \frac{r_t' \Sigma^{-1} e_m}{N}$, and $m = 1, 2$. When models are misspecified, the equality $\delta_1^2 = \delta_2^2$ can happen under two circumstances. The first is when $y_1 = y_2$. In this case, the previous distribution degenerates to 0. The second is when $y_1 \neq y_2$, and the aggregate errors are the same $\delta_1^2 = \delta_2^2$. This observation leads to a sequential test to compare the performance of two models. If the models are nested ($\mathcal{F}_1 \subset \mathcal{F}_2$ or $\mathcal{F}_2 \subset \mathcal{F}_1$), or overlapping ($\mathcal{F}_1 \not\subset \mathcal{F}_2$ or $\mathcal{F}_2 \not\subset \mathcal{F}_1$, $\mathcal{F}_1 \cap \mathcal{F}_2 \neq \emptyset$), we first test whether $y_1 = y_2$ through a Wald test of the nullity of the uncommon factors between the two models using distribution (2.4). If $y_1 = y_2$ is rejected, we compare the pricing errors of the two models using (2.7). However if $y_1 = y_2$ cannot be rejected, we conclude the equal performance of the two models. When the two models are strictly non-nested ($\mathcal{F}_1 \cap \mathcal{F}_2 \neq \emptyset$), we directly use (2.7).

When considering the comparison of multiple models, the significance level of a series of pairwise tests is not known. Instead, we can use a multiple-model comparison test of Wolak (1989). The test determines whether a benchmark model with δ_1^2 performs at least as well as the $q = p - 1$ others with squared HJ given by $\delta_i^2, i = 2, \ldots, p$. So, we test $H_0 : d = (d_2, \ldots, d_p) \leq 0_q$ against $H_1 : d \in \mathbb{R}^q$, where $d_i = \delta_1^2 - \delta_i^2$. The differences of HJ are normally distributed according to (2.7). We note Ω_d the covariance matrix of the vector d. The likelihood-ratio type statistics of the null hypothesis is

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

where $\hat{\Omega}_{d,\alpha}$ is a consistent estimator of Ω_d and $\tilde{d}_\alpha = \underset{d}{argmin} \ (\hat{d}_\alpha - d)' \hat{\Omega}_{d,\alpha}^{-1} (\hat{d}_\alpha - d) \quad s.t. \ d \leq 0_p$. As $N, T \to \infty$ and $\alpha \to 0$, this statistics converges to $\sum_{i=0}^{q} w_{q-i}(\Omega_d) \chi^2(i)$, where $\chi^2(i)$ are independent χ^2 random variables with i degrees of freedom, $\chi^2(0)$ is zero, and the weights w_i sum up to one[4]. This distribution can be used to obtain valid p-values.

2.2.2 Conditional models

We describe the estimation of conditional asset pricing models. The previous approach has a major drawback. Estimated parameters do not vary with time. It forces prices of risk to be invariant to business cycles. This might limit the ability of the models to price correctly the assets. So, we also estimate and compare conditional asset pricing models.

4. Appendix C of Gospodinov et al. (2013) gives the procedure to compute $w_i(\Omega_d)$ and the p-value of the test

We take into account business cycles by explicitly modeling the SDF parameters as linear function of J macroeconomic variables represented by the column vector x_t

$$y_t = 1 - \theta(x_{t-1})' \left(f_t - E(f_t)\right),$$

and

$$\theta_k(x_{t-1}) = \theta_{0k} + \theta_{1k}x_{1t-1} + \cdots + \theta_{jk}x_{Jt-1} \; for \; k = 1, \cdots, K.$$

This yields the following SDF $y_t = 1 - \theta_0'\tilde{f}_t - \theta_{-0}'(\tilde{f}_t \otimes x_{t-1})$, where $\otimes$ is the Kronecker product, θ_0 is a K-dimensional column vector associated to the demeaned factors, θ_{-0} is a KJ-dimensional column vector associated to the interaction variables, and $\tilde{f}_t = f_t - E(f_t)$. We demeaned the interaction terms to keep the mean of the SDF to 1. So empirically, we use $y_t = 1 - \theta_0'\tilde{f}_t - \theta_{-0}'(\tilde{f}_t \otimes x_{t-1} - E(\tilde{f}_t \otimes x_{t-1}))$. Therefore, estimating this type of conditional model is the same as estimating an unconditional model with scaled factors.

It is noteworthy to see that the conditioning variables do not directly influence the SDF as factors. It would have been the case if the model had a constant SDF parameter, as is when one uses gross returns to estimate asset pricing models. The conditioning variables differ from the SDF factors in that they can encompass any variable that tracks business cycles or specific financial events. Conversely, the SDF factors typically arise from portfolio analyses for the case of anomaly-driven models or a theoretical framework for the macroeconomic or theoretical models. Our modeling approach emphasizes and clarifies this distinction.

We also incorporate conditioning informations by adding scaled returns. This approach uses the conditional version of the fundamental Equation of asset pricing which is

$$E\left[m_t r_t \mid I_{t-1}\right] = 0_N.$$

where I_{t-1} is the information set available at time $t-1$. We can multiply both side of the previous equation by $x_{t-1} \in I_{t-1}$ and take the unconditional expectation to have $E\left[m_t r_t x_{t-1}\right] = 0_N$. $r_t x_{t-1}$ are now the payoffs to managed portfolios. Characteristic-sorted portfolios, i.e., portfolio sorted using financial indicator such as size, book-to-market, industry, beta, are all managed portfolios. Therefore, we use the Characteristic-sorted portfolios to add conditioning informations to the main test assets.

2.3 Test Assets and conditioning variables

We describe the test assets r_t used to evaluate the models. It is usual to regroup individual returns into portfolios. This reduces the volatility of the returns and the estimation errors of the SDF parameters. The mainstream approach consists in sorting individual returns by observed financial characteristics such as size and averaging the returns. This yields the characteristics-sorted portfolios. They can also be considered as managed portfolios because their composition changes per year. However, the number of individual returns per portfolio is considerable, which might dilute crucial information about returns in the financial market. We supplement them with portfolios of small number of individual returns called micro portfolios.

2.3.1 Micro portfolios formation

We form several thousands micro portfolios using the approach of Barras (2019). The latter is different from the mainstream approach of portfolio sorting and flexible enough to accommodate several characteristics at the same time. Furthermore, it strikes a balance between analyzing returns from individual companies and those from portfolio sorting and is based on the local averaging methods of Efron (2010, chap. 9).

We link the Computstat and the CRSP dataset using the cusip of the companies. We work with non-financial firms. We match all the financial informations available at the fiscal yearend $Ye - 1$ with returns from July Ye to June $Ye + 1$. This ensures that these variables are known to the returns they are trying to explain (see Fama and French (1992)). For each Year Ye, we divide the companies into three size groups using the 20th and 50th percentiles of their NYSE market capitalization. The first group represents the tiny companies, the second and third group include the small and big companies.

We repeat the following process for each size group. We estimate the expected return of each company i at time t as follows $\hat{\mu}_{it} = \hat{\rho}_t c_{it}$. In the previous equation, c_{it} is a vector of observed characteristics including a constant and $\hat{\rho}_t$ is the vector of coefficients from the Fama and MacBeth (1973) cross-sectional regression of the returns prior to t on the characteristics

$$r_{is} = \rho_s^{'} c_{is} + u_{is},$$

$s < t$[5]. Similar to Barras (2019), we use book-to-market ratio, investment, and the

5. The Fama and MacBeth cross-sectional regression is a two-step process. First, we run a cross-

profitability of the firms as characteristics to track the expected returns. For each formation year t_a, we rank the companies using their standardized estimated expected returns $\hat{\mu}^s_{it_a}$ – the monthly expected returns are cross-sectionally demeaned and scaled. Then, we take the nine closest stocks in terms of expected return and build an equal-weighted micro portfolio for each stock i. Finally, we link micro portfolio returns over time. Specifically, for each pair (i, j) of micro portfolios in years t_a and $t_a + 1$, we compute $\mid \hat{\mu}^s_{it_a} - \hat{\mu}^s_{jt_a+1} \mid$ and link the portfolios with the lowest value.

The portfolios obtained through this process present missing values for months we do not have returns. We remove portfolios with less than 60 returns and use a matrix completion method to obtain a balanced micro portfolio data. The matrix completion approach relies on an assumption that the full matrix can be written as a low-rank version of a noisy matrix. The low-rank assumption is reasonable in asset pricing models, where a factor model is usually assumed for the return matrix. The method has been used in Giglio et al. (2021).

Specifically, the goal is to recover the $N \times T$ low-rank matrix R of the micro portfolios, where N is the number of micro portfolios and T is the number of time points. Suppose that Y, a $N \times T$ matrix, is a "noisy version" of R and can be written as $Y = R + E$, and E is the noise. In addition, Y is not fully observed. We introduce X as the $N \times T$ matrix whose (i, t) element is 1 if the Y_{it} is observed and 0 otherwise. So $Y^\star = Y \circ X$ is the observed noisy data, where $\circ$ represents the element-wise matrix product. We employ the following nuclear-norm penalized regression approach to recover the full micro portfolio data R :

$$\hat{R} = \underset{R}{argmin} \parallel Y^\star - R \circ X \parallel^2 + \nu \parallel R \parallel_n, \tag{2.8}$$

where $\parallel R \parallel_n$ is the matrix nuclear norm of R and $\nu > 0$ is a tuning parameter.

The solution to problem (2.8) is given by

$$\hat{R} = S_{\tau\nu/2}(\hat{R}_l - \tau X \circ (\Omega \circ \hat{R}_l - Y^\star)) \tag{2.9}$$

for $\tau > 0$. For a matrix Y, $S_a(Y) = UD_aV'$, where U,V are the left and right matrix of singular value decomposition of Y, and D_a is a modified version of the diagonal singular value matrix. D_a replaces the singular values, D_{ii} by $max(D_{ii} - a, 0)$. So D_a applies a

sectional regression for each $s < t$ to obtain $\hat{\rho}_s$. This lead to a time series of $\hat{\rho}$. Then, we compute $\hat{\rho}_t = \frac{1}{t-1} \sum_{s=1}^{s=t-1} \hat{\rho}_s$.

soft-thresholding on the singular value of Y.

Solution (2.9) suggests an iterative approach to retrieve the final matrix $\hat{R}$ as follows

$$\hat{R}_{l+1} = S_{\tau\nu/2}(\hat{R}_l - \tau X \circ (\Omega \circ \hat{R}_l - Y^\star)), \tag{2.10}$$

where $\hat{R}_l$ is the solution of the l-th iteration. The parameter τcontrols the step size of each iteration. We choose $\tau = 0.9$. We take $\hat{R}_0 = Y^\star$ as the starting value.

The tuning parameter ν controls the rank of the recovered matrix. A high value of ν reduces the rank of the recovered matrix, while a low value does the opposite. As a result, it is important to choose it carefully. We choose it by relying on the method of Chernozhukov et al. (2019), where

$$\nu = 2(1 + c)\bar{Q}(\| X \circ Z \|_n; 1 - \delta_{NT}). \tag{2.11}$$

$\bar{Q}(W; m)$ is the mth quantile of a random variable W, Z is an $N \times T$ matrix whose elements z_{it} are generated as $\mathcal{N}(0, \sigma_{ei}^2)$ independent across (i, t), σ_{ei}^2 is the variance of E_{it}, and $\delta_{NT} = 0.05$.

We employ the empirical variance of $Y_{it}^\star$ as the initial estimate of σ_{ei}^2 to compute the first value of the tuning parameter. The initial value is used to compute the solution $\hat{R}_{init}$ of (2.9) using the iterative equation (2.10). Then, we recompute a new value ν using the recovered solution. For this value of ν, $E = Y^\star - \hat{R}_{init} \circ X$. We repeat this process until the difference between the old and new solution is small enough. We present the major steps of the matrix completion algorithm below.

Algorithm 2.1 Matrix completion

```
Compute initial ν using (2.11)
Compute initial solution R̂_init using (2.10)
R̂ = R̂_init
while (diff < crit) do
    R̂_old = R̂
    Compute new λ using (2.11)
    Compute R̂ using (2.10)
    diff = ‖R̂ − R̂_old‖
end while
```

Using this approach, we end up with 2465 micro portfolios (1596 tiny, 441 small, and 428 big portfolios) with returns ranging from July 1973 to June 2018 . Panel A of Table 2.1 presents the cross-sectional quantiles of the return moments and the observed characteristics for the micro portfolios. The average return and volatility are annualized. The median of the average returns and volatilities are 11.06% and 19.44%. Returns are mostly leptokurtic. Tiny-cap micro portfolios have slightly higher return and volatility while the big-cap portfolios present lower return and volatility.

2.3.2 Characteristic-sorted Portfolios

We also use test assets issued from portfolio sorts – characteristic-sorted portfolios. As stated in Subsection 2.2.2, they are managed portfolios and can be used to introduce conditioning informations. The 25 portfolios sorted independently on size and book-to-market are commonly used to test factors or evaluate asset pricing models for they offer a sufficient range of returns to be explained. To increase the number of returns to be explained, we also include all the available portfolios from the Kenneth French website. We suppressed 5 portfolios with missing data. This amounts to 733 portfolios. The complete list of portfolios used is given in Table 2.10. Panel B of Table 2.1 shows the cross-sectional quantiles of the return moments for the combined characteristics-sorted portfolios.

2.3.3 Conditioning variables

We describe the variables used to scale the factors in the conditional asset pricing models. We draw upon the economic uncertainty literature, which introduces a range of new indicators intricately linked to the real business cycles. As pointed out by Bloom (2014), uncertainty is a concept related to the agents' uncertainty about future possible states of the economy or the uncertainty over the path of a particular macro variable like the GDP and micro variable such as the growth rate of firms. Usually, these indicators rise during recessions and drop during economic booms. We employ a pair of indicators as conditioning variables.

We use the macroeconomic uncertainty index of Jurado, Ludvigson, and Ng (2015) to account for the real business cycles into the asset pricing models. The uncertainty of a series is defined as the conditional volatility of the h-period ahead of its forecast errors. So, this definition of uncertainty quantifies the level of unpredictability of the economy's future states. Their macroeconomic uncertainty index is the equal-weighted aggregate of 132 macroeconomic activity measures.

Panel A Figure 2.1 plots the macro uncertainty index. Researchers have typically used the cyclical part of the natural logarithm of the US Industrial Production Index as a conditioning variable. Hodrick and Zhang (2001) find significant predictive power of the variable for the value-weighted return from the CRSP. So, we also plot the cyclical part of the US industrial production (IP). The two indicators exhibit a strong negative correlation during recessions. The cyclical component of US IP goes down while uncertainty rises. However, during these periods, uncertainty seems to be timelier as its spikes come before the decrease of the IP. The contemporaneous correlation between the series is -0.19, while the correlation between the cyclical component with the one and two-period lagged uncertainty is -0.25 and -0.31.

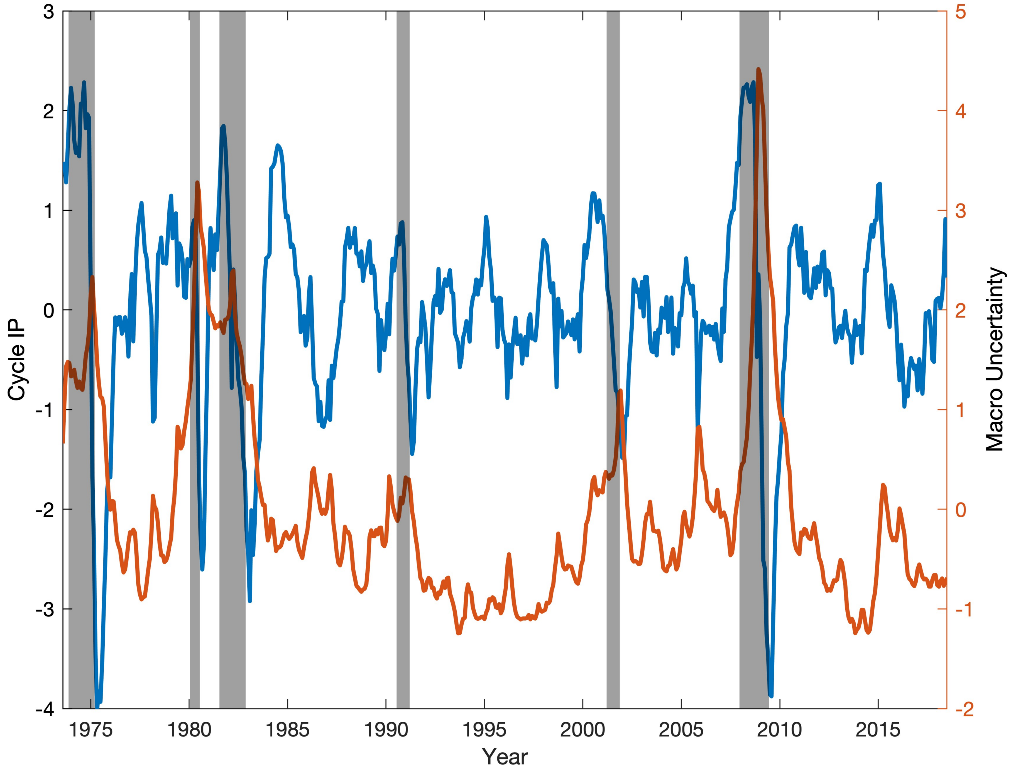
As Real/Macro uncertainty does not always translate to the financial market and vice versa, we also use the financial uncertainty index of Ludvigson, Ma, and Ng (2021). The indicator is the aggregate uncertainty of 148 financial series, such as the dividend-price and earnings-price ratios, spreads, yields, and sorted portfolios such as size, book-to-market, and momentum. Panel B Figure 2.1 plots the financial uncertainty index. In addition to the trough spikes, we observe additional ones outside these periods. A notable example is the spike of the 1987 stock market crash.

Table 2.1 – Descriptive statistics of the micro and characteristic portfolios

Panel A : All micro portfolios											
Quantile	0.05	0.25	0.50	0.75	0.95		0.05	0.25	0.50	0.75	0.95
Mean	5.56	8.43	11.06	13.30	16.39						
Std	13.27	17.48	19.44	21.09	23.40	Profitability	-0.96	-0.12	0.02	0.08	0.21
Kurtosis	5.11	5.85	6.40	7.07	12.61	Investment	-0.24	0.02	0.11	0.26	1.02
skewness	-0.52	-0.34	-0.11	0.08	0.75	Book to market	0.13	0.37	0.67	1.13	2.46
Panel A1 : Tiny-cap micro portfolios											
Quantile	0.05	0.25	0.50	0.75	0.95		0.05	0.25	0.50	0.75	0.95
Mean	8.42	10.78	12.54	14.22	17.11						
Std	16.86	18.88	20.22	21.49	23.52	Profitability	-1.04	-0.12	0.01	0.07	0.20
Kurtosis	5.85	6.33	6.71	7.60	16.20	Investment	-0.27	-0.01	0.09	0.24	1.02
Skewness	-0.28	-0.11	0.02	0.18	1.01	Book to market	0.13	0.41	0.73	1.20	2.59
Panel A2 : Small-cap micro Pportfolios											
Quantile	0.05	0.25	0.50	0.75	0.95		0.05	0.25	0.50	0.75	0.95
Mean	5.83	7.70	8.83	10.49	12.58						
Std	15.04	16.71	18.33	20.32	23.25	Profitability	-0.69	0.00	0.06	0.10	0.23
Kurtosis	5.19	5.59	5.84	6.24	8.84	Investment	-0.06	0.07	0.18	0.38	1.29
Skewness	-0.61	-0.48	-0.42	-0.35	-0.01	Book to market	0.10	0.27	0.46	0.76	1.61
Panel A3 : Big-cap micro portfolios											
Quantile	0.05	0.25	0.50	0.75	0.95		0.05	0.25	0.50	0.75	0.95
Mean	4.58	5.69	6.48	8.07	9.60						
Std	12.11	13.17	14.86	17.28	21.67	Profitability	-0.45	0.03	0.07	0.11	0.23
Kurtosis	4.88	5.21	5.39	5.67	6.59	Investment	-0.01	0.09	0.17	0.36	1.41
Skewness	-0.51	-0.41	-0.36	-0.31	-0.16	Book to market	0.08	0.22	0.38	0.66	1.29
Panel B : Characteristic-sorted portfolios											
Quantile	0.05	0.25	0.50	0.75	0.95						
Mean	5.37	8.24	10.20	12.23	14.77						
Std	15.13	17.58	19.36	21.69	26.86						
Kurtosis	4.34	5.06	5.63	6.31	8.23						
Skewness	-0.66	-0.49	-0.36	-0.22	0.34						

This leftmost side of this table presents the cross-sectional quantiles (0.05, 0.25, 0.5, 0.75, 0.95) of the average annualized excess returns (mean), annualized volatility (std), kurtosis and skewness of the micro portfolios. The rightmost side presents the cross-sectional quantiles of the characteristics used to construct the micro portfolios.

