Université de Montréal

Essays in empirical asset pricing

par

Eric Marius Pondi Endengle

Département de sciences économiques

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures et postdoctorales
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en sciences économiques

Aout, 2018

Université de Montréal
Faculté des études supérieures et postdoctorales

Cette thèse intitulée :

Essays in empirical asset pricing

présentée par :

Eric Marius Pondi Endengle

a été évaluée par un jury composé des personnes suivantes :

Marine Carrasco	Président-rapporteur
René Garcia	Directeur de recherche
Benoit Perron	Membre du jury
Nikolay Gospodinov	Examinateur externe
Marine Carrasco	Représentant du doyen de la FAS

Thèse acceptée le 23 Novembre 2018

Résumé

Cette thèse comprend trois chapitres dans lesquels sont développés des outils de comparaison et d'analyse dynamique des modèles linéaires d'évaluation d'actifs.

Dans le premier chapitre, j'introduis la notion de facteurs inutiles par intermittence; il s'agit de facteurs dont la pertinence n'est pas figée dans le temps (parfois utiles, parfois inutiles). Sous ce nouveau cadre théorique, je développe une méthode d'inférence sur les primes de risque. A chaque période, ma méthode permet d'estimer de façon consistante la prime de risque des facteurs utiles, tout en étant robuste à la présence de facteurs inutiles par intermittence. Empiriquement, j'analyse le modèle Fama-French à cinq facteurs. Il apparait qu'à l'exception du marché, tous les facteurs de ce modèle sont inutiles par intermittence, même s'ils demeurent pertinents 90% du temps.

Dans le second chapitre, je développe une méthode d'inférence sur les paramètres dynamiques d'un facteur d'actualisation stochastique (SDF) mal spécifié. J'étends au cadre des SDF conditionnels, l'analyse de Gospodinov, Kan & Robotti (2014); les coefficients et les covariances varient ici dans le temps. Cette nouvelle méthode permet d'éliminer les effets négatifs des facteurs inutiles, et de restaurer la pertinence des facteurs importants, le tout en étant robuste aux erreurs de spécification du modèle. Empiriquement, j'analyse l'évolution de 1963 à 2016, de la pertinence de certains modèles d'évaluation d'actifs. Il apparait que les modèles Fama-French à trois et à cinq facteurs sont les deux meilleurs modèles sur les 50 dernières années. Cependant depuis 2000, le meilleur modèle est le modèle à quatre facteurs de Carhart, suivi du modèle Fama-French à cinq facteurs. Une analyse des modèles possédant des facteurs non échangeables sur les marchés montre que certains de ces facteurs possèdent aussi un pouvoir explicatif sur les rendements observés. La pertinence d'un modèle à capital humain, inspiré de Lettau & Ludvigson (2001) et Gospodinov et al. (2014), est à ce propos mise en évidence.

Le troisième chapitre propose une méthode de classification des modèles Fama-French, en fonction du niveau de préférence des investisseurs pour les moments d'ordre élevé. Les résultats indiquent que l'ajout de facteurs comme stratégie d'amélioration des performances des modèles d'évaluation, n'est efficace que lorsque les investisseurs ont un niveau de préférence assez faible pour les moments d'ordre élevé. Lorsque la préférence pour ces moments devient importante, le modèle à quatre facteurs de Carhart (1997) est plus performant que tous les modèles Fama-French. Les résultats indiquent par ailleurs que le modèle à capital humain analysé dans le deuxième chapitre possède un pouvoir explicatif sur les rendements observés, uniquement pour les investisseurs dont le niveau de préférence pour les moments d'ordre élevé est nul ou très faible.

Mots clés: Facteurs inutiles, Prime de risque, Facteurs d'actualisation stochastique, Modèles mal spécifiés, Distance de Hansen-Jagannathan, Fonctions de divergence, Moments d'ordre élevé.

Abstract

This thesis has three chapters in which I develop tools for comparisons and dynamic analysis of linear asset pricing models.

In the first chapter, I introduce the notion of dynamically useless factors: factors that may be useless (uncorrelated with the assets returns) at some periods of time, while relevant at other periods of time. This notion bridges the literature on classical empirical asset pricing and the literature on useless factors, where both assume that the relevance of a factor remains constant through time. In this new framework, I propose a modified Fama-Macbeth procedure to estimate the time-varying risk premia from conditional linear asset pricing models. At each date, my estimator consistently estimates the conditional risk premium for every useful factor and is robust to the presence of the dynamically useless ones. I apply this methodology to the Fama-French five-factor model and find that, with the exception of the market, all the factors of this model are dynamically useless, although they remain useful 90% of the time.