Panel A : Cyclical component of the US industrial production and macroeconomic uncertainty



Panel B :Financial uncertainty

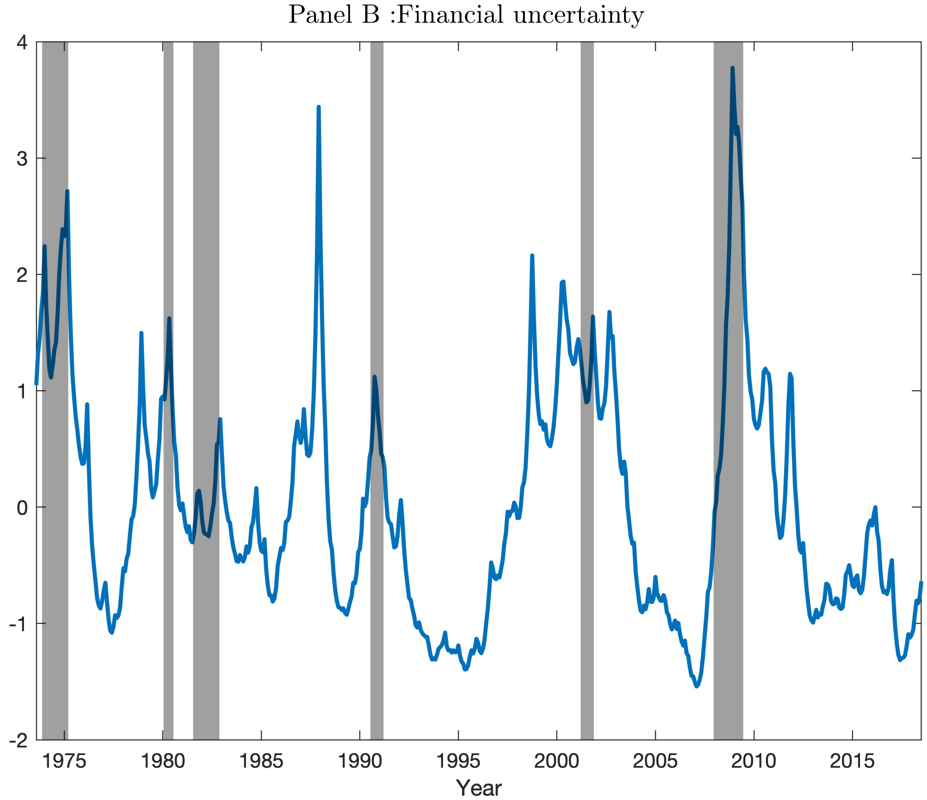


FIGURE 2.1 – Conditioning variables

2.4 Competing Models

We estimate and compare ten asset pricing models proposed in the literature. They are categorized in two types : theoretical or macroeconomic and anomaly-driven ones. Theoretical models often incorporate macroeconomic factors, whether derived from general equilibrium models or not, while anomalies-driven models directly derive factors from accounting-related data. We succinctly summarize each one of them.

Consumption Capital Asset Pricing Model (CCAPM). This model, presented in Hansen and Singleton (1982), assumes a representative agent with power utility function $u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}$, where γ is the relative risk aversion. The model yields a SDF of the form $Y_t = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}$, where β is the discount rate. Using the fact that non-linear models can be approximated by first order log linear function as

$$\frac{Y_t}{E[Y_t]} \approx 1 + y_t - E[y_t],$$

where y_t is the logarithm of Y_t, the initial SDF can be rewritten as $\frac{Y_t}{E[Y_t]} \approx 1 - \gamma \left(log(\frac{C_{t+1}}{C_t}) - E(log(\frac{C_{t+1}}{C_t}))\right)$. Therefore, in our specification, we use

$$y = 1 - \theta_{ndur} \left(\Delta c_{ndur} - E\left[\Delta c_{ndur}\right]\right),$$

where Δc_{ndur} is the log consumption growth rate of non durable goods (seasonally adjusted at annual rates). Nondurable consumption is the sum of real personal consumption expenditures on nondurable goods and services obtained from the US Bureau of Economic Analysis (BEA). θ_{ndur} is directly interpretable as the relative risk aversion coefficient.

Durable Consumption Capital Asset Pricing Model (DCAPM). In this model, as described by Yogo (2006), representative household has an intra period utility function that depends on durable and non durable goods. Non durable goods are consumed immediately after purchase, while durable goods last for more than one period. The model has an intra period utility given by $u(C,D) = \left[(1-\alpha)C^{1-\frac{1}{\rho}} + \alpha D^{1-\frac{1}{\rho}}\right]^{\frac{1}{1-\frac{1}{\rho}}}$,where C is the units of non durable goods, D is the stock of durable goods. $\alpha \in (0,1)$ and $\rho \geq 0$ is the elasticity of substitution between the two consumption goods. D follows the following equation

$$D_{t+1} = (1-\delta)D_t + E_t, \tag{2.12}$$

where δ is the depreciation rate and E_t is units of a durable consumption good. The inter temporal utility function is defined recursively as $U_{t+1} = \{(1-\beta)u(C_t, D_t)^{1-\frac{1}{\sigma}} + \beta E_t(U_{t+1}^{\frac{1}{\kappa}})\}^{1/(1-\frac{1}{\sigma})}$, $\kappa = (1-\gamma)/(1-1/\sigma)$. The parameter $\beta \in (0,1)$ is the household's subjective discount factor, $\sigma \geq 0$ is the Intertemporal Elasticity of Substitution, and $\gamma > 0$ determines the relative risk aversion. In addition, the household invests in an economy with N risky assets. The model yields the following SDF $Y_{t+1} = \left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\sigma}}\left(\frac{v(D_{t+1}/C_{t+1})}{v(D_t/C_t)}\right)^{1/\rho-1/\sigma} R_{W,t+1}^{1-\frac{1}{\kappa}}\right]^{\kappa}$, where $v(D/C) = \left[1-\alpha+\alpha(\frac{D}{C})^{1-\frac{1}{\rho}}\right]^{\frac{1}{1-\frac{1}{\rho}}}$. The SDF can be approximated as

$$\frac{Y_t}{E[Y_t]} \approx 1 - b_1(\Delta c_t - E[\Delta c_t]) + b_2(\Delta d_t - E[\Delta d_t]) + b_3(r_{wt} - E[\Delta d_t]),$$

where b_1, b_2, and b_3 are non linear function of α, κ, σ and ρ. So, their interpretation is not as straightforward as in CCAPM. We use this specification in our linear SDF as follows :

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

where r_{mkt} is the excess return on the value-weighted combined NYSE-AMEX-NASDAQ index, Δc_{dur} is log consumption growth rate of the per capita stock of durable goods. We construct the stock of durable goods by using the flow of personal consumption expenditures of durable goods from the BEA and Equation (2.12) with $\delta = 0.06$ and $D_0 = \frac{E_0}{g+\delta}$. g is the average growth rate of the durable personal consumption expenditure.

Conditional Consumption Capital Asset Pricing Model (CCAY). The third model is the conditional CCAPM of Lettau and Ludvigson (2001). It improves on the CCAPM by scaling the consumption variable with the log consumption–aggregate (human and non-human) wealth ratio. The latter summarizes the agent's expectations of future returns and has good forecasting power for excess returns. High consumption-wealth ratio goes with high wealth return in the future or low consumption growth rates and vice-versa. In practice, they use the cointegrating residual between log consumption,

c, log asset (nonhuman) wealth, a, and log labor income, y as the observable version of the ratio, hence the cay notation. The SDF of this model is

$$y = 1 - \theta_{ndur}\left(\Delta c_{ndur} - E\left[\Delta c_{ndur}\right]\right) - \theta_{ndur,cay}\left(\Delta c_{ndur} * cay - E\left[\Delta c_{ndur} * cay\right]\right),$$

where $\Delta c_{ndur} * cay$ is the interaction between nondurable consumption and the log consumption–aggregate wealth ratio.

Intermediary Asset Pricing Model (IAPM). All the theoretical models presented so far place household decisions at the center of the SDF. Intermediary asset pricing theory proposes a new perspective to understand risk in financial markets. In its setting, financial intermediaries such as traditional commercial banks, investment banks, and hedge funds act as marginal investors (see He and Krishnamurthy (2013)). Given the households' relative lack of expertise in trading assets, this view is quite cogent. He, Kelly, and Manela (2017) consider a two-agent economy populated by Households and Financial intermediaries. The Financial intermediaries are the marginal investors and maximize a power utility function $E\int_0^\infty e^{\rho t}u(c_t)dt$, where ρ is the discount rate and $u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}$. As the financial intermediaries' consumption is a fraction of its wealth, the SDF that arises in general equilibrium is $\Lambda_t = e^{\rho t}(\alpha W_t^I)^{-\gamma}$, where α is a positive constant and W_t^I is the wealth of the financial intermediaries. Let η_t be the intermediary sector's share of aggregate wealth in the economy. Then $\Lambda_t = e^{\rho t}(d\eta_t W_t)^{-\gamma}$. Using the fact that for any asset i with instantaneous return dR_t^i, $E_t(dR_t^i) - r_{f,t}dt = -E_t(dR_t^i.\frac{d\Lambda_{t+1}}{\Lambda_t})$, we have the following beta representation

$$E_t(dR_t^i) - r_{f,t}dt = \beta_{W,t}^i\lambda_W + \beta_{\eta,t}^i\lambda_\eta,$$

where β is the exposure of the factors to the returns. He, Kelly, and Manela (2017) use the capital ratio $\frac{Equity_t}{Assets_t}$ of the Federal Reserve Bank of New York primary dealers as η_t. They employ the growth rate of capital ratio, η_t^Δ, as a factor to implement the model. The latter is $\eta_t^\Delta = \frac{u_t}{\eta_{t-1}}$, where u_t is the innovations from the following $AR(1)$ model $\eta_t = \rho_0 + \rho\eta_{t-1} + u_t$. This implementation is equivalent to using an SDF of the form

$$y_t = 1 - \theta_W(r_{W,t} - E[r_{W,t}]) - \theta_\eta(\eta_t^\Delta - E[\eta_t^\Delta]).$$

Finally, they also use the Federal Reserve Bank of New York primary dealers' value-weighted return as a factor and find similar pricing performance of the model.

Human Capital Capital Asset Pricing Model (HCAPM). The fifth model is the Human capital CAPM of Jagannathan and Wang (1996). This model mainly tries to improve on the CAPM model by including a variable related to the business cycle through the yield spread between BAA- and AAA-rated bonds and a variable related to the aggregate wealth through the growth rate of the per capita labor income. Its SDF is

$$y = 1 - \theta_{jmkt} \left(r_{jmkt} - E\left[r_{jmkt}\right]\right) - \theta_{prem}(r_{prem} - E\left[r_{prem}\right]) - \theta_{lab}(r_{lab} - E\left[r_{lab}\right]),$$

where r_{jmkt} is the return on the valued-weighted combined NYSE-AMEX-NASDAQ index, r_{prem} is the lagged yield spread between BAA and AAA rated corporate bonds, and r_{lab} is the growth rate in per capita labor income. The latter is a two month moving average of the per capita labor income L_t. So, $r_{lab,t} = \frac{L_{t-1}+L_{t-2}}{L_{t-2}+L_{t-3}} - 1$. The two month moving average reduces the influence of measurement errors.

FF3. Besides the market, Fama and French (1993) uncovered two additional factors related to firms' size and value that can explain the cross-section of returns. Value is measured using the book-to-market ratio (BM), defined as the division of a firm's common stock's book value (BE) by its market value (ME). To build the factors, they sort the universe of stocks into two groups (small and big) using the NYSE stocks' median market value of Equity. In addition, they also break the stocks into three book-to-market (BM) groups. The breakpoints are the 30th (low), 40th (medium), and 70th (high) percentile of the ranked values of BM for the NYSE stocks. This process yields six value-weighted portfolios sorted independently on size and book-to-market. The size factor (Small Minus Big) is the difference between the average return on the three small portfolios and the average return on the three big portfolios, while the book-to-market factor (High Minus Low) factor is the difference between the average return on the two portfolios with low BM (value portfolios) and the average return on the two portfolios with high BM (growth portfolios). The SDF of this model is given by

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

where r_{mkt} is the excess return on the value-weighted combined NYSE-AMEX-NASDAQ index, r_{SMB} is the size factor and r_{HML} is the value factor.

Carhart. Researchers have proposed several anomalies besides the size and value factor. Jegadeesh and Titman (1993) documented the tendency of past well-performing stocks to outperform in the future. So, we include the Carhart (1997) model, which adds the momentum factor. This factor captures the returns associated with investing in winning stocks in contrast to losing stocks. For a month t, the momentum of a stock is the cumulative return of the stock during the 11 months covering month $t-11$ to $t-1$. In the literature, researchers build the momentum factor using the sorting method of Fama and French (1993), except that the sort takes place each month, and momentum replaces BM. The SDF of this model is given by

$$\begin{aligned} y &= 1-\theta_{mkt}\left(r_{mkt}-E\left[r_{mkt}\right]\right)-\theta_{SMB}\left(r_{SMB}-E\left[r_{SMB}\right]\right)-\theta_{HML}\left(r_{HML}-E\left[r_{HML}\right]\right) \\ &\quad -\theta_{MOM}\left(r_{MOM}-E\left[r_{MOM}\right]\right), \end{aligned}$$

where r_{MOM} is the momentum factor.

Liquidity. We also estimate the liquidity model of Pástor and Stambaugh (2003). It adds the liquidity factor to the Cahart model. Liquidity refers to the facility with which one can trade a security without significant price impact. A liquid stock tends to be available to trade. As such, compared to an illiquid stock, it experiences lower price change. Pástor and Stambaugh (2003) measure the liquidity of a stock i in month t, γ_{it}, using the following model :

$$r_{i,d+1,t}-r_{m,d+1}=\theta_{i,t}+\phi_{i,t}r_{i,d,t}+\gamma_{i,t}sign(r_{i,d,t}-r_{m,d,t})v_{i,d,t}+\epsilon_{i,d+1,t},$$

where $r_{i,d,t}$ is the return on stock i on day d in month t, $r_{m,d+1}$ is the return on the CRSP value-weighted market return on day d in month t, and $v_{i,d,t}$ is the dollar volume for stock i on day d in month t. For a market with N stocks, the aggregate market liquidity is given by $\hat{\gamma}_t=\left(\frac{V_t}{V_1}\right)\frac{1}{N}\sum_{i=1}^{N}\hat{\gamma}_{i,t}$. V_t is the total dollar value at the end of month $t-1$ of the stocks included in the previous average in month t, and month 1 corresponds to August 1962. In their model, they use a factor related to the innovations or unexpected changes of aggregate liquidity. The aggregate liquidity innovation is given by $\Delta\hat{\gamma}_t=\left(\frac{V_t}{V_1}\right)\frac{1}{N_t}\sum_{i=1}^{N_t}(\hat{\gamma}_{i,t}-\hat{\gamma}_{i,t-1})$. Instead of using the previous variable, they remove the long-run mean and potential correlation by running the following regression $\Delta\hat{\gamma}_t=a+b\Delta\hat{\gamma}_{t-1}+c\left(\frac{V_t}{V_1}\right)\hat{\gamma}_{t-1}+u_t$ and then define $L_t=\frac{1}{100}\hat{u}_t$ as the main aggregate liquidity innovation. In addition, each year, they sort stocks into 10 groups using their liquidation innovation beta of the previous year from the model below

$$r_{i,t} = \beta_i^0 + \beta_{i,MKT} r_{mkt,t} + \beta_{i,SMB} r_{SMB,t} + \beta_{i,HML} r_{HML,t} + \beta_{i,L} L_t + \epsilon_{i,t},$$

where $r_{i,t}$ is the excess return of stock i during period t and $r_{mkt,t}$, $r_{SMB,t}$, $r_{HML,t}$, and L_t are the market factor, size factor, value factor, and liquidity innovation, respectively, in month t. The liquidity factor r_{LIQ} is the difference between decile ten portfolio and decile one. Finally, the SDF of the model is given by

$$\begin{aligned} y &= 1 - \theta_{mkt}\left(r_{mkt} - E\left[r_{mkt}\right]\right) - \theta_{SMB}\left(r_{SMB} - E\left[r_{SMB}\right]\right) - \theta_{HML}\left(r_{HML} - E\left[r_{HML}\right]\right) \\ &\quad - \theta_{MOM}\left(r_{MOM} - E\left[r_{MOM}\right]\right) - \theta_{LIQ}\left(r_{LIQ} - E\left[r_{LIQ}\right]\right). \end{aligned}$$

FF5. The Fama and French (2015) model adds a profitability and an investment factor to FF3 and has the following SDF.

$$\begin{aligned} y &= 1 - \theta_{mkt}\left(r_{mkt} - E\left[r_{mkt}\right]\right) - \theta_{SMB}\left(r_{SMB} - E\left[r_{SMB}\right]\right) - \theta_{HML}\left(r_{HML} - E\left[r_{HML}\right]\right) \\ &\quad - \theta_{RMW}\left(r_{RMW} - E\left[r_{RMW}\right]\right) - \theta_{CMA}\left(r_{CMA} - E\left[r_{CMA}\right]\right). \end{aligned}$$

The variable, r_{RMW}, is the RMW (Robust Minus Weak) or profitability factor. Its construction is similar to the value factor, only operating profitability replaces BM. So, RMW is the difference between the return on the robust operating profitability portfolios minus the return on the weak operating profitability portfolios. In the same vein, the variable, r_{CMA}, is the CMA (Conservative Minus Aggressive) factor or investment factor. It is the difference between the return on the low investment portfolios (conservative) minus the return on the high investment portfolios (aggressive).

q5. The tenth model is the q-factor model of Hou et al. (2021), which assumes that

$$\begin{aligned} y &= 1 - \theta_{mkt}\left(r_{mkt} - E\left[r_{mkt}\right]\right) - \theta_{ME}\left(r_{ME} - E\left[r_{ME}\right]\right) - \theta_{I/A}\left(r_{I/A} - E\left[r_{I/A}\right]\right) \\ &\quad - \theta_{ROE}\left(r_{ROE} - E\left[r_{ROE}\right]\right) - \theta_{Eg}\left(r_{Eg} - E\left[r_{Eg}\right]\right). \end{aligned}$$

The variable, r_{ME}, is the difference between the return on small size portfolios and the return on the big size portfolios. The factor, $r_{I/A}$, represents the investment of companies. It is the difference (low-minus-high) between the return on low

I/A (investment-to-assets, which is the annual change in total assets divided by 1-year-lagged total assets) portfolios and the returns on the high I/A portfolios. The variable, r_{ROE}, is the return on equity factor, and is the difference (high-minus-low) between the return on high ROE portfolios and the return on the low ROE portfolios. Hou et al. (2021) construct the three factors using a triple 2-by-3-by-3 sort on size, investment, and ROE (profitability). Finally, r_{Eg} is the expected growth factor. It is the difference (high-minus-low) between the simple average of the returns on high expected investment-to-assets changes portfolios and the returns on the low expected investment-to-assets changes portfolios.

Panel A of Table 2.2 presents the characteristics of the traded factors of the different models described earlier : mean, volatility and cross-correlation. For macroeconomic variables, we build a mimicking portfolio using (2.6). Every factor carries a significant positive risk premium. Therefore, an investment in portfolios that follows these factors will yield positive returns in excess of the risk free rate. The expected growth factor has the highest risk premium (0.62%) followed by the momentum factor (0.60%), and the market factor (0.59%). The Financial intermediary factor provides the smallest positive return followed by the macroeconomic factors. However, macroeconomic factors present the highest t-stat. This is mainly due to their low volatilities.

Panel B describes the correlation between factors. Numerous factors within FF3 and q5 exhibit strong correlations. The correlation between the size factors are normal given they use the same characteristics variable. IA is very correlated with HML (0.68) and CMA (0.90). ROE and RMW have a correlation of 0.67. So the two models might span the same SDF. In the intermediary asset pricing model of He, Kelly, and Manela (2017), the intermediary factor is highly correlated with the market (0.80). Macroeconomic factors do not display high correlations with other variables but exhibit a moderate level of correlation among themselves. Particularly, durable and non durable consumption have a correlation of 0.58.

Table 2.2 – Characteristics of the factors

Panel A : Descriptive statistics

	Mean	Std	t-stat
MKT	0.594	4.515	3.059
SMB	0.269	3.005	2.083
HML	0.320	2.914	2.554
RMW	0.274	2.329	2.737
CMA	0.316	1.945	3.776
ME	0.226	3.057	1.720
IA	0.362	1.861	4.515
ROE	0.547	2.564	4.960
Eg	0.809	1.873	10.044
MOM	0.620	4.385	3.288
LIQ	0.416	3.510	2.754
HKMF	0.010	0.068	3.610
NCONS	0.048	0.104	10.724
DCONS	0.079	0.237	7.772
LAB	0.117	0.152	17.880

Panel B : Correlations

	SMB	HML	RMW	CMA	ME	IA	ROE	Eg	MOM	LIQ	HKMF	NCONS	DCONS	LAB
MKT	0.226	-0.273	-0.252	-0.396	0.252	-0.366	-0.195	-0.419	-0.135	0.009	0.815	0.279	0.143	0.273
SMB		-0.053	-0.388	-0.038	0.981	-0.130	-0.385	-0.381	-0.026	0.000	0.122	0.298	0.285	0.231
HML			0.159	0.687	-0.195	0.675	-0.100	0.163	-0.189	0.032	0.021	0.048	0.116	0.139
RMW				0.057	-0.456	0.160	0.665	0.475	0.092	-0.008	-0.205	-0.113	-0.160	-0.038
CMA					-0.123	0.905	-0.054	0.269	0.000	-0.001	-0.217	0.005	0.068	0.118
ME						-0.219	-0.393	-0.406	-0.002	0.007	0.106	0.295	0.284	0.214
IA							0.070	0.328	0.030	0.008	-0.160	0.041	0.078	0.140
ROE								0.542	0.495	-0.085	-0.207	-0.140	-0.214	-0.124
Eg									0.338	-0.048	-0.341	-0.112	-0.116	-0.066
MOM										0.017	-0.268	-0.002	-0.076	-0.051
LIQ											-0.029	0.030	0.114	0.013
HKMF												0.249	0.125	0.295
NCONS													0.431	0.581
DCONS														0.447

2.5 Results

First, we provide the results of the estimation of the different asset pricing models using the regularized HJ distance (2.3). Second, we include additional information through the uncertainty indices to estimate the conditional version of the models. Third, we compare the unconditional and conditional asset pricing models.

2.5.1 SDF parameter estimates

In this segment, we are mainly interested in the SDF parameters estimates $\hat{\theta}$ as well as their t-ratios. The t-ratio of a particular factor provides the statistics of the test $H_0 : \theta = 0$. It is computed using the distribution (2.4). A statistically significant non-null parameter is evidence that the factor is a priced source of risk in the financial market and investor cares about it. Table 2.3 presents the results when the micro portfolios are the sole test assets. For each model, we report the SDF parameter estimates $(\hat{\theta})$, the t-ratio, the regularized HJ, and the level of penalization (α). One should keep in mind that the higher the penalization is, the lower the regularize HJ distance is.

For the consumption-based models, the nondurable consumption is significantly priced in the CCAPM at 10% level. It is interesting to see how this factor behaves in the different size classes of micro-portfolios [6]. The SDF parameter is significant at 1% with the tiny and small capitalization micro-portfolios, whereas it is not priced in the big firms. These observations might be explained by the fact that consumption variable and returns of smaller capitalization firms are more pro-cyclical. In the DCAPM of Yogo (2006), after taking into account the market factor, the nondurable and durable consumption are not priced anymore. Finally, the conditional variable in the model of Lettau and Ludvigson (2001) is only statistically significant in the small-cap micro portfolios.

For the IAPM model of He, Kelly, and Manela (2017), the intermediary factor (capital ratio growth rate) does not carry a significant value. However, the SDF parameter is significant when we use the value weighted return of the financial intermediaries. In addition, despite not being priced globally, the capital ratio growth rate is priced in the tiny and big-cap portfolios. The evidence is stronger in the big-cap portfolios.

6. We report the tables for tiny, small, and big-cap portfolios in the appendix for ease of presentation. See Section 2.9.

TABLE 2.3 – Estimates and misspecification-robust t-ratios of SDF parameters : Micro portfolios and unscaled factors

	CCAPM				DCAPM				HCAPM		
	$\hat{\theta}_{ndur}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{ndur}$	$\hat{\theta}_{dur}$			$\hat{\theta}_{hmkt}$	$\hat{\theta}_{prem}$	$\hat{\theta}_{lab}$
SDF	2.684			0.025	0.757	0.908			0.027	0.043	-0.771
t-ratio	1.771			1.750	1.181	1.135			1.512	0.048	-0.435
HJ	0.034			0.218					0.103		
α	0.200			0.0001					0.004		
	CCAY				IAPM				FF3		
	$\hat{\theta}_{ndur}$	$\hat{\theta}_{ndur.cay}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{\eta}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$
SDF	0.051	88.101			-0.047	7.339			0.043	0.027	0.088
t-ratio	0.046	1.357			-0.948	1.625			2.830	1.414	3.068
HJ	0.475				0.220				0.201		
α	0.0001				0.0001				0.0001		
	Cahart						Liquidity				
	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{MOM}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{MOM}$	$\hat{\theta}_{LIQ}$
SDF	0.053	0.023	0.102	0.033			0.053	0.024	0.102	0.032	0.669
t-ratio	3.331	1.311	3.881	1.451			3.312	1.331	3.898	1.457	0.213
HJ	0.193						0.193				
α	0.0001						0.0001				
	FF5						Q5				
	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{RMW}$	$\hat{\theta}_{CMA}$		$\hat{\theta}_{mkt}$	$\hat{\theta}_{ME}$	$\hat{\theta}_{I/A}$	$\hat{\theta}_{ROE}$	$\hat{\theta}_{Eg}$
SDF	0.052	0.025	0.031	-0.001	0.116		0.075	0.050	0.135	-0.069	0.216
t-ratio	3.442	1.165	0.744	-0.015	2.226		3.458	2.579	3.079	-1.299	1.940
HJ	0.192						0.175				
α	0.0001						0.0001				

The table presents the estimation results of ten asset pricing models with unscaled factors. The models are estimated using the monthly returns of the 2159 micro portfolios. Data are monthly from July 1973 to June 2018. CCAY model uses quarterly data. We report parameter estimates $\hat{\theta}$, t-ratios under model misspecification, the regularized HJ and the penalization level α.

In the FF3 model, results are very similar to what we observe in the literature. Globally, only the market and the value factors are significantly priced at 1% level. These factors are also the ones that are priced in the big-cap portfolios. The value factor is important regardless of the size of the micro portfolios. For the tiny-cap portfolios, in addition to the value factor, the size is significant. We estimate several extensions of the FF3. In the CAHART model, adding the momentum factor does not lead to much change. However, the factor is significant at 10% when we remove the tiny-cap portfolios. Therefore, the phenomenon is driven mainly by the big firms. Furthermore, the inclusion of the liquidity factor to the CAHART model does not improve the models. We have the same level of pricing errors. In the FF5, pricing errors improve slightly. In addition, the CMA factor drives out the other risk factors. The profile of the model is different across the micro portfolio sizes. We obtain the same results with the small-cap returns. The value factor is the sole priced variable in the tiny-cap returns. For the big-cap returns, all the factors have significant SDF parameters except the value factor.

In the q5 model, except for the ROE factor, all the factors are significant sources of risk in the micro portfolios. Contrary to the FF5, the q5 factors perform well in the tiny-cap returns. The SDF parameters have absolute t-ratios higher than 1.67. Therefore, they are priced at least at a 10% level. For the small and big-cap returns, market, size and investment are the main sources of risk.

So far, our attention has been directed toward the micro portfolios, yet characteristic-sorted portfolios represent an alternative class of test assets commonly employed for assessing asset pricing models. Table 2.4 presents the SDF parameter estimates and t-ratios of the ten models using them.

Table 2.4 – Estimates and misspecification-robust t-ratios of SDF parameters : Characteristic-sorted portfolios and unscaled factors

	CCAPM			Yogo					HCAPM		
	$\hat{\theta}_{ndur}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{ndur}$	$\hat{\theta}_{dur}$			$\hat{\theta}_{hmkt}$	$\hat{\theta}_{prem}$	$\hat{\theta}_{lab}$
SDF	2.171			0.030	0.086	0.293			0.033	0.129	0.314
t-ratio	2.106			2.499	0.340	0.640			2.581	0.425	0.941
HJ	0.088			0.650					0.647		
α	0.033			0.0001					0.0001		
	CCAY				IAPM				FF3		
	$\hat{\theta}_{ndur}$	$\hat{\theta}_{ndur.cay}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{\eta}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$
SDF param	0.116	-80.943			-0.003	2.687			0.039	0.016	0.070
t-ratio	0.175	-2.126			-0.083	0.858			2.754	0.976	2.818
HJ	0.090				0.082				0.022		
α	0.2				0.041				0.200		
	Cahart						Liquidity				
	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{MOM}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{MOM}$	$\hat{\theta}_{LIQ}$
SDF	0.050	0.020	0.084	0.060			0.045	0.021	0.067	0.047	3.563
t-ratio	3.373	1.287	3.780	3.467			3.093	1.443	3.201	3.193	1.622
HJ	0.051						0.582				
α	0.016						0.0001				
	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{RMW}$	$\hat{\theta}_{CMA}$		$\hat{\theta}_{mkt}$	$\hat{\theta}_{ME}$	$\hat{\theta}_{I/A}$	$\hat{\theta}_{ROE}$	$\hat{\theta}_{Eg}$
SDF	0.056	0.041	-0.014	0.086	0.135		0.073	0.076	0.145	0.054	0.164
t-ratio	4.143	2.428	-0.470	2.899	3.360		3.385	3.531	2.406	0.890	1.153
HJ	0.116						0.004				
α	0.004						0.200				

The table presents the estimation results of ten asset pricing models with unscaled factors. The models are estimated using the monthly returns of the 738 characteristics-sorted portfolios. Data are monthly from July 1973 to June 2018. CCAY model uses quarterly data. We report parameter estimates $\hat{\theta}$, t-ratios under model misspecification, the regularized HJ and the penalization level α.

For the characteristic-sorted portfolios, the nondurable consumption is priced in the CCAPM and when it is associated with the consumption–aggregate wealth ratio in CCAY. However, the durable consumption is not a source of risk in DCAPM. In HCAPM, the market is the only priced risk factor. The intermediary factor in IAPM does not have a significant value. This is the case even when we use traded version of the factor (value-weighted return of the financial intermediaries). For the FF3 and its extensions, compared to the micro portfolios, we note that the momentum factor is strongly priced. Overall, the results are similar to the big-cap portfolios analyzed previously. This might be due to the use of the value-weighted portfolios.

All the portfolios, despite being different in their construction, should still satisfy the fundamental Equation of Asset pricing. Therefore, we estimate the models using the combined portfolios. Using the combined portfolios as test assets yields more robust results. Furthermore, as the characteristic-sorted portfolios are also managed payoffs, one can consider them as conditional asset pricing models, where the conditional information is incorporated in the test assets. The SDF parameters and their t-stats are in Table 2.5.

TABLE 2.5 – Estimates and misspecification-robust t-ratios of SDF parameters : All the portfolios and unscaled factors

	CCAPM	DCAPM			HCAPM		
	$\hat{\theta}_{ndur}$	$\hat{\theta}_{mkt}$	$\hat{\theta}_{ndur}$	$\hat{\theta}_{dur}$	$\hat{\theta}_{hmkt}$	$\hat{\theta}_{prem}$	$\hat{\theta}_{lab}$
SDF	1.877	0.027	0.254	0.708	0.026	0.308	-0.453
t-ratio	1.823	2.130	0.682	1.022	1.725	0.336	-0.301
HJ	0.107	0.444			0.105		
α	0.008	0.0001			0.008		

	CCAY		IAPM		FF3		
	$\hat{\theta}_{ndur}$	$\hat{\theta}_{ndur.cay}$	$\hat{\theta}_{mkt}$	$\hat{\theta}_{\eta}$	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$
SDF	0.287	-90.224	-0.042	6.410	0.037	0.037	0.103
t-ratio	0.435	-1.823	-1.056	1.788	2.453	1.860	3.562
HJ	0.038		0.117		0.009		
α	0.200		0.004		0.200		

	Cahart				Liquidity				
	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{MOM}$	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{MOM}$	$\hat{\theta}_{LIQ}$
SDF	0.043	0.040	0.110	0.028	0.043	0.040	0.110	0.028	-0.409
t-ratio	2.652	2.284	4.079	1.046	2.767	2.168	4.809	0.800	-0.020
HJ	0.007				0.007				
α	0.200				0.200				

	FF5					Q5				
	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{RMW}$	$\hat{\theta}_{CMA}$	$\hat{\theta}_{mkt}$	$\hat{\theta}_{ME}$	$\hat{\theta}_{I/A}$	$\hat{\theta}_{ROE}$	$\hat{\theta}_{Eg}$
SDF	0.064	0.036	-0.057	0.025	0.293	0.105	0.067	0.122	-0.168	0.494
t-ratio	3.143	1.557	-0.757	0.435	2.343	3.807	2.379	2.106	-2.452	2.937
HJ	0.005					0.003				
α	0.200					0.200				

The table presents the estimation results of ten asset pricing models with unscaled factors. The models are estimated using the monthly returns of the 3198 characteristics-sorted and micro portfolios. Data are from July 1973 to June 2018. CCAY model uses quarterly data. We report parameter estimates $\hat{\theta}$, t-ratios under model misspecification, the regularized HJ and the penalization level α.

Overall, we note that the nondurable consumption have significant SDF parameter in the CCAPM and CCAY models. But, when one takes into account the market, the factor and its durable counterpart are no longer priced. The labor variable is not priced in the HCAPM model, while the intermediary factor of IAPM is significant at 10%.

In the FF3, the size factor is now priced at 10% in addition to the market and value factor. The additional factors of the CAHART and LIQ models are not priced anymore. In the FF5, market and CMA are significant, while all the factors in the Q5 have non null SDF parameters.

2.5.2 Conditional models

So far, the analysis assumes that the SDF parameters remain constant over time. We can relax this assumption by modeling them as a function of observable variables called conditioning variables. We use the Ludvigson, Ma, and Ng (2021) macro and financial uncertainty indicators as conditioning variables. Conditional models have many SDF parameters. Therefore, we only present the Wald statistics of the test that the SDF parameters associated with the conditioning variables are simultaneously null. This test is equivalent to a HJ equality test between the conditional and the unconditional models. As a result, rejecting the null hypothesis implies a significant HJ difference between the conditional and unconditional models and conditional variables provide useful additional information.

Table 2.6 presents the results of Wald tests on the factors of the conditional models. In general, six out of the nine models present significant improvements upon conditioning with the uncertainty indicators for the micro-portfolios. The Financial uncertainty indicator seems to be the crucial conditioning variable, as most models are not improved when we use the macro uncertainty indicator or the cycle component of US IP alone. Its importance emanates from the fact it also tracks events specific to the financial market. Except for CCAPM, all the macroeconomic models have significant smaller pricing errors after conditioning. For the anomalies-based models, FF3 and CAHART do not improve. Therefore, increasing the number of variables is not a panacea for significantly decreasing pricing errors. When we consider the full conditional models with the characteristic-sorted and micro portfolios as the test assets, all the models improved except CCAPM and IAPM.

Three models display interesting results (see Table 2.14 in Section 2.9.4 for more details). The first is the DCAPM, where the durable consumption scaled by the macroeconomic uncertainty index is significantly priced at 5%. There is also the case for HCAPM. In this model, the scaled labor variable shows a t-stat of 1.957. Finally, the momentum factor of the CAHART model has a non-null value. It remains significant even when we add the liquidity variable.

Table 2.6 – Wald test of conditional models : factors scaled by Macro and Financial Uncertainty indicators

	CCAPM	DCAPM	HCAPM	IAPM	FF3	CAHART	LIQ	FF3	Q5
Micro portfolios									
α	0.004	0.0001	0.0042	0.0001	0.0001	0.0001	0.2000	0.0083	0.2000
HJ	0.100	0.144	0.034	0.178	0.173	0.160	0.001	0.013	0.000
Wald	0.023	15.539	19.819	10.578	8.666	8.943	19.239	23.191	32.484
p-val	0.989	0.017	0.003	0.032	0.193	0.347	0.037	0.010	0.000
Tiny-cap Micro portfolios									
α	0.004	0.0001	0.0001	0.2000	0.0001	0.0001	0.0001	0.0001	0.0001
HJ	0.111	0.144	0.084	0.003	0.171	0.154	0.113	0.140	0.086
Wald	0.126	14.775	25.502	17.719	6.008	9.362	20.752	20.029	33.604
p-val	0.939	0.022	0.000	0.001	0.422	0.313	0.023	0.029	0.000
Small-cap Micro portfolios									
α	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
HJ	0.071	0.031	0.036	0.050	0.049	0.030	0.022	0.026	0.025
Wald	0.265	13.277	16.531	6.253	4.742	4.992	6.708	9.263	9.494
p-val	0.876	0.039	0.011	0.181	0.577	0.758	0.753	0.507	0.486
Big-cap Micro portfolios									
α	0.0001	0.0001	0.0001	0.0001	0.200	0.0001	0.0001	0.0001	0.0001
HJ	0.101	0.037	0.036	0.059	0.000	0.023	0.022	0.022	0.014
Wald	0.072	9.420	4.538	1.289	5.617	8.439	6.903	11.191	11.115
p-val	0.965	0.151	0.604	0.863	0.467	0.392	0.735	0.343	0.349
Characteristic-sorted portfolios									
α	0.1062	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
HJ	0.043	0.605	0.614	0.646	0.601	0.551	0.545	0.491	0.477
Wald	3.183	14.539	7.373	3.180	15.363	18.651	16.463	24.565	17.987
p-val	0.204	0.024	0.288	0.528	0.018	0.017	0.087	0.006	0.055
Micro and Characteristic-sorted portfolios									
α	0.0123	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
HJ	0.092	0.395	0.397	0.447	0.408	0.367	0.365	0.314	0.282
Wald	1.564	11.780	13.786	3.396	11.405	21.218	23.438	16.362	29.329
p-val	0.458	0.067	0.032	0.494	0.077	0.007	0.009	0.090	0.001

The table presents the estimation results of nine conditional asset pricing models. Factors are scaled with the macroeconomic and financial uncertainty indices. Data are monthly from July 1973 to June 2018. We report the Wald statistics of the test that the SDF parameters associated with the scaled variables are simultaneously null, the regularized HJ and the penalization level α.

2.5.3 Model comparison

In this section, we investigate whether the models estimated above present differences in expected returns' explanatory power, as measured by squared HJ distance.

The difference in squared HJ of two models is analog to the difference between the squared Sharpe Ratios of the factor-mimicking portfolio generated by each model. As a result, this test examines which model can extract higher investment gains from the test assets. We compare the models by assuming none can generate the exact squared Sharpe Ratio if one invests directly in the test assets : this is the misspecification assumption. In addition, we choose a single value of α to compare models. Changing α for each comparison leads to different errors in the fundamental Equation of Asset pricing, whereas the comparisons necessitate putting all the models on the same scale. We choose the smallest α ($\alpha = 0.0001$) to put the emphasis on the Fundamental Equation of Asset Pricing. Panel A of Table 2.7 presents the difference in sample squared Sharpe Ratio or regularized square HJ of the models and Panel B gives the p-value of the test.

TABLE 2.7 – Comparison of regularized squared HJ distances : characteristic-sorted and micro portfolios with non scaled factors

Panel A : Sample regularized squared HJ distance difference between models								
	DCAPM	HCAPM	HKM	FF3	CAHART	LIQ	FF5	Q5
CCAPM	0.025	0.012	0.011	0.038	0.067	0.070	0.087	0.114
DCAPM		-0.013	-0.014	0.013	0.043	0.046	0.062	0.090
HCAPM			0.000	0.026	0.056	0.059	0.075	0.103
HKM				0.027	0.056	0.059	0.075	0.103
FF3					0.030	0.033	0.049	0.077
CAHART						0.003	0.019	0.047
LIQ							0.016	0.044
FF5								0.028
Panel B : P-value of the equality test of two regularized square HJ distances								
	DCAPM	HCAPM	HKM	FF3	CAHART	LIQ	FF5	Q5
CCAPM	0.273	0.362	0.391	0.101	0.013	0.012	0.002	0.000
DCAPM		0.517	0.685	0.641	0.130	0.108	0.078	0.005
HCAPM			0.934	0.178	0.021	0.020	0.004	0.000
HKM				0.135	0.018	0.017	0.002	0.000
FF3					0.054	0.038	0.009	0.006
CAHART						0.248	0.432	0.055
LIQ							0.520	0.089
FF5								0.316

Panel A presents the sample difference in squared HJ distance between models in row i and models in column j, $\hat{\delta}^2_{\alpha,i} - \hat{\delta}^2_{\alpha,i}$. Factors are not scaled. The models are estimated using the monthly returns of the 3198 characteristics-sorted and micro portfolios. Data are from July 1973 to June 2018. Panel B reports the the p-value of the test $H_0 : \delta^2_i = \delta^2_j$. $\alpha = 0.0001$.

The comparison between two models depends on the form of their SDFs. If the SDF of one model 1, y_1, nests the SDF of model 2, y_2, we first conduct a pretest of whether the factors of the bigger model (model 2) have null SDF parameters. We do that because the normality test (2.7) requires that $y_1 \neq y_2$. For example, to compare CCAPM and DCAPM, we first test that the SDF parameter of the market and durable consumption are non-null. We employ a Wald test that comes from the asymptotic normal distribution of the SDF parameters. If the null hypothesis of the test is accepted, we conclude directly that the two models are the same. However, if we reject the null hypothesis, then $y_{CCAPM} \neq y_{DCAPM}$. Then, we proceed with test (2.7), for which the rejection of the null hypothesis entails that the model with lower pricing errors performs better.

When the models are non-nested but have common factors, we also use a sequential test. For example, IAPM and FF3 are non-nested models. To compare these two models, we first test whether the non overlapping factors of the models are simultaneously non-null, i.e, the intermediary, size, and value factor are non-null. We use the normality test only when the null hypothesis is rejected.

As displayed in Table 2.7, the result of the pairwise unconditional model comparisons can be summarized as follows. First, macroeconomic models do not exhibit significant differences in pricing errors. The differences in regularized squared HJ have p-values higher than 0.273. Furthermore, the models have the same performance as FF3. Second, dissimilarities between the macroeconomic and anomalies-based models emerge when one adds factors to the FF3. For example, CAHART dominates all the macroeconomic models with p-values higher than 0.021, except DCAPM. Third, for anomalies-based models, FF3 is dominated by CAHART and LIQ. Therefore, adding the momentum and liquidity factors, despite non-significant SDF parameters, improves the pricing performance of the models. The two models have the same pricing errors as FF5. However, q5 outperformed them. Finally, we do not observe a significant difference between FF5 and q5.

We also compare the conditional models as their performance is better than their unconditional counterparts. Table 2.8 presents the difference in regularized squared HJ and the p-value of the equality tests. Compared to unconditional pairwise comparisons, new results emerge. Notably, the conditional models of DCAPM and HCAPM improved such that, except for q5, their differences with anomaly-based models are insignificant. In addition, they dominate CCAPM and HKM at 5% and 10%, respectively. The conditional FF3 is significantly different from CCAPM and IAPM.

Until now, we have resorted to pairwise comparison tests that lead to many tests per model. To compare one model against the rest, we perform eight tests. The size of the series of tests might be larger than the theoretical one we are looking at. A solution to this problem is the multiple models' comparison test of Wolak (1989) presented in Section 2.2. It simultaneously tests the null hypothesis that a given model, the benchmark, performs at least as well as a group of models against the hypothesis that one of the models has a lower squared HJ distance (or perform better).

As in the previous pairwise comparisons, we take certain precautions before using the multiple comparison test. We remove alternative models nested by the benchmark model as they already have 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 these models have at least a lower square HJ distance. For example, to compare CCAPM against the other models, we removed DCAPM, which nests it. Within the alternatives, we remove FF3 nested by FF5 and CAHART nested by LIQ. As a result, we compare CCAPM against the five remaining models.