In the second chapter, I infer the time-varying parameters of a potentially misspecified stochastic discount factor (SDF) model. I extend the model of Gospodinov et al. (2014) to the framework of conditional SDF models, as the coefficients and the covariances are allowed to vary over time. The proposed misspecification-robust inference is able to eliminate the negative effects of potential useless factors, while maintaining the relevance of the useful ones. Empirically, I analyze the dynamical relevance of each factor in seven common asset pricing models from 1963 to 2016. The Fama-French's three-factor model (FF3) and five-factor model (FF5) have been the overall best SDFs in the last 50 years. However, since 2000, the best SDF is CARH (FF3 + momentum factor), followed by FF5 as the second best. Apart from traded factors, the results bring a nuance on non-traded factors. We analyze the relevance, for linear pricing, of a human capital model inspired by Lettau & Ludvigson (2001) and Gospodinov

et al. (2014).

The third chapter proposes a method for ranking Fama-French linear factor models according to investors' preference for higher-order moments. I show that adding a new Fama-French factor to a prior Fama-French model systematically leads to a better model, only when the preference for higher-order moments is moderate (in absolute value). When the preference for higher-order moments is important or extreme, the four-factor model of Carhart (1997) has a better pricing ability than all the Fama-French models. An analysis of models with non-traded factors confirms the relevance, for linear pricing, of the human capital model analyzed in the second chapter. However, I show that this relevance is effective only for investors with null or very low preferences for higher-order moments.

Keywords: Useless factors, Risk premium, Stochastic discount factors, Misspecified models, Hansen-Jagannathan distance, Discrepancy functions, Higher-order moments.

Table des matières

Table of contents

Liste des Tableaux

List of Tables

Liste des Figures

List of Figures

à mes parents, Jeanne et Gabriel

Remerciements

Acknowledgments

J'aimerais tout d'abord remercier mon directeur de recherche, René Garcia, pour la confiance qu'il a placée en moi en acceptant de m'encadrer mais surtout, pour sa disponibilité, son suivi continu et sa gentillesse. Merci René, d'avoir été mon mentor pendant toutes ces années.

Je remercie par ailleurs Ilze Kalnina, pour tout le soutien qu'elle m'a apporté au début de mes études doctorales. Elle est le professeur avec qui j'ai partagé en premier, les réflexions qui auront menées à la rédaction du premier chapitre de cette thèse.

Toute ma gratitude à Marine Carrasco pour sa disponibilité, à Benoit Perron pour ses remarques et conseils, à tous les professeurs du département d'économie de l'Université de Montréal, ainsi qu'à l'ensemble du personnel administratif dudit département.

Mes sincères remerciements à tous les doctorants en économie de l'Université de Montréal, et plus particulièrement à mes camarades de promotion Georges, Hervé, Louphou et Michel, ainsi qu'à Ada et Kokouvi. Merci pour votre présence et vos encouragements. Je remercie par ailleurs Junior et Nery, qui ont été à mes côtés dans les moments de doute.

Je tiens enfin à remercier le Département de sciences économiques de l'Université de Montréal, la Faculté des arts et des sciences, le Centre Inter-universitaire de Recherche en Economie Quantitative (CIREQ) et la Faculté des études supérieures et postdoctorales, pour leur soutien financier et logistique.

Avant-propos

Foreword

Expliquer l'évolution des rendements des actifs financiers est une préoccupation récurrente depuis au moins les années 1960. Sharpe (1964) et Lintner (1965) ont posé les bases de ce qui est aujourd'hui un champ de recherche florissant en asset pricing: l'explication des rendements à partir d'un nombre réduit de facteurs observés et d'un modèle linéaire. Le CAPM a été le premier modèle de ce genre, et a été suivi par une multitude de modèles à facteurs dont les plus populaires sont entre autres: les modèles Fama-French, le CAPM intertemporel, le CAPM conditionnel, le modèle avec momentum de Carhart, etc...

L'une des motivations souvent évoquées pour justifier le développement de nouveaux modèles est l'aspect dynamique des marchés. Un modèle à facteur doit prendre en compte l'évolution de l'information disponible afin d'être le plus précis possible. Cette préoccupation est au centre de cette thèse.

La prise en compte de la dynamique des marchés impose le développement de nouveaux outils d'analyse et d'explication des rendements. Les modèles avec des paramètres constants doivent ainsi faire place à des modèles avec des paramètres qui varient dans le temps. Aussi, la significativité des facteurs ne doit plus être analysée de façon absolue, mais plutôt avec une certaine nuance. En effet, un facteur important hier peut ne plus l'être aujourd'hui, et vice-versa. Dès lors, les questions relatives à l'estimation de primes de risque dynamiques ou encore à l'évaluation dynamique de l'importance des facteurs, deviennent d'un intérêt certain. Ces deux points font l'objet des chapitres 1 et 2.

La spécificité des marchés et celle des agents qui y opèrent est aussi un élément de différenciation des modèles d'asset pricing. Elle est donc de ce fait, une motivation crédible à l'élaboration de modèles à facteurs. Ceux-ci peuvent être dès lors pour la circonstance, opérationnels pour certains marchés et éventuellement moins adaptés pour d'autres. De ce point de vue, il est intéressant de savoir comment évoluent les performances des modèles à facteurs en fonction des caractéristiques des agents qui les appliquent. Nous analysons cette question dans le chapitre 3. Dans ce chapitre, les performances des modèles Fama-French sont évaluées en fonction des niveaux de préférence des agents pour les moments d'ordre élevé.

Les trois articles de cette thèse s'inscrivent dans une logique de développement de nouveaux outils économétriques, permettant de déterminer les conditions sous lesquelles un modèle à facteur donné est plus performant qu'un autre. Les conditions d'analyse considérées reposent sur la dynamique des marchés et l'hétérogénéité des agents. Nous ne proposons pas de nouveau modèle ici, mais présentons une analyse qui devrait permettre de choisir le meilleur modèle à facteurs selon les circonstances prises en compte (différentes périodes d'analyse ou encore différentes fonctions d'utilité). Tous les modèles linéaires à facteurs connus ne sont pas considérés; nous sélectionnons quelques-uns comme modèles de référence dans les trois articles. Néanmoins, les différentes analyses présentées dans cette thèse peuvent être appliquées à tous les modèles linéaires à facteurs utilisés en asset pricing.

Chapter 1

Time-varying risk premia with intermittently useless factors *

1.1 Introduction

Following the initial CAPM model of Sharpe (1964) and Lintner (1965), numerous papers emphasize evidence in favor of time variation in betas (Bollerslev, Engle & Wooldridge, 1988; Jagannathan & Wang, 1996).[1] As stressed by Engle, Lilien & Robins (1987), the time-variation of uncertainty in assets returns leads to a variation of compensation required by risk averse agents, as the payoffs of the risky assets change through time. The time variation in betas is therefore a natural motivation for time variation in risk premia.

When the betas are assumed to be constant, the risk premia are generally estimated through the two-step procedure proposed by Fama & MacBeth (1973). Recently, Gagliardini, Ossola & Scaillet (2016) (hereafter GOS) have developed a new method for inference on the time-varying risk premium, as the classical Fama-MacBeth procedure is not suitable when the betas are assumed to be time-varying. The method proposed by GOS is also a two-step procedure, with the estimation of the matrices that link the betas to the instruments as the first step. As for the Fama-MacBeth procedure, this new method also depends on how good

* I am grateful to René Garcia and Ilze Kalnina for their invaluable guidance. For useful suggestions and comments, I would like to thank Marine Carrasco, Benoit Perron, Kokouvi Tewou and the participants of the CESG 2017.

[1] For additional evidence, see also Ferson & Harvey (1991), Petkova & Zhang (2005) and Ang & Chen (2007).

the betas are estimated. Therefore, having a factor with null or small betas in the model could jeopardize inference abilities in GOS' method and lead to unreliable results for the time-varying risk premium estimation (see Kan & Zhang, 1999b; Kleibergen, 2009, for the link between small-beta factors and the Fama-MacBeth procedure).