TABLE 2.8 – Comparison of regularized squared HJ distances : Characteristic-sorted and micro portfolios with factors scaled by macro and financial uncertainty indicators

Panel A : Sample regularized squared HJ distance difference between models								
	DCAPM	HCAPM	IAPM	FF3	CAHART	LIQ	FF5	Q5
CCAPM	0.070	0.068	0.017	0.056	0.098	0.100	0.151	0.183
DCAPM		-0.002	-0.055	-0.013	0.028	0.030	0.081	0.113
HCAPM			-0.053	-0.011	0.030	0.032	0.083	0.115
IAPM				0.039	0.081	0.082	0.133	0.166
FF3					0.041	0.043	0.094	0.127
CAHART						0.002	0.053	0.085
LIQ							0.051	0.083
FF5								0.032
Panel B : P-value of the equality test of two regularized square HJ distances								
	DCAPM	HCAPM	IAPM	FF3	CAHART	LIQ	FF5	Q5
CCAPM	0.017	0.023	0.364	0.067	0.001	0.001	0.001	0.000
DCAPM		0.944	0.071	0.696	0.457	0.419	0.095	0.004
HCAPM			0.082	0.750	0.447	0.406	0.121	0.008
IAPM				0.099	0.000	0.000	0.001	0.000
FF3					0.028	0.022	0.005	0.001
CAHART						0.735	0.116	0.008
LIQ							0.132	0.011
FF5								0.434

Panel A presents the sample difference in squared HJ distance between models in row i and models in column j, $\hat{\delta}^2_{\alpha,i} - \hat{\delta}^2_{\alpha,i}$. Factors are scaled with the macroeconomic and financial uncertainty indices. The models are estimated using the monthly returns of the 3198 characteristics-sorted and micro portfolios. Data are from July 1973 to June 2018. Panel B reports the the p-value of the test $H_0 : \delta^2_i = \delta^2_j$. $\alpha = 0.0001$.

Table 2.9 presents the results of the multiple comparison tests for the unconditional and conditional models. The test determines whether the benchmark model (each line) performs at least as well as the remaining models. With unconditional models, we note that the null hypothesis is rejected for all the theoretical models, i.e., these models do not perform at least as well as the alternatives. We reached the same conclusion for FF3 and CAHART but at 5 and 10 percent. Therefore, LIQ, FF5, and q5 best all the other models. However, we cannot significantly differentiate the three models. We have the same results when we use scaled factors (conditional models), except that LIQ is significantly dominated by FF5 and q5.

TABLE 2.9 – Multiple comparison tests of conditional and unconditional models

Benchmark	q	Unconditional			Conditional		
		$\hat{\delta}^2_\alpha$	LR	p-Value	$\hat{\delta}^2_\alpha$	LR	p-Value
CCAPM	5	0.468	7.685	0.000	0.463	13.076	0.000
DCAPM	5	0.444	5.731	0.003	0.395	9.623	0.000
HCAPM	5	0.457	7.529	0.001	0.397	3.564	0.015
IAPM	5	0.458	10.092	0.000	0.447	16.906	0.000
FF3	4	0.431	3.718	0.012	0.408	5.169	0.003
CAHART	5	0.401	1.852	0.099	0.367	4.041	0.018
LIQ	5	0.398	1.449	0.141	0.365	3.698	0.025
FF5	5	0.382	0.503	0.363	0.314	0.306	0.411
q5	5	0.354	0.503	0.363	0.282	0.210	0.790

The table presents the sample squared HJ distance for unconditional and conditional models, the Wolak LR statistics, and its P-value. q is the number of alternative models. The models are estimated using the monthly returns of the 3198 characteristics-sorted and micro portfolios. Data are from July 1973 to June 2018. $\alpha = 0.0001$.

2.5.4 Robustness analysis

We perform additional comparison analyses by changing the variables used to form the micro portfolios and the number of stocks in each micro portfolio. First, we reform micro portfolios of 10 stocks using the earning-price ratio, investment and return on equity (Section 2.10.1). Second, we reform the micro portfolios by using portfolios of 5 stocks (Section 2.10.2). There are slight differences among the macroeconomic/theoretical models. However, we still observe the dichotomy between the macroeconomic/theoretical and the anomaly-based models, which display better performance. The multiple comparisons display that FF5 and q5 dominate all the models, but we cannot differentiate them. These two models dominate all the others.

2.6 Conclusion

In this paper, we use 3198 test assets that combined the characteristics-sorted and micro portfolios to estimate and compare ten unconditional and conditional asset pricing models. The micro portfolios are formed using a small group of stocks (5 to 10 stocks). Hence they are analogous to individuals stocks. We use a regularized version of the well-known HJ distance. The extension allows for the use of a large number of test assets by replacing the covariance matrix with a regularized version. In addition, it is equivalent to allowing pricing errors for the true SDF. In the case of excess returns, using the regularized HJ distance to compare models is identical to comparing their Sharpe ratios or their cross-sectional R^2.

We studied four macroeconomic/theoretical models such as the Consumption Capital Asset Pricing Model (CCAPM), the Durable Consumption Capital Asset Pricing Model (DCAPM) , the Human Capital Asset Pricing Model, and the Intermediary Asset pricing model (IAPM). In addition, we included five anomaly-driven models such as the three (FF3) and Five-factor (FF5) Model, the Carhart model, which adds momentum to FF3, the Liquidity Model, and the Augmented q-Factor Model. We also looked at the conditional version of these models by modeling the SDF parameters as a function of the recent macroeconomic and financial uncertainty indices. Our results show that conditional models are much better that their unconditional counterpart. Furthermore, the anomaly-driven models have bigger explanatory power than the macroeconomic/theoretical models. Finally, the Five-factor (FF5) and the Augmented q-Factor Models are similar and dominate all the models.

Our analysis can be extended in a number of ways. For example, it would be desirable to develop approaches to compare linear and non linear models with a large number of test assets. Future research should also address the use of the HJ distance in unbalanced datasets as we had recourse to a matrix completion to balance the micro portfolios data.

2.7 Appendix A : definition of the financial ratios

We use the CRSP and quarterly Computstat datasets. The characteristics used to compute the stocks' average returns are : book-to-market ratio, profitability, investment, earning-price ratio and return on equity. We use the same definitions as Barras (2019) and present them below for completeness.

Book-to-market ratio. It is equal to the ratio of the book value of equity to the market value of equity. Bali et al. (2016, P. 177) give a detailed implementation of this ratio. The book value for year is defined as total assets minus liabilities, plus balance sheet deferred taxes and investment tax credit (if available), minus preferred shares stock liquidating values (if available), or carrying value (if available) in the fiscal year ending in the calendar year $t-1$. The market value for year t equals the price times shares outstanding at the end of December of year $t-1$.

Profitability. It is defined as revenues minus cost of goods sold, minus selling, general, and administrative expenses, minus interest expense all divided by the book value of equity. Each of these variables is computed using data in the fiscal year ending in the calendar year $t-1$.

Investment. Investment for year t is computed as the relative change in total assets between the fiscal years ending in calendar years $t-2$ and $t-1$.

Earning-price ratio. It is the ratio of income before extraordinary items in the fiscal year ending in the calendar year $t-1$ to market value measured at the end of December of year $t-1$.

Return on equity. ROE for year t is defined as income before extraordinary items in the fiscal year ending in the calendar year $t-1$ divided by book equity in the fiscal year ending in the calendar year $t-2$.

2.8 Appendix B : list of characteristic portfolios

TABLE 2.10 – List of characteristic-sorted portfolios

100 Portfolios Formed on Size and Book-to-Market
100 Portfolios Formed on Size and Operating Profitability
100 Portfolios Formed on Size and Investment
25 Portfolios Formed on Book-to-Market and Operating Profitability
25 Portfolios Formed on Book-to-Market and Investment
25 Portfolios Formed on Operating Profitability and Investment
32 Portfolios Formed on Size, Book-to-Market, and Operating Profitability
32 Portfolios Formed on Size, Book-to-Market, and Investment
32 Portfolios Formed on Size, Operating Profitability, and Investment
6 Portfolios Formed on Size and Earnings/Price
6 Portfolios Formed on Size and Cashflow/Price
6 Portfolios Formed on Size and Dividend Yield
25 Portfolios Formed on Size and Momentum
25 Portfolios Formed on Size and Short-Term Reversal
25 Portfolios Formed on Size and Long-Term Reversal
25 Portfolios Formed on Size and Accruals
25 Portfolios Formed on Size and Market Beta
25 Portfolios Formed on Size and Net Share Issues
25 Portfolios Formed on Size and Variance
25 Portfolios Formed on Size and Residual Variance
49 Portfolios Formed Industry

2.9 Appendix C : Asset pricing models by type of micro portfolios

2.9.1 Asset pricing models for tiny-cap micro portfolios

TABLE 2.11 – Estimates and misspecification-robust t-ratios of SDF parameters : Tiny-cap Micro portfolios and unscaled factors

	CCAPM	DCAPM			HCAPM		
	$\hat{\theta}_{ndur}$	$\hat{\theta}_{mkt}$	$\hat{\theta}_{ndur}$	$\hat{\theta}_{dur}$	$\hat{\theta}_{hmkt}$	$\hat{\theta}_{prem}$	$\hat{\theta}_{lab}$
SDF	2.664	0.034	0.590	1.111	0.041	-0.658	-1.075
t-ratio	2.600	1.817	0.778	1.116	2.412	-1.146	-1.235
HJ	0.077	0.229			0.238	0.238	0.238
α	0.016	0.0001			0.0001		

	CCAY		IAPM		FF3		
	$\hat{\theta}_{ndur}$	$\hat{\theta}_{ndur.cay}$	$\hat{\theta}_{mkt}$	$\hat{\theta}_{\eta}$	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$
SDF	2.796	-97.169	-0.053	8.124	0.019	0.078	0.115
t-ratio	1.293	-0.461	-0.943	1.649	0.862	2.422	3.510
HJ	0.020		0.231		0.197		
α	0.200	0.200	0.0001		0.0001		

	Cahart				Liquidity				
	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{MOM}$	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{MOM}$	$\hat{\theta}_{LIQ}$
SDF	0.034	0.064	0.126	0.027	0.034	0.065	0.129	0.026	6.529
t-ratio	1.107	1.678	3.542	0.900	1.154	1.895	3.753	0.961	1.596
HJ	0.193				0.186				
α	0.0001				0.0001				

	FF5					Q5				
	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{RMW}$	$\hat{\theta}_{CMA}$	$\hat{\theta}_{mkt}$	$\hat{\theta}_{ME}$	$\hat{\theta}_{I/A}$	$\hat{\theta}_{ROE}$	$\hat{\theta}_{Eg}$
SDF param	0.002	0.095	0.129	-0.045	0.041	0.060	0.070	0.205	-0.120	0.226
t-ratio	0.070	1.420	1.866	-0.566	0.412	1.736	2.346	3.080	-1.670	1.726
HJ	0.055					0.170				
α	0.004					0.0001				

The table presents the estimation results of ten asset pricing models with unscaled factors. The models are estimated using the monthly returns of the 1596 tiny-cap micro portfolios. Data are monthly from July 1973 to June 2018. CCAY model uses quarterly data. We report parameter estimates $\hat{\theta}$, t-ratios under model misspecification, the regularized HJ and the penalization level α.

2.9.2 Asset pricing models for small-cap micro portfolios

Table 2.12 – Estimates and misspecification-robust t-ratios of SDF parameters : small-cap Micro portfolios and unscaled factors

	CCAPM			DCAPM					HCAPM		
	$\hat{\theta}_{ndur}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{ndur}$	$\hat{\theta}_{dur}$			$\hat{\theta}_{hmkt}$	$\hat{\theta}_{prem}$	$\hat{\theta}_{lab}$
SDF	1.799			0.029	0.946	0.473			1.836	9.361	0.238
t-ratio	3.186			1.884	1.507	0.522			1.822	0.183	0.071
HJ	0.073			0.064					0.154		
α	0.0001			0.0001					0.0001		
	CCAY				IAPM				FF3		
	$\hat{\theta}_{ndur}$	$\hat{\theta}_{ndur.cay}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{\eta}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$
SDF	1.836	9.361			0.017	2.133			0.032	0.039	0.050
t-ratio	1.822	0.183			0.477	0.643			1.579	1.047	1.809
HJ	0.154				0.070				0.063		
α	0.0001				0.0001				0.0001		
	Cahart						Liquidity				
	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{MOM}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{MOM}$	$\hat{\theta}_{LIQ}$
SDF	0.058	0.006	0.066	0.045			0.055	0.012	0.061	0.043	-8.057
t-ratio	2.224	0.132	2.474	1.526			1.936	0.251	2.287	1.400	-1.598
HJ	0.055						0.047				
α	0.0001						0.0001				
	FF5						Q5				
	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{RMW}$	$\hat{\theta}_{CMA}$		$\hat{\theta}_{mkt}$	$\hat{\theta}_{ME}$	$\hat{\theta}_{I/A}$	$\hat{\theta}_{ROE}$	$\hat{\theta}_{Eg}$
SDF	0.046	0.052	-0.059	0.065	0.152		0.061	0.060	0.093	0.006	0.131
ts	2.293	1.360	-1.033	1.343	2.039		2.508	1.848	1.888	0.115	1.219
HJ	0.054						0.051				
α	0.0001						0.0001				

The table presents the estimation results of ten asset pricing models with unscaled factors. The models are estimated using the monthly returns of the 441 small-cap micro portfolios. Data are monthly from July 1973 to June 2018. CCAY model uses quarterly data. We report parameter estimates $\hat{\theta}$, t-ratios under model misspecification, the regularized HJ and the penalization level α.

2.9.3 Asset pricing models for big-cap micro portfolios

TABLE 2.13 – Estimates and misspecification-robust t-ratios of SDF parameters : Big-cap Micro portfolios and unscaled factors

	CCAPM			DCAPM					HCAPM		
	$\hat{\theta}_{ndur}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{ndur}$	$\hat{\theta}_{dur}$			$\hat{\theta}_{hmkt}$	$\hat{\theta}_{prem}$	$\hat{\theta}_{lab}$
SDF	1.570			0.050	-1.155	1.177			0.037	0.297	0.485
t-ratio	1.482			2.438	-0.921	1.246			1.947	0.230	0.276
HJ	0.047			0.070					0.078		
α	0.004			0.000					0.0001		
	CCAY				IAPM				FF3		
	$\hat{\theta}_{ndur}$	$\hat{\theta}_{ndur.cay}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{\eta}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$
SDF	-0.699	161.399			-0.058	8.458			0.039	0.050	0.088
t-ratio	-0.326	1.196			-1.302	2.094			2.497	1.176	2.527
HJ	0.187				0.066				0.065		
α	0.0001				0.0001				0.0001		
	Cahart						Liquidity				
	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{MOM}$			$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{MOM}$	$\hat{\theta}_{LIQ}$
SDF	0.077	-0.002	0.112	0.108			0.058	0.060	0.141	0.067	-7.444
t-ratio	2.727	-0.025	2.549	1.636			1.698	0.659	2.174	0.816	-0.631
HJ	0.001						0.000				
α	0.200						0.2				
	FF5						Q5				
	$\hat{\theta}_{mkt}$	$\hat{\theta}_{SMB}$	$\hat{\theta}_{HML}$	$\hat{\theta}_{RMW}$	$\hat{\theta}_{CMA}$		$\hat{\theta}_{mkt}$	$\hat{\theta}_{ME}$	$\hat{\theta}_{I/A}$	$\hat{\theta}_{ROE}$	$\hat{\theta}_{Eg}$
SDF	0.045	0.079	-0.016	0.087	0.120		0.030	0.117	0.188	0.136	-0.112
t-ratio	3.200	2.737	-0.315	2.395	1.907		0.750	2.772	2.033	1.193	-0.379
HJ	0.052						0.007				
α	0.0001						0.004				

The table presents the estimation results of ten asset pricing models with unscaled factors. The models are estimated using the monthly returns of the 428 big-cap micro portfolios. Data are monthly from July 1973 to June 2018. CCAY model uses quarterly data. We report parameter estimates $\hat{\theta}$, t-ratios under model misspecification, the regularized HJ and the penalization level α.

2.9.4 Conditional models of DCAPM, HCAPM, and CARHART

TABLE 2.14 – Conditional models of DCAPM, HCAPM, and CARHART

DCAPM			HCAPM			MOM		
	$\hat{\theta}$	t-ratio		$\hat{\theta}$	t-ratio		$\hat{\theta}$	t-ratio
θ_{mkt}	-0.070	-0.511	θ_{jmkt}	-0.026	-0.216	θ_{mkt}	0.008	0.059
θ_{ndur}	3.883	1.436	θ_{prem}	1.934	1.342	θ_{SMB}	-0.219	-0.900
θ_{dur}	-0.992	-0.306	θ_{lab}	3.189	1.141	θ_{HML}	0.431	3.410
θ_{mkt*MU}	0.190	1.040	θ_{mkt*MU}	0.157	0.589	θ_{MOM}	0.232	2.361
$\theta_{ndur*MU}$	-4.819	-1.572	$\theta_{prem*MU}$	0.860	0.485	θ_{mkt*MU}	0.448	2.155
θ_{dur*MU}	7.019	2.886	θ_{lab*MU}	4.409	1.365	θ_{SMB*MU}	0.104	0.477
θ_{mkt*FU}	-0.134	-0.549	$\theta_{jmkt*FU}$	-0.161	-0.441	θ_{HML*MU}	0.099	0.393
$\theta_{ndur*FU}$	1.436	0.343	$\theta_{prem*FU}$	-2.903	-1.173	θ_{MOM*MU}	-0.022	-0.144
θ_{dur*FU}	-8.291	-1.628	θ_{lab*FU}	-10.323	-1.957	θ_{mkt*FU}	-0.619	-2.120
						θ_{SMB*FU}	0.209	0.403
						θ_{HML*FU}	-0.662	-2.034
						θ_{MOM*FU}	-0.220	-0.965
α			α			α		
0.0001			0.0001			0.0001		
$\hat{\delta}^2_{\alpha}$			$\hat{\delta}^2_{\alpha}$			$\hat{\delta}^2_{\alpha}$		
0.395			0.397			0.367		

The table presents the estimation results of DCAPM, HCAPM, and CARHART with factors scaled by the macroeconomic (MU) and financial uncertainty (FU) indices. The models are estimated using the monthly returns of the 3198 characteristics-sorted and micro portfolios. Data are monthly from July 1973 to June 2018. We report parameter estimates $\hat{\theta}$, t-ratios under model misspecification, the regularized HJ and the penalization level α.

2.10 Appendix D : tables for the robustness analysis

We present the tables related to the robustness analysis.

2.10.1 Analysis with 10 stocks per micro portfolios using the earning-price ratio, investment and return on equity

Table 2.15 presents the pairwise comparisons of asset pricing models using the characteristics-sorted and micro portfolios. The micro portfolios contains 10 stocks and are formed using the earning-price ratio, investment and return on equity.

TABLE 2.15 – Comparison of regularized squared HJ distances : Characteristic-sorted and micro portfolios with factors scaled by macro and financial uncertainty indicators

Panel A : Sample regularized squared HJ distance difference between models

	DCAPM	HCAPM	IAPM	FF3	CAHART	LIQ	FF5	Q5
CCAPM	0.096	0.081	0.032	0.077	0.118	0.124	0.173	0.197
DCAPM		-0.015	-0.064	-0.019	0.022	0.028	0.077	0.101
HCAPM			-0.049	-0.004	0.037	0.043	0.092	0.116
IAPM				0.045	0.086	0.092	0.141	0.165
FF3					0.041	0.047	0.096	0.120
CAHART						0.006	0.055	0.079
LIQ							0.049	0.073
FF5								0.024

Panel B : P-value of the equality test of two regularized square HJ distances

	DCAPM	HCAPM	IAPM	FF3	CAHART	LIQ	FF5	Q5
CCAPM	0.004	0.012	0.170	0.030	0.000	0.000	0.000	0.000
DCAPM		0.651	0.026	0.553	0.534	0.445	0.087	0.016
HCAPM			0.091	0.910	0.336	0.275	0.070	0.014
IAPM				0.084	0.000	0.000	0.001	0.000
FF3					0.065	0.061	0.008	0.007
CAHART						0.519	0.089	0.018
LIQ							0.121	0.026
FF5								0.449

Table 2.16 presents the multiple comparison tests of asset pricing models using the characteristics-sorted and micro portfolios. The test determines whether the benchmark model (each line) performs at least as well as the remaining models. The micro portfolios contains 10 stocks and are formed using the earning-price ratio, investment and return on equity.

TABLE 2.16 – Multiple comparison tests of conditional and unconditional models

Benchmark	q	Unconditional			Conditional		
		$\hat{\delta}_{\alpha}^{2}$	LR	p-Value	$\hat{\delta}_{\alpha}^{2}$	LR	p-Value
CCAPM	5	0.485	10.013	0.000	0.475	14.093	0.000
DCAPM	5	0.459	4.556	0.006	0.385	2.928	0.040
HCAPM	5	0.468	8.556	0.000	0.399	3.029	0.031
IAPM	5	0.468	8.929	0.000	0.449	11.114	0.000
FF3	4	0.444	4.557	0.006	0.403	3.627	0.018
CAHART	5	0.414	2.842	0.044	0.363	2.906	0.056
LIQ	5	0.412	2.451	0.029	0.357	2.580	0.073
FF5	5	0.389	0.755	0.283	0.308	0.287	0.483
q5	5	0.355	0.755	0.283	0.284	0.287	0.224

2.10.2 Analysis with 5 stocks per micro portfolios using the book-to-market ratio, investment and profitability

Table 2.17 presents the pairwise comparisons of asset pricing models using the characteristics-sorted and micro portfolios. The micro portfolios contains 5 stocks and are formed using the book-to-market ratio, investment and profitability.

TABLE 2.17 – Comparison of regularized squared HJ distances : Characteristic-sorted and micro portfolios with factors scaled by macro and financial uncertainty indicators

Panel A : Sample regularized squared HJ distance difference between models

	DCAPM	HCAPM	IAPM	FF3	CAHART	LIQ	FF5	Q5
CCAPM	0.059	0.049	0.019	0.057	0.099	0.100	0.144	0.175
DCAPM		-0.010	-0.041	-0.002	0.041	0.041	0.085	0.116
HCAPM			-0.031	0.009	0.051	0.051	0.095	0.127
IAPM				0.038	0.080	0.080	0.124	0.156
FF3					0.042	0.042	0.087	0.118
CAHART						0.000	0.044	0.076
LIQ							0.044	0.076
FF5								0.031

Panel B : P-value of the equality test of two regularized square HJ distances

	DCAPM	HCAPM	IAPM	FF3	CAHART	LIQ	FF5	Q5
CCAPM	0.030	0.055	0.374	0.085	0.002	0.002	0.001	0.000
DCAPM		0.697	0.129	0.960	0.259	0.258	0.061	0.003
HCAPM			0.208	0.787	0.153	0.153	0.059	0.003
IAPM				0.112	0.001	0.001	0.002	0.000
FF3					0.027	0.026	0.011	0.002
CAHART						0.876	0.178	0.016
LIQ							0.181	0.017
FF5								0.363

Table 2.18 presents the multiple comparison of asset pricing models using the characteristics-sorted and micro portfolios. The test determines whether the benchmark model (each line) performs at least as well as the remaining models. The micro portfolios contains 5 stocks and are formed using the book-to-market ratio, investment and profitability.