The literature associates small betas with useless factors (as opposed to useful factors), defined as factors that are uncorrelated with all the asset returns. Kan & Zhang (1999b) and Kleibergen (2009) show that when the betas are null or when they converge quickly enough to 0, the Fama-MacBeth procedure provides misleading results. As Kleibergen (2009) proposes statistics that are reliable regardless of whether the model includes a useless factor, some papers propose to correct the Fama-MacBeth procedure in the presence of useless factors (see, for example, Bryzgalova, 2016; Kleibergen & Zhan, 2018). However, they all assume that the relevance of factors is constant through time, which means these corrections cannot be applied when the betas are time-varying.

Time-variation in the relevance of factors is a matter of course when we assume that the loadings are time-varying and when we want to estimate time-varying risk premia. As the agents update their beliefs, the returns are continuously updated and the sensitivity of the assets with respect to the risk factors changes accordingly over time. Then, due to phenomena such as crisis, asymmetries, or momenta, it is entirely possible to have a factor that is useful in a given time period and useless in another one. This motivates us to introduce the notion of *dynamically useless factors*: in essence, risk factors that are useless at least in one time period. Contrary to the classical literature on useless factors, where the relevance (or the irrelevance) of every factor is constant through time, this new notion opens up to the possibility of having a factor that is useful in a given time period and useless in another one. Capturing the intermittent uselessness property of a factor is important because it tells us when the factor is relevant and when it is not. Thus, it opens-up the possibility for an investor to build stronger dynamic strategies and correctly estimate the premia that will prevent him from dynamic risks.

We propose a method to infer the time-varying risk premia of factors in the presence of a dynamically useless factor. For that purpose, we adapt the shrinking procedure introduced by Bryzgalova (2016) to a conditional factor model, in order to consistently estimate the time-varying risk premia of factors for periods where they are useful and achieve robustness to the presence of potential dynamically useless factors. In our theoretical framework, the number of assets is large, which means our approach can be applied on large equity data sets, thereby eliminating data snooping (see, for example, Lo & Mackinlay, 1990; White, 2000; Gagliardini et al., 2016).

Concretely, we propose the following modified Fama-Macbeth procedure to estimate the risk premia. In the first step, we estimate the betas at every time period using high frequency data. This step allows us to identify the time periods for which any factor is useless. This in turn allows us to define a penalty function for the second step. The second step uses this penalty function to estimate a semi-parametric dynamic model for risk premia, while also imposing the cross-sectional asset pricing restrictions. Thus, we use two types of data according to their sampling frequency: a low-frequency, e.g. monthly, data set for the estimation of the risk premia, and a high-frequency data set for the estimation of the betas. In a framework with a large number of assets, a large number of periods (months), and a large number of high-frequency observations in every period, our estimator consistently estimates the risk premia of any factor when it is useful, and shrinks its risk premium to zero when it is useless.

In our empirical application, we consider three models: the three-factor model of Fama & French (1993), the four-factor model of Carhart (1997), and the five-factor model of Fama & French (2015). We estimate monthly risk premia by using both daily and monthly data. Our results show that all the factors from the aforementioned models display months where they are not priced, even if they are all found to be relevant. Overall, we find that the principal factors of the Fama-French models are likely dynamically useless, as they all incur times where their risk premia are insignificant (even if they all remain useful at least 90% of the time). Nevertheless, the market return can be considered as an "always useful" factor, as the times where it has an insignificant risk premium are negligible. We also find that none of the five factors of the Fama-French five-factor model is redundant when daily information on factors are taken into account. This result is particularly interesting, as the main argument used against the Fama-French five-factor model is that some factors would become redundant when we add the operating profitability factor (RMW) and the investment factor (CMA) to the original Fama-French three factors. We find some additional information from those new factors. First, the risk premium of the profitability factor RMW (Robust Minus Weak) is never insignificant during recessions. Hence, investors should continue to build strategies on profitability during recessions. Second, the risk premium of the investment factor CMA (Conservative Minus Aggressive) is insignificant mostly at dates where the dynamic of the default spread changes sign. Thus, for agents, the most important consideration regarding investment strategies is not the gap between the top-rated stocks' returns and the bottom-rated stocks' returns; rather, it is the dynamics of that gap. The strategies on investment should be executed as long as that gap maintains a monotonic dynamic.