Table 2.18 – Multiple comparison tests of conditional and unconditional models

Benchmark	q	Unconditional			Conditional		
		$\hat{\delta}^2_\alpha$	LR	p-Value	$\hat{\delta}^2_\alpha$	LR	p-Value
CCAPM	5	0.449	7.913	0.000	0.446	12.670	0.000
DCAPM	5	0.439	5.673	0.003	0.393	4.422	0.010
HCAPM	5	0.445	7.997	0.000	0.403	4.404	0.009
IAPM	5	0.445	8.417	0.001	0.431	12.564	0.000
FF3	4	0.419	3.754	0.011	0.394	4.600	0.006
CAHART	5	0.391	2.047	0.084	0.352	2.988	0.048
LIQ	5	0.387	1.622	0.070	0.352	2.903	0.051
FF5	5	0.369	0.551	0.342	0.308	0.414	0.386
q5	5	0.341	0.551	0.342	0.276	0.414	0.182

Chapitre 3

Pseudo-true SDF parameters estimation with many assets : a comparison between gross and excess returns *

3.1 Introduction

Researchers in empirical asset pricing vastly use linear models. In an asset pricing model, the linear models assign a free parameter to each factor of the Stochastic Discount factor (SDF), in contrast to nonlinear ones. Therefore, they are easily interpretable. In addition, the ability to linearize virtually any model makes them the workhorse of the empirical literature.

The SDF parameter estimation often relies on the minimization of the Hansen and Jagannathan (1997) distance, which assesses how close the proposed linear SDF is to the theoretical SDF that perfectly prices the set of assets in the financial market. As the models are just an approximation of reality, they are usually considered misspecified, i.e., the model's SDF never equals the true SDF that prices all the assets.

There are two approaches to estimating the SDF parameters depending on the type of returns (gross or excess returns) one employs, as shown in Cochrane (2005). When the test assets are gross returns, the HJ distance is a Generalized Method of Moments (GMM) distance on the pricing errors of the model with a weighting matrix represented

*. I thank Marine Carrasco for her invaluable guidance.

by the inverse of the second-moment matrix of the returns. The risk-free rate is usually among the test assets if available. On the other hand, the use of excess returns is prevalent. In this case, Kan and Robotti (2008) proposed a modification of the initial HJ distance. The new misspecification measure is also a GMM distance of the pricing errors but with the covariance matrix of the excess returns as the weighting matrix. This result comes from the fact that, in the case of excess returns, the mean of the linear SDF is fixed beforehand. In either case, the approach uses a GMM framework. However, as the models are considered misspecified, the inference centers around the SDF parameter that minimizes the GMM distances called the pseudo-true parameter (see Hall and Inoue (2003), Gospodinov et al. (2013), and Antoine et al. (2020)). Finally, these two approaches are generally used in a setting of a small number of test assets.

This paper proposes a novel approach for estimating the SDF parameters in the misspecified asset pricing models (pseudo-true parameters) with many assets. Our main contributions can be summarized as follows. We regularize the Hansen-Jagannathan distance, stabilizing the covariance/second-moment matrix of returns. This regularization relaxes the Fundamental Equation of Asset pricing models and permits an arbitrage-free market with frictions. See Korsaye, Quaini, and Trojani (2019) who introduced the latter as Smart-SDF. In contrast to Carrasco and Nokho (2022) where only excess returns are used as test assets, the approach proposed here applies both gross and excess returns as test assets. We use a mean value expansion of the regularized Hansen-Jagannathan distance, along with the misspecified GMM framework studied by Hall and Inoue (2003), to derive the asymptotic distribution of the SDF estimator as the number of test assets and time points approach infinity. We also discuss the specificities that apply to each type of return. Finally, we document the finite sample properties of the SDF estimator with gross and excess returns. Specifically, we compare the size and power of the test of whether the SDF parameter is null. This test verifies whether a particular factor helps to price the test assets.

Our findings indicate that as the number of assets increases, the estimation of the SDF parameter, achieved by regularizing the inverse of the excess returns covariance matrix, demonstrates superior size control compared to the gross returns. This result stems from the inherent instability of the second-moment matrix of gross returns. Additionally, the risk-free asset's gross return displays minimal variability, leading to pronounced collinearity with the other test assets, which the regularization fails to mitigate.

The remaining of the paper is organized as follows. The next section introduces the

HJ distance in a general setting with gross and excess returns with their regularized versions. Section 3.3 discusses the estimation of the SDF parameter using the regularized HJ distance. It presents the basic assumptions and the asymptotic properties of the sample SDF parameter. Section 3.4 presents the choice of the regularization parameter. Finally, Section 3.5 explores the finite sample properties of the SDF parameter tests, utilizing gross and excess returns as test assets.

3.2 Minimum discrepancy distance for empirical asset pricing models with many assets

Let x_t be a vector of N **gross returns.** We include the risk-free $(x_{rf,t})$ among the assets. We define $q = [1, \dots, 1]^{'}$ as the cost of the N assets. Let m_t be an admissible Stochastic Discount factor (SDF) for the gross returns. $m_t \in \mathcal{M}$, the set of square integrable random variables. The SDF m_t is admissible if $E[m_t x_t] = q$. We suppose a proxy SDF $y_t(\theta)$, where $\theta \in \Theta$ is an unknown parameter and Θ is the parameter space. We consider the following regularized Hansen and Jagannathan (1997) problem :

$$\delta_\alpha^2 = \underset{m \in \mathcal{M}}{inf}\, E\left[(m_t - y_t(\theta))^2\right] + \frac{1}{\alpha} \parallel E[m_t x_t] - q \parallel_{N,\Omega}^2, \tag{3.1}$$

where $\alpha > 0$, $\Omega = \frac{V_{xx}}{N}$, $V_{xx} = E[x_t x_t^{'}]$ and $\parallel l \parallel_{N,\Omega}^2 = \frac{l^{'}\Omega l}{N}$ for any $l \in \mathbb{R}^N$. δ_α^2 evaluates the performance of the proxy SDF y_t with respect to the true SDF m. However, we relax the constraint $E[m_t x_t] = q$ due to measurement errors or market frictions as discussed in Korsaye, Quaini, and Trojani (2019). The regularization parameter α controls the pricing errors $E[m_t x_t] - q$ allowed. Low values of α emphasize on lower pricing errors, while high values of α allow higher pricing errors.

This optimization problem (3.1) is cast in an infinite dimension space. Making use of Borwein and Lewis (1992), we consider the simpler dual problem [2] given by

$$\delta_\alpha^2 \quad = \quad \underset{\lambda \in \mathbb{R}^N}{max}\, E\left[2\frac{x_t^{'}\lambda}{N} y_t(\theta) - \frac{\lambda^{'} x_t x_t^{'} \lambda}{N^2} - \frac{2\lambda^{'} q}{N}\right] - \alpha \parallel \lambda \parallel_{N,\Omega^{-1}}^2 . \tag{3.2}$$

We note $\Omega_\alpha^{-1} = (\Omega^2 + \alpha I_N)^{-1}\Omega$, and $e(\theta) = E[x_t y_t(\theta) - q]$. The Lagrange multiplier is $\lambda_\alpha = \Omega_\alpha^{-1} e(\theta)$. The squared distance between the theoretical and empirical SDF is

2. For ease of presentation, we present the regularity conditions of the dual in section 3.8.

given by

$$\delta_\alpha^2 = \frac{e(\theta)'\Omega_\alpha^{-1}e(\theta)}{N}. \tag{3.3}$$

From (3.3), we can observe that the misspecification measure δ_α^2 is a GMM distance with moments given by the vector of pricing errors $e(\theta)$ and weighting matrix given by the regularized inverse of the second-moment matrix of the gross returns Ω_α^{-1}. As a result, the main effect of the regularization parameter α is to stabilize the weighting matrix.

One can also use **excess returns** to evaluate the distance between the proxy SDF and m. Define $r_{it} = x_{it} - x_{rf,t}$, the difference between each payoff and the risk-free payoff, and r_t, the vector of $\tilde{N} = N - 1$ excess returns. In the case of excess returns, $q = 0_{\tilde{N}}$. The mean of SDF is not identified, and one must specify it. Kan and Robotti (2008) recommend centering the factor in the SDF such that $E[y_t(\theta)] = 1$. As a result, the initial problem requires an additional constraint on the mean of the SDF. As pointed out by Kan and Robotti (2008), the latter leads to the use of the covariance matrix of excess return $V_{rr} = E[\tilde{r}_t\tilde{r}_t'],\ \tilde{r}_t = r_t - E[r_t]$ as a weighting matrix. Therefore, the modified problem is

$$\tilde{\delta}_\alpha^2 = \underset{m\in\mathcal{M},E[m]=1}{inf} E\left[(m_t - y_t(\theta))^2\right] + \frac{1}{\alpha} \parallel E[m_t r_t] \parallel_{\tilde{N},\Sigma}^2,$$

where $\Sigma = \frac{V_{rr}}{\tilde{N}}$. The regularization parameter α plays the same role as in the initial problem (3.1).

The modified problem has been solved in Carrasco and Nokho (2022). The dual is given by

$$\underset{\lambda\in\mathbb{R}^{\tilde{N}}}{max} E\left\{2\frac{\lambda' r_t}{\tilde{N}}y_t(\theta) - \frac{\lambda\tilde{r}_t\tilde{r}_t'\lambda}{\tilde{N}^2}\right\} - \alpha \parallel \lambda \parallel_{\tilde{N},\Sigma^{-1}}^2 . \tag{3.4}$$

The Lagrange multiplier is $\lambda_\alpha = \Sigma_\alpha^{-1}e(\theta)$, where $\Sigma_\alpha^{-1} = (\Sigma^2 + \alpha I_{\tilde{N}})^{-1}\Sigma$. It leads to the following regularized distance

$$\tilde{\delta}_\alpha^2 = \frac{e(\theta)'\Sigma_\alpha^{-1}e(\theta)}{\tilde{N}}. \tag{3.5}$$

$\tilde{\delta}_\alpha^2$ is the regularized modified squared HJ distance. It is also a GMM distance with the weighting matrix given by the covariance matrix of the excess returns.

For the remaining of the paper, we endow $\mathbb{R}^N$ with the norm $\parallel \phi \parallel_N^2 = \frac{\phi_1'\phi_2}{N}$ with associated inner product $< \phi_1, \phi_2 >_N = \frac{\phi_1'\phi_2}{N}$, and $\mathbb{R}^T$ with norm $\parallel v \parallel_T^2 = \frac{v'v}{T}$ generated by inner product $< v_1, v_2 >_T = \frac{v_1^{'} v_2}{T}$.

3.3 Estimation and asymptotic properties of estimators

In this section, we present the estimation of SDF parameters using the regularized HJ distance for gross and excess returns in the framework of linear asset pricing models. We describe the form of the SDF for each type of return as well as the assumptions needed to obtain their distribution when N and T go to ∞.

3.3.1 Gross returns specification

We assume the availability of a vector of K observed factors f_t. When we use **gross returns** as test asset, the linear specification of the proxy SDF is $y_t(\theta) = \theta_0 + \theta_1' f_t$, where $\theta = [\theta_0, \theta_1]$. With this specification, the parameter space Θ is $\mathbb{R}^{K+1}$ and $e(\theta) = D\theta - q$, with $D = E[x_t F_t^{'}]$ and $F_t = \begin{bmatrix} 1 & f_t^{'} \end{bmatrix}^{'}$. In addition,

$$\theta_\alpha = (D^{'}\Omega_\alpha^{-1}D)^{-1}D^{'}\Omega_\alpha^{-1}q$$

is the minimizer of (3.3). So θ_α is the result of a cross-sectional regression of q on D.

To take into account many test assets, we estimate θ as following

$$\hat{\theta}_\alpha = \underset{\theta}{argmin} \underset{\lambda \in \mathbb{R}^N}{max} Q_{N,T}(\theta, \lambda; \alpha)$$

where $Q_{N,T}(\lambda, \theta; \alpha) = \frac{1}{T}\sum_{t=1}^{T} \varphi_{N,t}(\lambda, \theta; \alpha)$ and $\varphi_{N,t}(\lambda, \gamma; \alpha) = 2\frac{x_t^{'}\lambda}{N} y_t(\theta) - \frac{\lambda^{'} x_t x_t^{'} \lambda}{N^2} - \frac{2\lambda^{'} q}{N} - \frac{1}{N}\lambda' \left(\hat{\Omega}_\alpha - \hat{\Omega}\right)\lambda$. $\hat{\theta}_\alpha$ can also be rewritten using a simpler analytic expressions :

$$\hat{\theta}_\alpha \quad = \quad \underset{\theta}{argmin} \frac{e_T(\theta)\hat{\Omega}_\alpha^{-1} e_T(\theta)}{N}, \tag{3.6}$$

where $e_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}(r_t F_t^{'}\theta - q) = \hat{D}\theta - q$ and $\hat{\Omega}_\alpha^{-1} = (\hat{\Omega}^2 + \alpha I_N)^{-1}\hat{\Omega}$.

In this framework, we have one free parameters α which is the penalization of the Lagrange multiplier. Solving (3.6) leads to the following SDF estimator

$$\hat{\theta}_\alpha = (\hat{D}'\hat{\Omega}_\alpha^{-1}\hat{D})^{-1}\hat{D}'\hat{\Omega}_\alpha^{-1}q.$$

The Lagrange multiplier is given by

$$\hat{\lambda}_\alpha = \hat{\Omega}_\alpha^{-1}e_T(\hat{\theta}_\alpha).$$

Without the regularization, we note $\theta_\star$, the unique pseudo-true value that solves

$$\begin{aligned}\theta_\star &= \underset{\theta}{argmin}\ \frac{e(\theta)\Omega^{-1}e(\theta)}{N} \\ &= (D'\Omega^{-1}D)^{-1}D'\Omega^{-1}q\end{aligned}$$

and

$$\lambda_\star = \Omega^{-1}e(\theta_\star).$$

Gospodinov et al. (2013) use the return specification (3.2) to derive the asymptotic distribution of $\hat{\theta}$ when N is finite and $\alpha = 0$. They show that

$$\sqrt{T}(\hat{\theta} - \theta_\star) \xrightarrow{d} N(0_k, V(\hat{\theta})), \tag{3.7}$$

where $V(\hat{\theta}) = \sum_{j=-\infty}^{\infty} E[l_t l_{t+j}']$, with

$$l_t = (D'V_{xx}^{-1}D)^{-1}\left[D'V_{xx}^{-1}e_t(\theta) + \left\{\frac{\partial y_t(\theta_{HJ})}{\partial\theta} - D'V_{xx}^{-1}x_t\right\}u_t\right]$$

$u_t = e(\theta_{HJ})'V_{xx}^{-1}x_t$, and $V_{xx} = E[x_t x_t']$.

We introduce the following assumptions to study the behavior (consistency and distribution) of $\hat{\theta}_\alpha$.

Assumption 7. *Assume that*

(i) $e(\theta) \neq 0,\ \forall\theta \in \mathbb{R}^{K+1}$.

(ii) *The process $\{x_t, f_t\}$ is stationary and strong mixing with mixing coefficients $\alpha(l)$ verifying*

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

for some $\rho > 0$.

(iii) $E[f_{kt}^{4+2\rho}] < \infty$, *for $k = 1, 2, \cdots K$.*

(iv) $E[x_{it}^{4+2\rho}] < c$, *for* $i = 1, 2, \cdots$, *where* c *is a constant.*

(v) $E[\| x_t \|_N^{4+2\rho}] < \infty$.

(vi) $\| \Omega^{-\omega} e \|_N = O(1)$, $\omega \geq 3$.

(vii) $\| \Omega^{-\omega} D_k \|_N = O(1)$, $\omega \geq 3$ *for* $k = 1, \cdots, K+1$. D_k *are the columns of* D.

(viii) $\| D \|_F = O(N)$ *and* $C_D = \frac{D'\Omega^{-1}D}{N} \to C_1$ *when* $N \to \infty$, *where* C_1 *is a full rank matrix of size* $K+1$.

Remark 12. Assumption 1 (i) imposes global misspecification usually observed in asset pricing models. Assumption 1 (iv) is a moment condition on the cross-section average of the squared returns. It implies that Ω is a trace-class matrix and the compactness of the operator Ω. Assumption 1 (v) is a moment restriction on the norm of the gross returns. Assumption 1 (vi) implies that $e \in \mathcal{R}(\Omega)$, i.e $\Omega^{-1}e$ exists ($\omega = 1$). This assumption will also characterize the rate of converge of the bias on e. It implies that $\| (\Omega_\alpha^{-1} - \Omega^{-1})\, e \|_N = O(\alpha^2)$ as $e \in \mathcal{R}(\Omega^3)$.

The following proposition establishes the consistency of $\hat{\theta}_\alpha$ when the number of assets increases to infinity.

Proposition 7. *If Assumption 7 is verified,* $\hat{\theta}_\alpha$ *converges in probability to* $\theta_\star$ *as* N,T, $\alpha^{\frac{1}{2}}T \to \infty$, $\alpha \to 0$.

From a mean value expansion of $\nabla_\theta Q_{N,T}(\hat{\theta}_\alpha, \hat{\lambda}_\alpha)$ around $(\theta_\star, \lambda_\star)$, we have the following

$$\begin{aligned} \hat{\theta}_\alpha - \theta_\star \;\; = \;\; & \left(\frac{\hat{D}'\hat{\Omega}_\alpha^{-1}\hat{D}}{N}\right)^{-1} \left(-\frac{1}{T}\sum_{t=1}^{T} F_t \frac{x_t'\lambda_\star}{N} + \right. \\ & \left. + \hat{D}'\hat{\Omega}_\alpha^{-1}\frac{1}{T}\sum_{t=1}^{T} \frac{x_t y_t(\theta_\star) - q}{N} - \frac{x_t x_t'\lambda_\star}{N^2} - \frac{(\hat{\Omega}_\alpha - \hat{\Omega})\lambda_\star}{N}\right). \end{aligned}$$

To derive the asymptotic distribution of the SDF parameter, we add the following assumptions.

Assumption 8. *Suppose*

1. $0 < \underset{N,T\to\infty}{lim} Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_t \frac{x_t'\lambda_\star}{N}) < \infty$.
2. $0 < \underset{N,T\to\infty}{lim} Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} D'\Omega^{-1}\frac{x_t y_t(\theta_\star)}{N}) < \infty$.

3. $0 < \underset{N,T\to\infty}{lim} Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} D^{'}\Omega^{-1}\frac{x_t x_t^{'}\lambda_\star}{N^2}) < \infty$.

Proposition 8. *Suppose Assumptions 7 and 8 are verified. As $N, T \to \infty$, if $\alpha \to 0$, $\alpha T \to \infty$, and $\alpha^2 T \to 0$, then*

$$\sqrt{T}\left(\hat{\theta}_\alpha - \theta_\star\right) \xrightarrow{d} N(0_K, V(\theta)), \tag{3.8}$$

where $V(\theta) = \underset{N,T\to\infty}{lim} Var\left[\left(\frac{D^{'}\Omega^{-1}D}{N}\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_t\frac{x_t^{'}\lambda_\star}{N} + D^{'}\Omega^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{x_t y_t(\theta_\star)-q}{N} - \frac{x_t x_t^{'}\lambda_\star}{N^2}\right)\right]$.

Remark 13. The previous proposition establishes the distribution of the SDF parameter estimator in a misspecified linear model when the regularized HJ distance with gross returns is used. For this distance, the inverse of the second-moment matrix is the main weighting matrix. The rates of convergence are identical to the one obtained in Carrasco and Nokho (2022).

3.3.2 Excess returns specification

In the case of **excess returns**, a common specification is $y_t(\theta) = 1 - \theta^{'} f_t$, with $E[f_t] = 0$. So, the stochastic discount factor depends on the factor innovations and has a mean equal to 1. The previous specification has been used, for example, in Giglio and Xiu (2021) and Feng et al. (2020). In this specification, $e(\theta) = \mu_r - V_{rf}\theta$, with $\mu_r = E[r_t]$ and $V_{rf} = E[r_t f_t^{'}]$. Furthermore,

$$\tilde{\theta}_\alpha = (V_{rf}^{'}\Sigma_\alpha^{-1}V_{rf})^{-1}V_{rf}\Sigma_\alpha^{-1}\mu_r$$

minimizes (3.5). Similar to the gross returns, we can estimate $\tilde{\theta}$ as below

$$\hat{\tilde{\theta}}_\alpha = \underset{\theta}{argmin}\ \underset{\lambda\in\mathbb{R}^N}{max}\tilde{Q}_{N,T}(\theta, \lambda; \alpha),$$

where $\tilde{Q}_{\tilde{N},T}(\lambda, \theta; \alpha) = \frac{1}{T}\sum_{t=1}^{T}\tilde{\varphi}_{\tilde{N},t}(\lambda, \theta; \alpha)$ and $\tilde{\varphi}_{\tilde{N},t}(\lambda, \gamma; \alpha) = 2\frac{\lambda^{'} r_t}{\tilde{N}} y_t(\theta) - \frac{\lambda \tilde{r}_t \tilde{r}_t^{'}\lambda}{\tilde{N}^2} - \frac{1}{\tilde{N}}\lambda^{\prime}\left(\hat{\Sigma}_\alpha - \hat{\Sigma}\right)\lambda$.