This paper is related to the literature that studies the effect of useless factors on the Fama-MacBeth estimation of risk premium in the arbitrage pricing theory framework. As Kan & Zhang (1999b) point out, the risk premium estimation by the Fama-MacBeth method is erroneous when there are factors with null betas. Moreover, Kleibergen (2009) shows that in the presence of such factors or when the number of assets is large, linear factor models based on the Fama-MacBeth estimator give misleading results. This finding is confirmed by Kleibergen & Zhan (2015) and Burnside (2016), as they document that when the model includes factors weakly correlated with the assets being priced, the standard estimation methods lead to unreliable risk premia estimation. Likewise, Gospodinov et al. (2014) analyze the effect of misspecification and factor irrelevance on asset-pricing models. They show that the inclusion of factors uncorrelated with the priced assets leads to unreliable statistical inference. Moreover, Gospodinov, Kan & Robotti (2017) show that using optimal and invariant estimators like GMM does not solve the inference problem stressed above, but rather makes estimations worse. All the papers referenced above assume constant betas in their theoretical framework; from this perspective, they are different from this paper.

Under the assumption of constant betas, some papers propose to restore the inference properties of the estimators when the model includes useless factors. Gospodinov et al. (2014) and Feng, Giglio & Xiu (2017) propose a selection procedure which eliminates the useless and the redundant factors from the model and restores the inference properties of the useful ones. Kleibergen (2009) proposes some statistics that are reliable regardless of whether the model includes a useless factor. Gospodinov et al. (2017) characterize the asymptotic properties of the stochastic discount factor parameters when the rank conditions are not satisfied. Bryzgalova (2016) proposes an improvement of the Fama-MacBeth approach by considering a penalized Lasso as second step, and by penalizing according to the nature of the factors. Finally, Giglio & Xiu (2017) propose a three-pass method in order to recover information from omitted factors and obtain valid risk premia estimations.

The rest of the paper is organized as follows. Section 2 provides the motivation behind the importance of considering dynamically useless factors. Section 3 presents the model, the estimation procedures, and the asymptotic properties of our estimator. Section 4 presents the results of the Monte-Carlo simulations. Section 5 presents the results of the empirical applications. Section 6 concludes.

1.2 Dynamically useless factors: motivation and definition

We present here the reasons of considering dynamically useless factors in a conditional estimation of the risk premium.

1.2.1 Empirical motivation

In the arbitrage pricing theory of Ross (1976), the expected return of an asset is explained by a linear combination of macroeconomic variables (risk factors) with their respective betas. One of the main empirical constraints of this theory is that there are no predetermined factors to use, making it difficult to know how many factors are needed and which to consider. Many papers analyze this issue and try to propose a pattern for the choice of the factors that would best explain the returns (see, for example, Fama & French, 1993; Carhart, 1997; Fama & French, 2015)

The difficulty in choosing the right factors becomes more important when we consider that the agents continuously update their beliefs and accordingly, their risk premia. Moreover due to anomalies such as momentum (in stocks) or asymmetries,[2] the set of risk factors that explain the equity returns may change over time. A factor that explains the returns today may not have the same explanatory power tomorrow and vice versa. Let us illustrate this behavior by analyzing the explanatory power of the *default spread* on equity returns (the default spread is proxied by the difference of yields between Moody's Baa-rated and Aaa-rated corporate bonds). For that purpose, we examine the dynamics of the correlation between the

[2]The "leverage effect" is for example a source of asymmetries on the market. A negative shock to an equity market leads to a much more important movement of the volatility than a positive shock. The reader can refer to Black (1976), Engle & Ng (1993) and Bekaert & Wu (2000) for a deeper analysis of the "leverage effect" on markets

default spread and some market portfolios. The default spread has been used as a risk factor in several papers, among which are Jagannathan & Wang (1996), Petkova (2006) and GOS.

Figure 1.1 shows the evolution of the covariances between the default spread and the returns of 6 Fama-French portfolios sorted on size and book-to-market. The covariances are estimated from January 1986 to December 2015 and are updated every 60 days. We see that across time, the default spread displays small covariances with the returns of the portfolios. However, there are periods where some jumps occur. Let us see if this leads to a modification of the explanatory power of the chosen factor on the returns. For that purpose, we perform in every 60 days period, an OLS regression of each portfolio returns on the factor. According to the p-values of the coefficients (see table 1.1), it appears that the default spread generally displays insignificant correlations with each of the 6 portfolios. However, around 2008 (following the financial crisis), these correlations have been significant. In that period, the default spread had an explanatory power on some of the 6 portfolios returns; this was not true for example between 07/08/2008 and 30/10/2008 or between 29/01/2009 and 14/10/2009. This example illustrates our previously introduced concept of a *dynamically useless factor*.[3]