As a result, the estimator of $\tilde{\theta}$ is given by

$$\hat{\tilde{\theta}}_\alpha = (\hat{V}_{rf}^{'}\hat{\Sigma}_\alpha^{-1}\hat{V}_{rf})^{-1}\hat{V}_{rf}\hat{\Sigma}_\alpha^{-1}\hat{\mu}_r,$$

where $\hat{\mu}_r = \frac{1}{T}\sum_{t=1}^{T} r_t$,$\hat{V}_{rf} = \frac{1}{T}\sum_{t=1}^{T} r_t\tilde{f}_t$, and $\hat{\Sigma}_\alpha^{-1} = (\hat{\Sigma}^2 + \alpha I_{\tilde{N}})^{-1}\hat{\Sigma}$.

The parameter α stabilizes the covariance matrix of the excess returns.

The Lagrange multiplier is given by

$$\hat{\hat{\lambda}}_\alpha = \hat{\Sigma}_\alpha^{-1} e_T(\hat{\hat{\theta}}_\alpha).$$

Without the regularization, we note $\tilde{\theta}_\star$, the unique pseudo-true value that solves

$$\begin{aligned}\tilde{\theta}_\star &= \underset{\theta}{argmin}\ \frac{e(\theta)\Sigma^{-1}e(\theta)}{\tilde{N}} \\ &= (V_{rf}^{'}\Sigma^{-1}V_{rf})^{-1}V_{rf}\Sigma^{-1}\mu_r\end{aligned}$$

and

$$\tilde{\lambda}_\star = \Sigma^{-1}e(\theta_\star).$$

We impose the following assumptions similar to Assumptions 7 and 8 to obtain the distribution of $\hat{\hat{\theta}}_\alpha$.

Assumption 9. *Assume that*

1. $\| \Sigma^{-\omega} e \|_{\tilde{N}} = O(1)$, *for* $\omega \geq 3$.
2. $\| \Sigma^{-\omega} V_{rf,k} \|_{\tilde{N}} = O(1)$, $\omega \geq 3$ *for* $k = 1, \cdots, K+1$. $V_{rf,k}$ *are the columns of* V_{rf}.
3. $\| V_{rf} \|_F = O(\tilde{N})$ *and* $C_v = \frac{V_{rf}^{'}\Sigma^{-1}V_{rf}}{\tilde{N}} \to C_2$ *when* $\tilde{N} \to \infty$, *where* C_2 *is a full rank matrix of size* K.

The preceding assumption closely resembles Assumption 3 proposed by Carrasco and Nokho (2022), with the distinction that it concerns V_{rf} instead of β in this context.

Assumption 10. *Suppose*

1. $\underset{N\to\infty}{lim} E[\| r_t \|_{\tilde{N}}^{4+2\rho}] < \infty$ for $\rho > 0$.
2. $0 < \underset{N,T\to\infty}{lim} Var(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t \frac{r_t^{'}\tilde{\lambda}_\star}{\tilde{N}}) < \infty$.
3. $0 < \underset{N,T\to\infty}{lim} Var(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{rf}^{'}\Sigma^{-1} \frac{r_t y_t(\tilde{\theta}_\star)}{\tilde{N}}) < \infty$.
4. $0 < \underset{N,T\to\infty}{lim} Var(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{rf}^{'}\Sigma^{-1} \frac{\tilde{r}_t\tilde{r}_t^{'}\tilde{\lambda}_\star}{\tilde{N}^2}) < \infty$.

Proposition 9. *Suppose Assumption 9 and 10 are verified. As* $\tilde{N}, T \to \infty$, *if* $\alpha \to 0$, $\alpha T \to \infty$, *and* $\alpha^2 T \to 0$, *then*

$$\sqrt{T}\left(\hat{\hat{\theta}}_\alpha - \tilde{\theta}_\star\right) \xrightarrow{d} N(0_K, V_1(\tilde{\theta})), \tag{3.9}$$

where $V_1(\tilde{\theta}) = \underset{N,T\to\infty}{lim} Var\left[\left(\frac{V_{rf}'\Sigma^{-1}V_{rf}}{\tilde{N}}\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t \frac{r_t'\tilde{\lambda}_\star}{\tilde{N}} + V_{rf}'\Sigma^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{r_t y_t(\tilde{\theta_\star})}{\tilde{N}} - \frac{\tilde{r}_t\tilde{r}_t'\tilde{\lambda}_\star}{\tilde{N}^2}\right)\right]$.

Remark 14. This previous result is the counterpart of Proposition 8. Particularly, it uses a stochastic discount factor given by $y_t(\theta) = 1 - \theta' f_t$, with $E[f_t] = 0$. The excess return vector r_t replaces x_t, V_{rf} replaces D and we use the covariance matrix Σ in place of the second moment matrix Ω. So the proof of the proposition is not repeated.

So far, we assumed that $E[f_t] = 0$. However, in the particular context of excess returns, it is frequent to estimate the mean of the factors for the purpose of demeaning them. Keep in mind that subtracting the factors' mean sets the mean of the SDF y_t to 1. We employ the Generalized Method of Moments (GMM) in a misspecified setting studied by Hall and Inoue (2003) to estimate the SDF parameter θ and take into account the estimation of the factors' mean μ_f. We have the following population moment condition

$$E[g_t(\gamma)] = E\begin{bmatrix} f_t - \mu_f \\ r_t(1 - \tilde{f}_t'\theta) \end{bmatrix} = 0_{\tilde{N}+K},$$

where $\mu_f = E[f_t]$ and $\gamma = \begin{bmatrix} \mu_f \\ \theta \end{bmatrix}$. The sample moments are given by $\bar{g}_T(\gamma) = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} f_t - \mu_f \\ r_t(1 - \tilde{f}_t'\theta) \end{bmatrix}$.

From Cochrane (2005, chap. 11), the estimator $\hat{\gamma}_\alpha = \begin{bmatrix} \hat{\mu}_f \\ \hat{\hat{\theta}}_\alpha \end{bmatrix}$ can be written as the solution of the equations

$$A_T\bar{g}_T(\gamma) = 0_{2K,1},$$

where $A_T = \begin{bmatrix} I_K & 0_{K,N} \\ 0_{K,K} & \hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1} \end{bmatrix}$. The preceding equation combines the first order conditions of $\underset{\theta}{min}\ (\hat{\mu}_r - \hat{V}_{rf}\theta)'\hat{\Sigma}_\alpha^{-1}(\hat{\mu}_r - \hat{V}_{rf}\theta)$ and $\underset{\mu_f}{min}\ (\hat{\mu}_f - \mu_f)'(\hat{\mu}_f - \mu_f)$.

A mean value expansion of $\bar{g}_T(\gamma)$ around $\gamma_\star = (\mu_f', \tilde{\theta}_\star')'$ yields the following

$$0 = A_T\bar{g}_T(\hat{\gamma}_\alpha) = A_T\bar{g}_T(\gamma_\star) + A_T G_T.(\hat{\gamma}_\alpha - \gamma_\star),$$

where $G_T = \frac{\partial \bar{g}_T}{\partial \theta'}(\theta_\star) = \begin{bmatrix} -I_K & 0_{K,K} \\ \hat{\mu}_r\theta_\star' & -\hat{V}_{rf} \end{bmatrix}$. This leads to

$$(\hat{\gamma}_\alpha - \gamma_\star) = -(A_T G_T)^{-1}A_T\bar{g}_T(\gamma_\star).$$

The following proposition gives the asymptotic distribution of $\hat{\theta}_\alpha$ in the case of the excess returns when the mean of the factors is estimated.

Proposition 10. *Suppose Assumption 9 and 10 are satisfied. For excess returns, assume* $\mu_f \neq 0$*. As* $\tilde{N}, T \to \infty$*, if* $\alpha \to 0$*,* $\alpha T \to \infty$*, and* $\alpha^2 T \to 0$*, then*

$$\sqrt{T}\left(\hat{\tilde{\theta}}_\alpha - \tilde{\theta}_\star\right) \xrightarrow{d} N(0_K, V_2(\tilde{\theta})), \tag{3.10}$$

where

$$V_2(\tilde{\theta}) = \lim_{N,T\to\infty} Var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_t\right],$$

and $l_t = \theta_\star(y_t(\theta_\star)-1) - C_v^{-1}\frac{V_{rf}'\Sigma^{-1}}{\tilde{N}}(r_t y_t(\theta_\star) - \tilde{r}_t\tilde{u}_t - e(\theta_\star)) - C_v^{-1}\tilde{f}_t\tilde{u}_t$ *and* $\tilde{u}_t = \frac{\tilde{r}_t'\Sigma^{-1}e(\theta_\star)}{\tilde{N}}$.

Remark 15. The previous proposition applies to a misspecified model with excess returns where one estimates the mean of the factors and has many assets. Using a misspecified GMM framework, it introduces an alternative approach distinct from that of Carrasco and Nokho (2022), who rely on the representation of linear asset pricing models in terms of β. The method also differs from the asymptotic analysis of Kan and Robotti (2008), where the authors use a delta method to derive the distribution of the SDF parameter.

3.3.3 Link between the HJ distance with gross returns and the modified version with excess returns

When the gross risk-free rate is available, there is a relation between the HJ distance computed on the gross returns and the modified version on the excess returns first proposed in Kan and Robotti (2008). To see the link, note that $E[x_t y_t(\theta)] = 0$ can be decomposed by the following two equalities :

$$\begin{aligned} E[r_t y_t(\theta)] &= 0 \\ E[x_{rf} y_t(\theta)] &= 1, \end{aligned}$$

where r_t is the vector of excess returns $(x_t - x_{rf})$ and x_{rf} is the gross return of the risk-free asset. The time dependence of the risk-free asset is removed as it does not vary a lot. The pricing errors of using the previous system to make inference on θ using a linear SDF is

$$e(\theta) = E\begin{bmatrix} r_t(\theta_0 - \theta_1 f_t) \\ x_{rf}(\theta_0 - \theta_1 f_t) - 1 \end{bmatrix},$$

and the squared HJ distance is given by

$$\delta^2 = e(\theta)'\tilde{\Omega}^{-1}e(\theta),$$

where $\tilde{\Omega}$ is the second moment matrix of $\begin{bmatrix} x_{rf} & r_t' \end{bmatrix}'$. Lemma 2 of Kan and Robotti (2008) shows that the previous HJ distance (δ^2) using the previous errors is related to the modified squared HJ distance based on excess return ($\tilde{\delta}^2$) as follows

$$\delta^2 = \frac{\tilde{\delta}^2}{x_{rf}}.$$

The squared distance $\tilde{\delta}^2$ is obtained from the SDF $y(\theta_1) = 1 - \theta_1'(f_t - E(f_t))$ and the pricing errors $e(\theta_1) = E[r_t y_t(\theta_1)] = \mu_r - V_{rf}\theta_1$.

In population, minimizing the distance with the gross returns is equivalent to minimizing the modified distance with the excess returns when risk-free asset return is constant. However, in reality, the return of the risk-free asset is not constant over time. Therefore, in a finite sample, the properties of the SDF parameters estimated using the two distances might differ. We investigate the differences in finite samples in 3.5.

3.4 Choice of the regularization parameters

The estimation of SDF parameters is contingent upon the regularization parameter α. We now describe the process of selecting its appropriate value. For a given sample size T, we divide the data in two parts. We use the first part to estimate θ. We choose α that maximizes the out-of-sample R-square R^2_{oos}, which expression depends on the type of return. For the gross returns,

$$R^2_{oos} = 1 - \frac{(\mu_r^o + D^o \frac{\hat{\theta}_{1,\alpha}'}{\hat{\theta}_{0,\alpha}} - \frac{q}{\hat{\theta}_{0,\alpha}})'(\mu_r^o + D^o \frac{\hat{\theta}_{1,\alpha}'}{\hat{\theta}_{0,\alpha}} - \frac{q}{\hat{\theta}_{0,\alpha}})}{\mu_r^{o'}\mu_r^o}$$

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

For the excess returns,

$$R^2_{oos} = 1 - \frac{(\mu_r^o - V_{rf}^o\hat{\tilde{\theta}}_\alpha)'(\mu_r^o - V_{rf}^o\hat{\tilde{\theta}}_\alpha)}{\mu_r^{o'}\mu_r^o}.$$

3.5 Simulations

In this section, we study the finite sample behavior of the t-test issued from the distribution given in the previous propositions by simulating linear asset pricing models.

We simulate a vector of 393 gross returns (392 portfolios gross returns plus the monthly gross risk-free rate) such that the linear asset pricing models are misspecified. The simulations follow the design of Gospodinov et al. (2013). For sake of completeness, we present the approach. We generate gross returns and factors from a multivariate normal distribution with mean μ and covariance V, where $\mu = E\begin{bmatrix} f_t \\ r_t \end{bmatrix} = \begin{bmatrix} \mu_f \\ \mu_r \end{bmatrix}$ and $V = Var\begin{bmatrix} f_t \\ r_t \end{bmatrix} = \begin{bmatrix} V_{ff} & V_{fr} \\ V_{rf} & V_{rr} \end{bmatrix}$.

The parameter of the distribution are calibrated to actual data to reflect the typical characteristic of linear asset pricing models in the literature. We use 392 monthly portfolio returns ranging from January 1964 to December 2019. For the factors, we use two sets. The first set is the three factors of Fama and French : the market return, the small minus big factor (SMB), and the High Minus Low factor (HML). The second set of factors are from the Durable Capital Asset Pricing Model of Yogo (2006) : the gross market return, the log consumption growth of non durable goods, and the log consumption growth of durable goods. The 392 portfolio returns and the Fama-French Factors comes from the Kenneth French's website, whereas the consumption aggregates are from the U.S Bureau of Economic Analysis.

For each model, V is equal to empirical covariance matrix of the data. In addition, the distribution of the SDF parameter given in (3.8) does not depend on the mean of the factors. As a result, we set $\mu_f = 0$.

For each model, we choose μ_r such that restrictions on $\theta_\star$ are verified and $e(\theta) \neq 0_N$, $\forall\theta$. To do so, notice that for an invertible covariance matrix, the pseudo-true SDF parameter is equivalent to

$\theta_\star = (D'\Sigma^{-1}D)^{-1}D'\Sigma^{-1}q$, with $D = E[r_t F_t] = \begin{bmatrix} \mu_r & V_{rf} \end{bmatrix}$. Using the expression of D, and multiplying $\theta_\star$ by $D'\Sigma^{-1}D$, we have the following system

$$\begin{cases} \mu_r'\Sigma^{-1}(\mu_r\theta_0 + V_{rf}\theta_1 - q) = 0 \\ V_{fr}\Sigma^{-1}(\mu_r\theta_0 + V_{rf}\theta_1 - q) = 0_{k-1} \end{cases}.$$

If the model is misspecified, then $e(\theta_\star) = z \neq 0_N$. Then, $\mu_r\theta_0 = q - V_{rf}\theta_1 + z$. As a result, the previous system becomes

$$\begin{cases} z^{'}\Sigma^{-1}q = -z^{'}\Sigma^{-1}z \\ V_{fr}\Sigma^{-1}z = 0_{k-1} \end{cases} . \tag{3.11}$$

To have a misspecified model, one needs to choose an adequate vector z that satisfies (3.11). A convenient choice is $z = \hat{e}$ with $\hat{e} = (\hat{D}^{'}\Sigma^{-1}\hat{D})^{-1}\hat{D}^{'}\Sigma^{-1}q - q$, and $\hat{D} = \begin{bmatrix}\hat{\mu}_r & V_{rf}\end{bmatrix}$. This is because $\hat{D}^{'}\Sigma^{-1}q = 0$. So, it also verifies $\hat{e}^{'}\Sigma^{-1}q = -\hat{e}^{'}\Sigma^{-1}\hat{e}$, and $V_{fr}\Sigma^{-1}\hat{e} = 0_{k-1}$.

So setting $\mu_r = \frac{q - V_{rf}\theta_1 + z}{\theta_0} = \begin{bmatrix} q & V_{rf}\end{bmatrix}\begin{bmatrix}1/\theta_0 \\ \theta_1/\theta_0\end{bmatrix} + \frac{z}{\theta_0} = X\eta + \eta_0\hat{e}$ gives a misspecified model with $\theta_\star$ as solution. To set a plausible value of $\theta_\star$, we can first pick η, by solving the following constraint problem

$$\underset{\eta}{min}\,(\hat{\mu}_r - X\eta)^{'}\Sigma^{-1}(\hat{\mu}_r - X\eta) \quad s.t\; g(\eta) = 0_k,$$

where the constraint $g(\eta) = 0$ is used to set a particular parameter to 0 to verify the size of a test. Let $\eta_1 = \begin{bmatrix}\eta_{1a}^{'} & \eta_{1b}^{'}\end{bmatrix}^{'}$, where η_{1a} is a $k_a \times 1$ vector and η_{1b} is a $k_b \times 1$ vector with $k_a + k_b = k - 1$. We can set $\eta_{1b} = 0$, by choosing $\eta = (X_c^{'}\Sigma^{-1}X_c)^{-1}X_c^{'}\Sigma^{-1}\hat{\mu}_r$, where $X_c = \begin{bmatrix} q & V_{rf,c}\end{bmatrix}$ with $V_{rf,c}$ is the matrix V_{rf} where the last k_b columns are equal to 0_{k_b}. Without the constraint, $\eta = (X^{'}\Sigma^{-1}X)^{-1}X^{'}\Sigma^{-1}\hat{\mu}_r$.

3.5.1 Models with the risk-free asset

Here, we present the finite sample properties of the t-test on the SDF parameters when the test assets include gross return and the risk free rate.

Table 3.1 presents the result of t-test on the SDF parameters using the asymptotic distribution given in (3.7) for each simulated model. This asymptotic distribution is only valid when the covariance matrix Ω is invertible. In our simulation, when $N > T$, we use the generalized inverse of the matrix. The generalized inverse of the matrix remove the null eigenvalues before computing the inverse.

TABLE 3.1 – Empirical size of the Gospodinov et al. (2013) test with 392 portfolios gross returns and the monthly risk-free asset

T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A : Three factors model of Fama and French (1993)									
	θ_{mkt}			θ_{SMB}			θ_{HML}		
150	0.169	0.091	0.019	0.108	0.057	0.012	0.106	0.052	0.01
350	0.116	0.063	0.013	0.12	0.063	0.012	0.157	0.091	0.026
650	0.107	0.051	0.01	0.102	0.051	0.009	0.111	0.055	0.013
Panel C : Durable consumption CAPM of Yogo (2006)									
	θ_{mkt}			θ_{ndur}			θ_{dur}		
150	0.132	0.071	0.014	0.188	0.109	0.032	0.132	0.071	0.014
350	0.174	0.105	0.030	0.698	0.641	0.533	0.174	0.105	0.030
650	0.113	0.059	0.012	0.236	0.153	0.059	0.113	0.059	0.012

For the three factors model of Fama and French (1993), the t-test presents adequate size when T is large ($T = 650$). This shows that for adequate sample size the SDF parameter test of Gospodinov et al. (2013) is well behaved. For smaller sample size, it exhibits distorted size for the market ($T = 150$) and HML ($T = 350$).

The results are worse for the Durable consumption CAPM of Yogo (2006). All the variables present distorted rejection rates. The market and the durable consumption factors present higher rejection when $N > T$. The test on the non durable consumption variable shows large over-rejections for all values of T. As a result, one will tend to falsely declare these factors as essential in pricing the returns.

In table 3.2, we show the finite sample properties of the SDF parameter t-test SDF using gross returns and the Tikhonov regularization given in proposition 8.

TABLE 3.2 – Empirical size of the Tikhonov test under misspecification with 392 portfolios gross returns and the monthly risk-free asset

T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A : Three factors model of Fama and French (1993)									
	θ_{mkt}			θ_{SMB}			θ_{HML}		
150	0.317	0.268	0.197	0.195	0.142	0.073	0.161	0.111	0.054
350	0.34	0.270	0.186	0.185	0.126	0.059	0.138	0.086	0.034
650	0.413	0.330	0.209	0.203	0.136	0.062	0.122	0.074	0.024
Panel B : Linear durable consumption CAPM of Yogo (2006)									
	θ_{mkt}			θ_{ndur}			θ_{dur}		
150	0.325	0.273	0.203	0.092	0.043	0.005	0.089	0.038	0.005
350	0.341	0.278	0.191	0.107	0.053	0.01	0.096	0.049	0.008
650	0.41	0.333	0.208	0.117	0.062	0.013	0.097	0.050	0.009

In general, when applying the Tikhonov regularization to the models, we have a noticeable over-rejection of the t-test. Therefore, the regularization does not bring any value. In fact, the properties are much worse than without regularization. This is chiefly due to the poor estimation of the second moment matrix of gross returns. In the matrix, the gross return of the risk-free asset does not vary a lot and is much closer to 1. Therefore, we have a near perfect collinearity between the risk-free and the remaining test assets, which adds to the collinearity between test assets. In the previous simulations, its effect is completely removed by the generalized inverse. The poor performance of the second moment matrix of returns, as noted by Kan and Zhou (2004, p. 5), is noticeable even when dealing with a low number of assets. In addition, Kan and Robotti (2009) use the covariance matrix, instead of the second moment matrix, for numerical purposes. In our case, the covariance matrix gives numerically the same result.

Instead of using the gross returns, one can resort to the excess returns. Therefore, we can transform the simulated gross returns into excess returns and utilize the modified HJ to conduct inference on the SDF parameters. This approach has the advantage of eliminating the need for the risk-free asset. In addition, the mean of the SDF is fixed to 1 (see Section 3.3.2).

So, in Table 3.3, we present the empirical size of the t-test using the Tikhonov penalization with excess returns (Proposition 9).

TABLE 3.3 – Empirical size of the Tikhonov test under misspecification with 392 excess returns.

T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A : Three factors model of Fama and French (1993)									
	θ_{mkt}			θ_{SMB}			θ_{HML}		
150	0.06	0.029	0.005	0.058	0.028	0.006	0.053	0.022	0.004
350	0.065	0.03	0.005	0.068	0.034	0.007	0.064	0.029	0.005
650	0.068	0.032	0.006	0.083	0.04	0.010	0.072	0.035	0.007
Panel B : Linear durable consumption CAPM of Yogo (2006)									
	θ_{mkt}			θ_{ndur}			θ_{dur}		
150	0.038	0.015	0.002	0.035	0.015	0.002	0.03	0.012	0.001
350	0.043	0.018	0.002	0.061	0.033	0.007	0.071	0.032	0.006
650	0.055	0.023	0.003	0.088	0.051	0.015	0.132	0.068	0.016

The t-test size of the factors in Fama and French (1993) remains below the theoretical level. So the test is slightly conservative. The market and non durable consumption factors in Yogo (2006) present the same characteristic. The size of the test for the durable consumption factors is slightly higher but still remains close to the theoretical value when T is large.