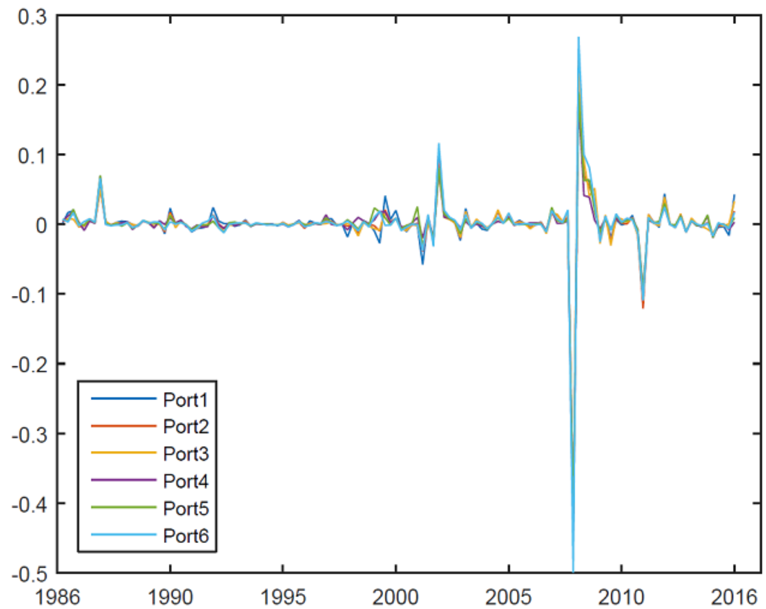
Table 1.1: OLS regressions of 6 Fama-French portfolio returns on the default spread

Date	Port1	Port2	Port3	Port4	Port5	Port6
07/08/2008–30/10/2008	-1.25	-1.13	-1.18	-1.17	-1.16	-1.37
31/10/2008–28/01/2009	1.10^{*}	0.98	1.14	0.86^{**}	0.92	1.31
29/01/2009–24/04/2009	5.21	5.45	7.03	3.16	4.79	7.66
27/04/2009–21/07/2009	0.40	0.35	0.26	0.25	0.41	0.52
22/07/2009-14/10/2009	0.67	0.77	1.12	0.12	0.40	0.66
15/10/2009–11/01/2010	-0.72	-1.02^{**}	-1.21^{*}	-0.29	-0.53	-1.11
12/01/2010–08/04/2010	11.86^{*}	12.78^{*}	14.35	8.89	8.72	12.69

The table presents the coefficients from the ols regressions of 6 Fama-French portfolio returns on the factor *default spread*.
The portfolios are 6 Fama-French portfolios sorted on size and book-to-market We estimate 6 linear models where the default spread is the regressor in each model, and the returns of the portfolio i is the dependent variable in the model i ($i = 1, \dots, 6$). The stars give the degree of significance under the null hypothesis that the coefficient is zero (we use the Bonferroni correction by dividing each significance level by 6, in order to avoid multi-test bias). The factor is useless with regard to the 6 portfolios if all the coefficients are insignificant. The standard errors are estimated using heteroscedasticity and autocorrelation consistent covariance estimators, following Newey & West (1987)
The 6 portfolios are the 6 Fama-French sorted by size and book-to-market; Port1=SMALL/LoBM, Port2=ME1/BM2, Port3=SMALL/HiBM, Port4=BIG/LoBM, Port5=ME2/BM2, Port6=BIG/HiBM.

[3] In order to add some diversification into the portfolios, we follow Lewellen, Nagel & Shanken (2010) and add five industry portfolios to the first six portfolios; we obtain similar results by doing so

Figure 1.1: 60 days covariances between 6 Fama-French portfolio returns and the default spread



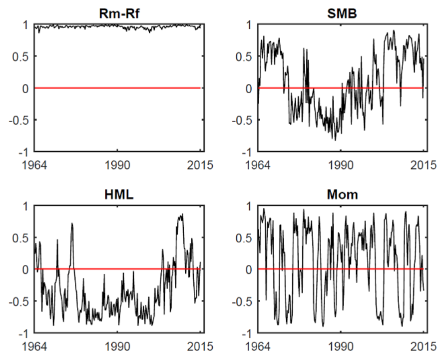
Evolution of the covariances between the default spread and the returns of 6 Fama-French portfolios sorted on size and book-to-market. The covariances are estimated from January 1986 to December 2015 and are updated every 60 days. Port1=SMALL/LoBM, Port2=ME1/BM2, Port3=SMALL/HiBM, Port4=BIG/LoBM, Port5=ME2/BM2, Port6=BIG/HiBM.