Table 3.4 presents the empirical power of the t-test based on the Tikhonov penalization of the SDF parameter and the excess returns. The test does not exhibit uniform power over all the factors in both models. For the Fama and French model, the rejection rate is very high for the market and the HML factors. However, it is lower for SMB. In the durable CAPM of Yogo (2006), the rejection rate of the market is still high. It is lower for the consumption factors. This result is mainly due to the strength of the factors.

TABLE 3.4 – Empirical power of the Tikhonov test under misspecification with 392 excess returns.

T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A : Three factors model of Fama and French (1993)									
	θ_{mkt}			θ_{SMB}			θ_{HML}		
150	0.389	0.267	0.105	0.067	0.032	0.005	0.314	0.204	0.066
350	0.73	0.602	0.341	0.107	0.055	0.012	0.683	0.538	0.270
650	0.943	0.888	0.702	0.168	0.09	0.021	0.923	0.856	0.627
Panel B : durable CAPM of Yogo (2006)									
	θ_{mkt}			θ_{ndur}			θ_{dur}		
150	0.176	0.091	0.021	0.034	0.014	0.002	0.034	0.013	0.002
350	0.33	0.203	0.059	0.07	0.037	0.009	0.065	0.028	0.006
650	0.485	0.357	0.148	0.11	0.061	0.021	0.118	0.062	0.015

3.5.2 Models without the risk-free asset

In this section, we conduct the same simulations for the gross returns without including the gross risk-free asset. The latter exhibits minimal variation, remains close to 1, and introduces additional correlations between the test assets. Table 3.5 displays the size of the t-test for both models.

Table 3.5 – Empirical size of the Tikhonov test with 392 portfolios gross returns without the monthly risk free

T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A : Three factors model of Fama and French (1993)									
	θ_{mkt}			θ_{SMB}			θ_{HML}		
150	0.072	0.031	0.005	0.081	0.036	0.006	0.092	0.042	0.008
350	0.111	0.056	0.010	0.094	0.045	0.008	0.144	0.082	0.019
650	0.147	0.083	0.019	0.106	0.053	0.011	0.207	0.123	0.040
Panel C : Durable consumption CAPM of Yogo (2006)									
	θ_{mkt}			θ_{ndur}			θ_{dur}		
150	0.02	0.006	0.000	0.033	0.009	0.000	0.027	0.006	0.000
350	0.051	0.019	0.002	0.105	0.047	0.005	0.062	0.028	0.002
650	0.16	0.085	0.014	0.239	0.146	0.028	0.107	0.056	0.011

Compared to Table 3.2, we observe lower rejection rates when we exclude the risk-free asset for the Fama and French model. However, we note a rejection rate higher than the theoretical one for the market and HML factors. In the consumption model by YOGO, we also observe a higher rejection rate for the market and non durable consumption.

3.6 Conclusion

In conclusion, this paper introduces a novel approach to estimating the Stochastic Discount Factor (SDF) parameters in misspecified asset pricing models, commonly called pseudo-true parameters. Leveraging regularization techniques on the Hansen-Jagannathan (HJ) distance, the proposed approach accommodates many test assets, addressing the limitations of existing approaches. In addition, unlike the prior work of Carrasco and Nokho (2022) on excess returns, our method applies to the gross and excess returns.

Typically, the HJ distance equates to a GMM distance with a particular weighing matrix and the pricing errors of the asset pricing model as the moments. When one uses gross returns, this weighting matrix is the inverse of the second-moment matrix, while it is the inverse of the covariance matrix for excess returns. The regularization of the distance equates to the stabilization of the appropriate weighting matrix. This paper explores the varying modeling approaches found in the literature, depending on whether researchers utilize gross or excess returns. Additionally, it establishes the asymptotic distribution of the Stochastic Discount Factor (SDF) estimator in each scenario as the number of assets increases. We also compare the finite sample performance of the t-test using the distribution of the estimated SDF parameter.

Our results show that when the number of assets increases, the SDF parameter estimation through the excess returns, using regularization, presents much better size control than the gross returns. This result emanates from the unstable nature of the second-moment matrix of gross returns. Furthermore, the gross return of the risk-free asset exhibits minimal variability. Consequently, this results in a significant collinearity between the risk-free asset and the remaining test assets that the regularization cannot dampen.

3.7 Appendix A : Notations for the proofs

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 $\| v \|_T^2 = \frac{v' v}{T}$ generated by inner product $< v_1, v_2 >_T = \frac{v_1' v_2}{T}$.
Ω is a trace-class matrix. We note $\{\lambda_j, \phi_j\}$, $j = 1, 2, ..., N$, the non zero eigenvalues and eigenvectors of Ω. The Tikhonov regularization consists of using $\Omega_\alpha^{-1} = (\Omega^2 + \alpha I_N)^{-1}\Omega$ or

$$\Omega_\alpha^{-1}\phi = \sum_{j=1}^{N} \frac{q(\alpha, \lambda_j^2)}{\lambda_j} < \phi, \phi_j >_N \phi_j,$$

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

To obtain the regularized inverse of $\hat{\Omega}_\alpha^{-1}$, it suffices to replace ϕ_j by $\hat{\phi}_j$ and λ_j by $\hat{\lambda}_j$, for $j = 1, \ldots, min(N, T)$.

Similarly, Ω_α denotes the generalized inverse of Ω_α^{-1}.

3.8 Appendix B : Penalized HJ-distance using gross returns in detail

In this section, we present a formal proof of (3.4). We remove the subscript t for ease of presentation. We assume that the set $\mathcal{D} = \{m \in \mathcal{M} :\| E[mx] - q \|_N^2 < \infty\}$ is not empty.

As in Carrasco and Nokho (2022), we define $X = \frac{2x}{N}$, Let $f_y : L^2 \to \mathbb{R}$ be the function defined by $f_y(x) = E[(x-y)^2]$ and $A : L^2 \to \mathbb{R}^N$ be the operator such that $A(m) = E[mX]$. In addition, $g(l) = \frac{N}{4\alpha} \| l - \frac{2q}{N} \|^2_{N,\Omega}$. So the initial problem (3.1) can be rewritten as below

$$\delta^2 = \underset{m\in L^2}{inf} \left\{ f_y(m) + g(A(m)) \right\}.$$

As $\mathcal{D} \neq \emptyset$, strong duality holds, i.e., $(ri\, dom(g)) \bigcap (ri\, A(dom(f_y))) \neq \emptyset$ [3](see Theorem 4.2 of Borwein and Lewis, 1992). Therefore,

$$\delta^2 = -\underset{\lambda\in\mathbb{R}^N}{min} \left\{ f_y^*(-A^*(\lambda)) + g^*(\lambda) \right\}$$

$f_y^*(z) = E[zy + \frac{1}{4}z^2]$, $A^* : R^N \to L^2$ with $A^*(\lambda) = X'\lambda$, and

$$\begin{aligned} g^*(z) &= \underset{w\in R^N}{sup} \left\{ w'z - \frac{N}{4\alpha} \| w - \frac{2q}{N} \|^2_{N,\Omega} \right\} \\ &= \frac{2q'z}{N} + \underset{h\in R^N}{sup} \left\{ h'z - \frac{N}{4\alpha} \| h \|^2_{N,\Omega} \right\} \\ &= \frac{2q'z}{N} + \frac{\alpha}{N} \| z \|^2_{N,\Omega^{-1}} \end{aligned}$$

by using $h = w - \frac{2q}{N}$.

The dual is given by

$$\begin{aligned} \delta^2 &= -\underset{\lambda\in\mathbb{R}^N}{min} \left\{ f_y^*(-A^*(\lambda)) + g^*(\lambda) \right\} \\ &= -\underset{\lambda\in\mathbb{R}^N}{min} \left\{ -2\frac{x'\lambda}{N} + \frac{\lambda'xx'\lambda}{N^2} + \frac{2q'\lambda}{N} + \frac{\alpha}{N} \| \lambda \|^2_{N,\Omega^{-1}} \right\} \\ &= \underset{\lambda\in\mathbb{R}^N}{max} \left\{ 2\frac{x'\lambda}{N} - \frac{\lambda'xx'\lambda}{N^2} - \frac{2q'\lambda}{N} - \frac{\alpha}{N} \| \lambda \|^2_{N,\Omega^{-1}} \right\}. \end{aligned}$$

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

3.9 Appendix C : Proofs

3.9.1 Preliminary lemmas

Lemma 8. *Under assumption 7,*

1. $\| \hat{D} - D \|_N^2 = O_p(\frac{1}{T})$.
2. $\| e_T(\theta) - e(\theta) \|_N^2 = O_p(\frac{1}{T})$ *for all* $\theta \in \mathbb{R}^{K+1}$.

Proof : First, we show that

$$\| \hat{D} - D \|_F^2 = O_p(\frac{N}{T}).$$

We note $\hat{D} - D = \frac{1}{T}\sum_{t=1}^{T} Y_{xF,t}$, where $Y_{xF,t} = x_t F_t^{'} - E[x_t F_t^{'}]$. We have

$$\begin{aligned}
E[\| \hat{D} - D \|_F^2] &= E\left[tr\left((\hat{D} - D)(\hat{D} - D)^{'}\right)\right] \\
&= E\left[tr\left((\hat{D} - D)^{'}(\hat{D} - D)\right)\right] \\
&= E\left[tr\left(\frac{1}{T}\sum_{t=1}^{T} Y_{xF,t}^{'}\right)\left(\frac{1}{T}\sum_{t=1}^{T} Y_{xF,t}\right)\right] \\
&= E\left[tr\left(\frac{1}{T^2}\sum_{t=1}^{T} Y_{xF,t}^{'} Y_{xF,t} + \frac{1}{T^2}\sum_{t\neq t^{'}}^{T} Y_{xF,t}^{'} Y_{xF,t^{'}}\right)\right] \\
&= \frac{1}{T}E\left[tr(Y_{xF,t}^{'} Y_{xF,t})\right] + \frac{2}{T}\sum_{l=1}^{T}(1-\frac{l}{T})E\left[tr(Y_{xF,1}^{'} Y_{xF,1+l})\right].
\end{aligned}$$

$$\begin{aligned}
trE[Y_{xF,t}^{'} Y_{xF,t}] &= trE[F_t x_t^{'} x_t F_t^{'}] - tr(D^{'}D) \\
&= E[F_t^{'} F_t x_t^{'} x_t] - tr(D^{'}D)
\end{aligned}$$

From Cauchy-Schwarz, $| E[F_t^{'} F_t x_t^{'} x_t] | \leq \sqrt{E[\| F_t \|^4] E[\| x_t \|^4]} = O(N)$. Therefore, $\frac{1}{T}E\left[tr(Y_{xF,t}^{'} Y_{xF,t})\right] = O(\frac{N}{T})$.
Using Davydov's inequality (Davydov (1968)),

$$\begin{aligned} trE\left[(Y_{xF,1}^{'}Y_{xF,1+l})\right] &= \sum_{i=1}^{N}\sum_{k=1}^{K+1}E\left[(F_{k1}x_{i1}-E(F_{k1}x_{i1}))\left(F_{k1+l}x_{i1+l}-E(F_{k1}x_{i1+l})\right)\right] \\ &= \sum_{i=1}^{N}\sum_{k=1}^{K+1}cov(F_{k1}x_{i1},F_{k1+l}x_{i1+l}) \\ &\leq 12\sum_{i=1}^{N}\sum_{k=1}^{K+1}\alpha(l)^{\frac{\rho}{2+\rho}}E[(F_{kt}x_{it})^{2+\rho}]^{\frac{2}{2+\rho}}. \end{aligned}$$

As a result,

$$\begin{aligned} \frac{2}{T}\sum_{l=1}^{T}(1-\frac{l}{T})E\left[tr(Y_{xF,1}^{'}Y_{xF,1+l})\right] &\leq \frac{24}{T}\sum_{i=1}^{N}\sum_{k=1}^{K+1}E[(F_{kt}x_{it})^{2+\rho}]^{\frac{2}{2+\rho}}\sum_{l=1}^{T}(1-\frac{l}{T})\alpha(l)^{\frac{\rho}{2+\rho}} \\ &\leq \frac{24}{T}\sum_{i=1}^{N}\sum_{k=1}^{K+1}E[(F_{kt}x_{it})^{2+\rho}]^{\frac{2}{2+\rho}}\sum_{l=1}^{T}l\alpha(l)^{\frac{\rho}{2+\rho}}. \end{aligned}$$

From Assumption **7(iv)**, and Cauchy-Schwarz, $\mid E[(F_{kt}x_{it})^{2+\rho}] \mid\leq [E(F_{kt}^{4+2\rho})]^{\frac{1}{2}}[E(x_{it}^{4+2\rho})]^{\frac{1}{2}} \leq c^{\frac{1}{2}}E[F_{kt}^{4+2\rho}]^{\frac{1}{2}}$.

So,

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

Furthermore, $E[\parallel \hat{D}-D \parallel_F^2] = O(\frac{N}{T})$. In conclusion,

$$\parallel \hat{D}-D \parallel_F^2 = O_p(\frac{N}{T}).$$

For 2., note that for all $\theta \in \mathbb{R}^{K+1}$, $e_T(\theta)-e(\theta) = \frac{1}{T}\sum_{t=1}^{T}Y_{xy,t}$, where $Y_{ry,t} = x_t y_t(\theta) - E(x_t y_t(\theta))$. We use the same approach as in 1. to show that $\parallel e_T(\theta)-e(\theta) \parallel_F^2 = O_p(\frac{N}{T})$.

Lemma 9. *Under Assumption 7 and 9,*

1. *As $N,T\to\infty$, $\alpha T\to\infty$, $\alpha\to 0$, $\parallel \hat{\Omega}_\alpha^{-1}\hat{D}_k - \Omega^{-1}D_k \parallel_N \xrightarrow{p} 0$.*
2. *As $N,T\to\infty$, $\alpha^{\frac{1}{2}}T\to\infty$, $\alpha\to 0$, $\parallel \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_k - \Omega^{-\frac{1}{2}}D_k \parallel_N \xrightarrow{p} 0$.*
3. *As $N,T\to\infty$, $\alpha^{\frac{1}{2}}T\to\infty$, $\alpha\to 0$, $\parallel \hat{\Omega}_\alpha^{-\frac{1}{2}}e_T(\theta) - \Omega^{-\frac{1}{2}}e(\theta) \parallel_N \xrightarrow{p} 0$.*
4. *As $N,T\to\infty$, $\alpha T\to\infty$, $\alpha\to 0$, $\parallel \hat{\Sigma}_\alpha^{-1}\hat{V}_{rf,k} - \Sigma^{-1}V_{rf,k} \parallel_{\tilde{N}} \xrightarrow{p} 0$.*
5. *As $N,T\to\infty$, $\alpha^{\frac{1}{2}}T\to\infty$, $\alpha\to 0$, $\parallel \hat{\Sigma}_\alpha^{-\frac{1}{2}}\hat{V}_{rf,k} - \Sigma^{-\frac{1}{2}}V_{rf,k} \parallel_N \xrightarrow{p} 0$.*

Proof : We follow the proof of Lemma 7 of Carrasco and Nokho (2022) as we have the same conditions. From Lemma 8, we have $\| \hat{D}_k - D_k \|_N^2 = O_p(\frac{1}{T})$, $\| \hat{V}_{rf,k} - V_{rf,k} \|_{\tilde{N}}^2 = O_p(\frac{1}{T})$, and $\| e_T(\theta) - e(\theta) \|_N^2 = O_p(\frac{1}{T})$ for all $\theta \in \mathbb{R}^{K+1}$. In addition, $\| \Omega_\alpha^{-1} \|_N = \underset{\|\phi\|_N \leq 1}{sup} \| \Omega_\alpha^{-1}\phi \|_N \leq \frac{1}{\alpha}$. The result follows.

Lemma 10. *1. Under Assumption 7, when $N, T \to \infty$, $\alpha \to 0$, and $\alpha^{\frac{1}{2}}T \to \infty$, $\frac{1}{N}\hat{D}'\hat{\Omega}_\alpha^{-1}\hat{D} \xrightarrow{P} C_1$.*

2. Under Assumption 9, when $N, T \to \infty$, $\alpha \to 0$, and $\alpha^{\frac{1}{2}}T \to \infty$, $\frac{1}{N}\hat{V}_{rf}'\hat{\Omega}_\alpha^{-1}\hat{V}_{rf} \xrightarrow{P} C_2$.

Proof : For 1), we can rewrite $\frac{1}{N}\hat{D}'\hat{\Omega}_\alpha^{-1}\hat{D}$ as follows

$$\left[\frac{1}{N}\hat{D}_{k_1}\hat{\Omega}_\alpha^{-1}\hat{D}_{k_2}\right]_{k_1,k_2=1,\cdots,K+1} =< \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_{k_1}, \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_{k_2} >_{N;k_1,k_2=1,\cdots,K+1} .$$

We have

$$\begin{aligned} < \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_{k_1}, \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_{k_2} >_N \; = \; & < \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_{k_1} - \Omega^{-\frac{1}{2}}D_{k_1}, \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_{k_2} - \Omega^{-\frac{1}{2}}D_{k_2} >_N + & (3.12)\\ & < \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_{k_1} - \Omega^{-\frac{1}{2}}D_{k_1}, \Omega^{-\frac{1}{2}}D_{k_2} >_N + & (3.13)\\ & < \Omega^{-\frac{1}{2}}D_{k_1}, \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_{k_2} - \Omega^{-\frac{1}{2}}D_{k_2} >_N + & (3.14)\\ & < \Omega^{-\frac{1}{2}}D_{k_1}, \Omega^{-\frac{1}{2}}D_{k_2} >_N - C_{1k_1,k_2} + & (3.15)\\ & C_{1k_1,k_2}. & (3.16) \end{aligned}$$

We have $|(3.12)| \leq \| \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_{k_1} - \Omega^{-\frac{1}{2}}D_{k_1} \|_N \| \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_{k_1} - \Omega^{-\frac{1}{2}}D_{k_1} \|_N \to 0$ as $N, T \to \infty$, and $\alpha^{\frac{1}{2}}T \to \infty$ using Lemma 9 (2).

For (3.13), we have

$$|< \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_{k_1} - \Omega^{-\frac{1}{2}}D_{k_1}, \Omega^{-\frac{1}{2}}D_{k_2} >_N| \leq \| \hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{D}_{k_1} - \Omega^{-\frac{1}{2}}D_{k_1} \|_N \| \Omega^{-\frac{1}{2}}D_{k_2} \|_N \to 0$$

as $N, T \to \infty$, and $\alpha^{\frac{1}{2}}T \to \infty$ using Lemma 9 (2).

The same is true for (3.14).

Finally, using Assumption **7(viii)**, (3.15) goes to 0 as N goes to ∞.

In conclusion, $\frac{\hat{D}'\hat{\Omega}_\alpha^{-1}\hat{D}}{N} \to C_1$ as $N, T \to \infty$, and $\alpha^{\frac{1}{2}}T \to \infty$, and, $\alpha \to 0$.

For 2), we follow the same approach as 1).

3.9.2 Proof of Proposition 7

We use Theorem 2.7 of Newey and McFadden (1994). We need to show that $Q_{T,N}(\theta) \xrightarrow{p} Q(\theta)$, where $Q_{T,N}(\theta) = \frac{e_T(\theta)\hat{\Omega}_\alpha^{-1}e_T(\theta)}{N}$, and $Q(\theta) = \delta(\theta)$ for all $\theta \in \mathbb{R}^{K+1}$.

We have

$$| Q_{T,N}(\theta) - Q(\theta) | \leq | \| \hat{\Omega}_\alpha^{-\frac{1}{2}} e_T(\theta) \|_N^2 - \| \Omega^{-\frac{1}{2}} e(\theta) \|_N^2 | .$$

As, $N, T \to \infty$, $\alpha^{\frac{1}{2}} T \to \infty$,

$$| \| \hat{\Omega}_\alpha^{-\frac{1}{2}} e_T(\theta) \|_N - \| \Omega^{-\frac{1}{2}} e(\theta) \|_N | \leq \| \hat{\Omega}_\alpha^{-\frac{1}{2}} e_T(\theta) - \Omega^{-\frac{1}{2}} e(\theta) \|_N \xrightarrow{P} 0$$

using Lemma 9 (3).

Therefore $| Q_{T,N}(\theta) - Q(\theta) | \xrightarrow{p} 0$ for all $\theta \in \mathbb{R}^{K+1}$. Concavity of $Q_{T,N}(\theta)$ leads to the desired result.

3.9.3 Proof of Proposition 8

To obtain the Mean value expansion, notice that we can obtain $\hat{\theta}_\alpha$ and $\hat{\lambda}_\alpha$ by solving the following problem

$$\underset{\theta}{argmin} \underset{\lambda \in \mathbb{R}^N}{max} Q_{N,T}(\theta, \lambda; \alpha),$$

where $Q_{N,T}(\theta, \lambda; \alpha) = \frac{1}{T} \sum_{t=1}^T \left[2\frac{x_t'\lambda}{N} y_t(\theta) - \frac{\lambda' x_t x_t' \lambda}{N^2} - \frac{2\lambda' q}{N} \right] - \frac{1}{N}\lambda' \left(\hat{\Omega}_\alpha - \hat{\Omega} \right) \lambda$, $\hat{\Omega}_\alpha$ is the generalized inverse of $\hat{\Omega}_\alpha^{-1}$. We note $\pi = \begin{bmatrix} \theta \\ \lambda \end{bmatrix}$. A Mean value expansion of $\frac{\partial Q_{N,T}}{\partial \pi}$ around $\pi_\star = \begin{bmatrix} \theta_\star \\ \lambda_\star \end{bmatrix}$ yields

$$0 = \frac{\partial Q_{N,T}}{\partial \pi}(\hat{\theta}, \hat{\lambda}) = \frac{\partial Q_{N,T}}{\partial \pi}(\theta_\star, \lambda_\star) + \frac{\partial^2 Q_{N,T}(\bar{\theta}, \bar{\lambda})}{\partial \pi \partial \pi'}(\hat{\pi} - \pi_\star).$$

We have $\frac{\partial Q_{N,T}}{\partial \pi}(\theta_\star, \lambda_\star) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^T 2\frac{\lambda_\star' x_t}{N} \frac{\partial y_t(\theta_\star)}{\partial \theta} \\ \frac{1}{T}\sum_{t=1}^T 2\frac{x_t y_t(\theta_\star) - q}{N} - 2\frac{x_t x_t' \lambda_\star}{N^2} - 2\frac{(\hat{\Omega}_\alpha - \hat{\Omega})\lambda_\star}{N} \end{bmatrix}$, and $\frac{\partial^2 Q_{N,T}(\bar{\theta}, \bar{\lambda})}{\partial \pi \partial \pi'} = \begin{bmatrix} I_k & -2\frac{\hat{D}}{N} \\ -2\frac{\hat{D}'}{N} & -2\frac{\hat{\Omega}_\alpha}{N} \end{bmatrix}$. As a result, the first row of $\left[\frac{\partial^2 Q_{N,T}(\bar{\theta}, \bar{\lambda})}{\partial \pi \partial \pi'}\right]^{-1} = \begin{bmatrix} I_k & -2\frac{\hat{D}}{N} \\ -2\frac{\hat{D}'}{N} & -2\frac{\hat{\Omega}_\alpha}{N} \end{bmatrix}^{-1}$ is given by

$$\left[\left(\frac{\hat{D}'\hat{\Omega}_\alpha^{-1}\hat{D}}{N}\right)^{-1} \quad -\left(\frac{\hat{D}'\hat{\Omega}_\alpha^{-1}\hat{D}}{N}\right)^{-1}\hat{D}'\hat{\Omega}_\alpha^{-1}\right],$$

and

$$\begin{aligned}\sqrt{T}\left(\hat{\theta}-\theta_\star\right) &= \left(\frac{\hat{D}'\hat{\Omega}_\alpha^{-1}\hat{D}}{N}\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}-\frac{\lambda_\star' x_t}{N}\frac{\partial y_t(\theta_\star)}{\partial\theta} \\ &- \left(\frac{\hat{D}'\hat{\Omega}_\alpha^{-1}\hat{D}}{N}\right)^{-1}\hat{D}'\hat{\Omega}_\alpha^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{x_t y_t(\theta_\star)-q}{N}-\frac{x_t x_t'\lambda_\star}{N^2}-\frac{\left(\hat{\Omega}_\alpha-\hat{\Omega}\right)\lambda_\star}{N}.\end{aligned}$$

Using the fact that $\frac{\partial Q}{\partial \pi}(\theta_\star,\lambda_\star) = \begin{bmatrix} E\left(\frac{\lambda_\star' x_t}{N}\frac{\partial y_t(\theta_\star)}{\partial\theta}\right) \\ E\left(\frac{x_t y_t(\theta_\star)-q}{N}-2\frac{x_t x_t'\lambda_\star}{N^2}\right)\end{bmatrix} = \begin{bmatrix}0_K\\0_N\end{bmatrix}$ (first order condition of the population dual),

$$\begin{aligned}\sqrt{T}\left(\hat{\theta}-\theta_\star\right) &= \left(\frac{\hat{D}'\hat{\Omega}_\alpha^{-1}\hat{D}}{N}\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}-\frac{F_t x_t'\lambda_\star - E(F_t x_t'\lambda_\star)}{N} \\ &- \left(\frac{\hat{D}'\hat{\Omega}_\alpha^{-1}\hat{D}}{N}\right)^{-1}\hat{D}'\hat{\Omega}_\alpha^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{x_t y_t(\theta_\star)-E(x_t y_t(\theta_\star))}{N}\right. \\ &\left.-\frac{x_t x_t'\lambda_\star - E(x_t x_t'\lambda_\star)}{N^2}-\frac{\left(\hat{\Omega}_\alpha-\hat{\Omega}\right)\lambda_\star}{N}\right). \qquad (3.17)\end{aligned}$$

We obtain the asymptotic distribution of the previous expression by determining the distribution of each component.

As $\alpha\to 0$, $\alpha^{\frac{1}{2}}T\to\infty$, $\frac{\hat{D}'\hat{\Omega}_\alpha^{-1}\hat{D}}{N} - C_D \xrightarrow{p} 0$. In addition, $\forall l\in\mathbb{R}^K$

$$l' C_D \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{F_t x_t'\lambda_\star - E(F_t x_t'\lambda_\star)}{N} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{y_{rf,t}'\lambda_\star}{N},$$

where $y_{rf,t} = l' C_D F_t x_t' - E[l' C_D F_t x_t']$. Using Lemma 5 of Carrasco and Nokho (2022), we have $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{y_{rf,t}'\lambda_\star}{N}$ converges to a normal distribution with variance $\underset{N,T\to\infty}{lim} Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{y_{rf,t}'\lambda_\star}{N})$.

In the same vein,

$$l' C_D D'\Omega^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{x_t y_t(\theta_\star)-E(x_t y_t(\theta_\star))}{N} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{y_{ry,t}' u}{N},$$

where $u = l' C_D D' \Omega^{-1}$ and $y_{ry,t} = r_t y_t(\theta_\star) - E(r_t y_t(\theta_\star))$. Therefore it has an asymptotic normal distribution.

Using Lemma 6 of Carrasco and Nokho (2022)

$$l' C_D D' \Omega^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{x_t x_t' \lambda_\star - E(x_t x_t' \lambda_\star)}{N^2}$$

has also a normal distribution.

Finally,

$$\frac{\sqrt{T}}{N} \hat{D}' \hat{\Omega}_\alpha^{-1} \frac{\left(\hat{\Omega}_\alpha - \hat{\Omega}\right) \Omega^{-1} e(\theta_\star)}{N} \tag{3.18}$$

is the bias. We note $e = e(\theta_\star)$. Notice that

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

So, $\| (\hat{\Omega}_\alpha - \hat{\Omega})\Omega^{-1} e \|_N^2 = \sum_{j/q \neq 0} \hat{\lambda}_j^2 \left(\frac{1 - q(\alpha, \hat{\lambda}_j^2)}{q(\alpha, \hat{\lambda}_j^2)} \right)^2 (\hat{\phi}_j, \Omega^{-1} e)_N^2$. We have

$$\begin{aligned} \sum_{j/q \neq 0} \hat{\lambda}_j^2 \left(\frac{1 - q(\alpha, \hat{\lambda}_j^2)}{q(\alpha, \hat{\lambda}_j^2)} \right)^2 (\hat{\phi}_j, \Omega^{-1} e)_N^2 &\leq \sup_{j/q \neq 0} \frac{1}{q(\alpha, \lambda_j^2)} \sum_j \hat{\lambda}_j^2 \left(q(\alpha, \hat{\lambda}_j^2) - 1 \right)^2 (\hat{\phi}_j, \Omega^{-1} e)_N^2. \\ &\leq \sup_{j/q \neq 0} \frac{\hat{\lambda}_j^2}{q(\alpha, \lambda_j^2)} \sum_j \hat{\lambda}_j^{2\omega} \left(q(\alpha, \hat{\lambda}_j^2) - 1 \right)^2 \frac{(\hat{\phi}_j, \Omega^{-1} e)_N^2}{\hat{\lambda}_j^{2w}}. \end{aligned}$$

We have $\sup_{j/q \neq 0} \frac{\lambda_j^2}{q(\alpha, \lambda_j^2)} = \hat{\lambda}_j^2 + \alpha$ is bounded. On the other hand,

$$\sum_j \hat{\lambda}_j^{2\omega} \left(q(\alpha, \hat{\lambda}_j^2) - 1 \right)^2 \frac{(\hat{\phi}_j, \Omega^{-1} e)_N^2}{\hat{\lambda}_j^{2w}} \leq \sup_j \hat{\lambda}_j^{2\omega} \left(q(\alpha, \hat{\lambda}_j^2) - 1 \right)^2 \sum_j \frac{(\hat{\phi}_j, \Omega^{-1} e)_N^2}{\hat{\lambda}_j^{2w}}$$

is $O_P(\alpha^2)$ if $\| \Omega^{-3} e \|_N^2 < \infty$. Therefore, the squared normed of term (3.18) is $O_p(\alpha^2 T)$ if $\| \Omega^{-3} e(\theta_\star) \|_N^2 < \infty$.

3.9.4 Proof of Proposition 10

We use the following equation

$$(\hat{\gamma}_\alpha - \gamma_\star) = -(A_T G_T)^{-1} A_T \bar{g}_T(\gamma_\star),$$

where $\gamma = \begin{bmatrix} \mu_f \\ \theta \end{bmatrix}$, $A_T = \begin{bmatrix} I_K & 0_{K,N} \\ 0_{K,K} & \frac{1}{\tilde{N}}\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1} \end{bmatrix}$, $\bar{g}_T(\gamma) = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} f_t - \mu_f \\ r_t(1 - \tilde{f}_t'\theta) \end{bmatrix}$, and $G_T = \frac{\partial \bar{g}_T}{\partial \gamma'}(\gamma_\star) = \begin{bmatrix} -I_K & 0_{K,K} \\ \hat{\mu}_r\theta_\star' & -\hat{V}_{rf} \end{bmatrix}$.

We have

$$A_T G_T = \begin{bmatrix} -I_K & 0_{K,K} \\ (\frac{1}{\tilde{N}}\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1}\hat{\mu}_r\theta_\star') & (-\frac{1}{\tilde{N}}\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1}\hat{V}_{rf}) \end{bmatrix}$$

and

$$\begin{aligned} (A_T G_T)^{-1} &= \begin{bmatrix} -I_K & 0_{K,K} \\ -(\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1}\hat{V}_{rf})\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1}\hat{\mu}_r\theta_\star' & -(\frac{\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1}\hat{V}_{rf}}{\tilde{N}})^{-1} \end{bmatrix}. \\ &= \begin{bmatrix} -I_K & 0_{K,K} \\ -\hat{\theta}_\alpha\theta_\star' & -(\frac{\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1}\hat{V}_{rf}}{\tilde{N}})^{-1} \end{bmatrix} \end{aligned}$$

Also, $A_T\bar{g}_T(\theta_\star) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} f_t - \mu_f \\ \frac{1}{\tilde{N}}\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1}\frac{1}{T}\sum_{t=1}^{T} r_t(1 - \tilde{f}_t'\theta_\star) \end{bmatrix}$. The second row can be rewritten as follows

$$\begin{aligned} \frac{1}{\tilde{N}}\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1}e_T(\theta_\star) &= \frac{1}{\tilde{N}}\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1}(e_T(\theta_\star) - e(\theta_\star)) \\ &+ \frac{1}{\tilde{N}}(\hat{V}_{rf}' - V_{rf}')\hat{\Sigma}_\alpha^{-1}e(\theta_\star) \\ &+ \frac{1}{\tilde{N}}V_{rf}'(\hat{\Sigma}_\alpha^{-1} - \Sigma^{-1})e(\theta_\star) \\ &+ \frac{1}{\tilde{N}}V_{rf}'\Sigma^{-1}e(\theta_\star). \end{aligned}$$

Therefore, using the fact that $V_{rf}'\Sigma^{-1}e(\theta_\star) = 0$, we have

$$\begin{aligned}
\sqrt{T}(\hat{\theta}_\alpha - \theta_\star) &= -(\hat{\theta}_\alpha\theta_\star')\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(f_t - \mu_f) \\
&\quad - \left(\frac{\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1}\hat{V}_{rf}}{\tilde{N}}\right)^{-1}\left[\frac{1}{\tilde{N}}\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1}\frac{1}{\sqrt{T}}(r_t y_t - E(r_t y_t))\right. \\
&\quad + \frac{1}{\tilde{N}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{f}_t\tilde{r}_t' - V_{rf})\hat{\Sigma}_\alpha^{-1}e(\theta_\star) \\
&\quad - \frac{1}{\tilde{N}}V_{rf}'\hat{\Sigma}_\alpha^{-1}\sqrt{T}(\hat{\Sigma} - \Sigma)\Sigma^{-1}e(\theta_\star) \\
&\quad \left. - \frac{1}{\tilde{N}}V_{rf}'\hat{\Sigma}_\alpha^{-1}\sqrt{T}(\hat{\Sigma}_\alpha - \hat{\Sigma})\Sigma^{-1}e(\theta_\star)\right].
\end{aligned}$$

Notice that we have the same components as in (3.17), plus the extra term taking into account the estimation of the factor mean. We can use the previous decomposition to prove the asymptotic normally of $\sqrt{T}(\hat{\theta}_\alpha - \theta_\star)$.

We obtain the asymptotic distribution of the previous expression by determining the distribution of each component.

As $N, T \to \infty$,$\alpha \to 0$, $\alpha^{\frac{1}{2}}T \to \infty$, $\frac{\hat{V}_{rf}'\hat{\Sigma}_\alpha^{-1}\hat{V}_{rf}}{\tilde{N}} - C_v \xrightarrow{p} 0$. In addition, $\forall l \in \mathbb{R}^K$

$$l'C_v\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{f_t r_t'\lambda_\star - E(f_t r_t'\lambda)}{\tilde{N}} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{y_{rf,t}'\lambda_\star}{\tilde{N}},$$

where $y_{rf,t} = l'C_v f_t r_t - E[l'C_v f_t r_t]$. Using Lemma 5 of Carrasco and Nokho (2022), we have $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{y_{rf,t}'\lambda_\star}{\tilde{N}}$ converges to a normal distribution with variance $\underset{N,T\to\infty}{lim} Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{y_{rf,t}'\lambda_\star}{\tilde{N}})$.

In the same vein,

$$l'C_v V_{rf}'\Sigma^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{r_t y_t(\theta_\star) - E(r_t y_t(\theta_\star))}{\tilde{N}} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{y_{ry,t}'u}{\tilde{N}},$$

where $u = l'C_v V_{rf}'\Sigma^{-1}$ and $y_{ry,t} = r_t y_t(\theta_\star) - E(r_t y_t(\theta_\star))$. Therefore it has an asymptotic normal distribution.

Using Lemma 6 of Carrasco and Nokho (2022)

$$l'C_v V_{rf}'\Sigma^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\bar{r}_t\bar{r}_t'\lambda_\star - E(\tilde{r}_t\tilde{r}_t'\lambda_\star)}{\tilde{N}^2}$$

has also a normal distribution.

Finally,

$$\frac{\sqrt{T}}{\tilde{N}} \hat{V}_{rf} \hat{\Sigma}_{\alpha}^{-1} \frac{\left(\hat{\Sigma}_{\alpha} - \hat{\Sigma}\right) \Sigma^{-1} e(\theta_{\star})}{\tilde{N}}$$

is the bias. The squared norm of this bias is $O_p(\alpha^2 T)$ if $\| \Sigma^{-3} e(\theta_{\star}) \|_{\tilde{N}}^2 < \infty$. As a result, we conclude the normal distribution when $N, T, \alpha T \to \infty$ and $\alpha, \alpha^2 T \to 0$.

The previous decomposition is equivalent to

$$\sqrt{T}(\hat{\theta}_{\alpha} - \theta_{\star}) - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} l_t = o_p(1),$$

as $N, T, \alpha T \to \infty$ and $\alpha, \alpha^2 T \to 0$, where

$$l_t = -(\theta_{\star} \theta_{\star}^{'}) \tilde{f}_t - C_v \frac{V_{rf}^{'} \Sigma^{-1}}{\tilde{N}} (r_t y_t(\theta_{\star}) - e(\theta_{\star}) - \frac{\tilde{r}_t \tilde{r}_t^{'}}{\tilde{N}} \Sigma^{-1} e(\theta_{\star}) + e(\theta_{\star})) \\ - C_v (\tilde{f}_t \tilde{r}_t \frac{\Sigma^{-1} e(\theta_{\star})}{\tilde{N}} - V_{rf} \frac{\Sigma^{-1} e(\theta_{\star})}{\tilde{N}}).$$

Using the fact that $y_t(\theta) = 1 - \tilde{f}_t^{'} \theta$ and $\tilde{u}_t = \frac{\tilde{r}_t^{'} \Sigma^{-1} e(\theta_{\star})}{\tilde{N}}$, l_t can be simplified as $l_t = \theta_{\star}(y_t(\theta_{\star}) - 1) - C_v \frac{V_{rf}^{'} \Sigma^{-1}}{\tilde{N}} (r_t y_t(\theta_{\star}) - \tilde{r}_t \tilde{u}_t - e(\theta_{\star})) - C_v \tilde{f}_t \tilde{u}_t$.

3.10 Appendix D : List of the Portfolios used in the simulations

TABLE 3.6 – List of portfolios

100 Portfolios Formed on Size and Book-to-Market
100 Portfolios Formed on Size and Operating Profitability
25 Portfolios Formed on Size and Operating Profitability
25 Portfolios Formed on Book-to-Market and Investment
25 Portfolios Formed on Operating Profitability and Investment
25 Portfolios Formed on Size and Momentum
25 Portfolios Formed on Size and Short-Term Reversal
25 Portfolios Formed on Size and Long-Term Reversal
49 Portfolios Formed Industry

Bibliographie

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

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.

Bali, T. G., R. F. Engle, and S. Murray (2016). *Empirical Asset Pricing : The Cross Section of Stock Returns*. Wiley.

Barillas, F., R. Kan, C. Robotti, and J. Shanken (2020). Model Comparison with Sharpe Ratios. *Journal of Financial and Quantitative Analysis 55*(6), 1840–1874.

Barillas, F. and J. Shanken (2017). Which Alpha ? *The Review of Financial Studies 30*(4), 1316–1338.

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

Barras, L. (2019). A large-scale approach for evaluating asset pricing models. *Journal of Financial Economics 134*(3), 549–569.

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

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

Bloom, N. (2014, May). Fluctuations in Uncertainty. *Journal of Economic Perspectives 28*(2), 153–176.

Blume, M. E. (1970). Portfolio Theory : A Step Toward Its Practical Application. *The Journal of Business 43*(2), 152–173. Publisher : University of Chicago 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.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance 52*(1), 57–82.

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*, Volume 6, pp. 5633–5751. Elsevier.

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

Carrasco, M. and C. Nokho (2022). Hansen-Jagannathan distance with many assets.

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

Chernozhukov, V., C. Hansen, Y. Liao, and Y. Zhu (2019). Inference for Heterogeneous Effects using Low-Rank Estimation of Factor Slopes. arXiv :1812.08089 [math, stat].

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

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

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.

Efron, B. (2010). *Large-scale inference : empirical Bayes methods for estimation, testing, and prediction*. Institute of mathematical statistics monographs ; 1. Cambridge ; : Cambridge University Press.

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

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium : Empirical tests. *Journal of political economy 81*(3), 607–636. Publisher : The University of Chicago Press.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the Factor Zoo : A Test of New Factors. *The Journal of Finance 75*(3), 1327–1370.

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.

Giglio, S., Y. Liao, and D. Xiu (2021, July). Thousands of Alpha Tests. *The Review of Financial Studies 34*(7), 3456–3496.

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

Gospodinov, N., R. Kan, and C. Robotti (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.

Hansen, L. P. and S. F. Richard (1987). The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica 55*(3), 587–613.

Hansen, L. P. and K. J. Singleton (1982). Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. *Econometrica 50*(5), 1269–1286.

He, Z., B. Kelly, and A. Manela (2017). Intermediary asset pricing : New evidence from many asset classes. *Journal of Financial Economics 126*(1), 1–35.

He, Z. and A. Krishnamurthy (2013). Intermediary Asset Pricing. *American Economic Review 103*(2), 732–770.

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

Hou, K., H. Mo, C. Xue, and L. Zhang (2021). An Augmented q-Factor Model with Expected Growth. *Review of Finance 25*(1), 1–41.

Jagannathan, R., G. Skoulakis, and Z. Wang (2010). CHAPTER 14 - The Analysis of the Cross-Section of Security Returns. In Y. Aït-Sahalia and L. P. Hansen

(Eds.), *Handbook of Financial Econometrics : Applications*, Volume 2, pp. 73–134. San Diego : Elsevier.

Jagannathan, R. and Z. Wang (1996). The Conditional CAPM and the Cross-Section of Expected Returns. *The Journal of Finance 51*(1), 3–53.

Jegadeesh, N. and S. Titman (1993). Returns to Buying Winners and Selling Losers : Implications for Stock Market Efficiency. *The Journal of Finance 48*(1), 65–91.

Jurado, K., S. C. Ludvigson, and S. Ng (2015, March). Measuring Uncertainty. *American Economic Review 105*(3), 1177–1216.

Kan, R. and C. Robotti (2008). Specification tests of asset pricing models using excess returns. *Journal of Empirical Finance 15*(5), 816–838.

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

Kan, R., C. Robotti, and J. Shanken (2013). Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology. *The Journal of Finance 68*(6), 2617–2649.

Kan, R. and G. Zhou (2004). *Hansen-Jagannathan distance : Geometry and exact distribution.* School of Economics and Finance, University of Hong Kong.

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

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

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.

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

Lettau, M. and S. Ludvigson (2001). Resurrecting the (C)CAPM : A Cross-Sectional Test When Risk Premia Are Time-Varying. *Journal of Political Economy 109*(6), 1238–1287. Publisher : The University of Chicago Press.

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*, Volume 2, pp. 799–906. Elsevier.

Ludvigson, S. C., S. Ma, and S. Ng (2021). Uncertainty and Business Cycles : Exogenous Impulse or Endogenous Response ? *American Economic Journal : Macroeconomics 13*(4), 369–410.

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

Newey, W. K. and D. McFadden (1994). Chapter 36 Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, Volume 4, pp. 2111–2245. Elsevier.

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*(3), 703–708.

Pástor, Ľ. and R. F. Stambaugh (2003). Liquidity Risk and Expected Stock Returns. *Journal of Political Economy 111*(3), 642–685. Publisher : The University of Chicago Press.

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*, Volume 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.