The previous illustration helped us to introduce the notion of *dynamically useless factors*. Now, we propose a general illustration of how the set of factors that explains equity returns may change across time. For that purpose, we consider 25 Fama-French portfolios sorted on size and book-to-market and perform a principal component analysis on their daily returns from July 1964 to December 2015. Then, we select the first three principal components and analyze their correlations with the four empirical factors of Carhart (1997) (the Fama-French three factors plus the momentum). Figure 1.2 illustrates the evolution of the correlations between the first two principal components and the selected empirical factors (the correlations are updated every 60 days). We see that the first principal component is highly correlated to the market excess return, and this correlation does not fluctuate. So, the market excess return is always a useful factor when explaining the returns of the chosen 25 Fama-French portfolios. On the other hand, the second principal component displays volatile correlations with all the four empirical factors, even if the correlations with the market excess return and the momentum seem to be lesser than those with the two other factors (SMB and HML).

In order to have more accurate information about how much the factors SMB and HML are linearly linked to the second principal component (regardless of the sign of the correlation), we represent the absolute values of the correlations in figure 1.3. There, we see that the magnitude of the correlation with the size factor SMB is generally higher. Nonetheless, there are some periods where this magnitude is very low (the lowest value is 0.08), and some others where the correlation with the value factor HML is higher (from 1964 to 1965). So, unlike the first principal component, the second one does not clearly have a perfect match with one of the four empirical factors. Thus, if for example we want to build an empirical two-factor model with the explanation of the returns of the selected 25 Fama-French portfolios as goal, the first factor has to be the excess market return and the second one has to be, depending on the period, the size factor SMB, or the value factor HML, or another one that we cannot clearly identify. This means that the second empirical factor should not be the same across time. Therefore, if we build a two-factor model by taking the market excess return as the first factor and a given factor as the second, the second factor could be dynamically useless. This result is confirmed by figure 1.4, which presents the evolution of the adjusted R-squared from the OLS regressions of the principal components on the four selected empirical factors (we perform rolling-window regressions on 60 days windows). The four empirical factors seem to explain quite well the first principal component, as the smallest adjusted R-squared is 0.83. On the other hand, these four empirical factors do not always explain very well the second and the third principal components; the smallest adjusted R-squared is -0.06 for the second and -0.07 for the third. This result tells us that if we are looking for a second and a third empirical factor to explain the dynamics of our 25 portfolio returns from July 1964 to December 2015 (in addition to the market excess return), the other three factors from Carhart (1997) may not be always exhaustive.

Figure 1.2: Correlations between the principal components and some empirical factors

(a) First principal component



(b) Second principal component

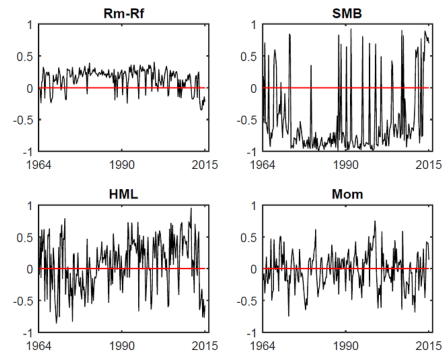
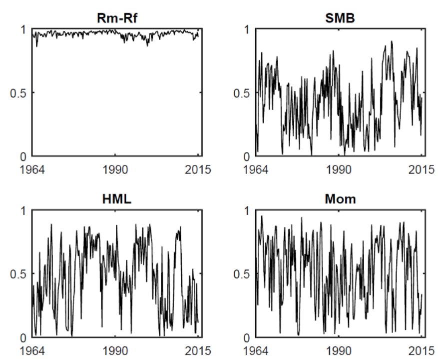


Figure 1.3: Magnitudes of the correlations

(a) First principal component



(b) Second principal component

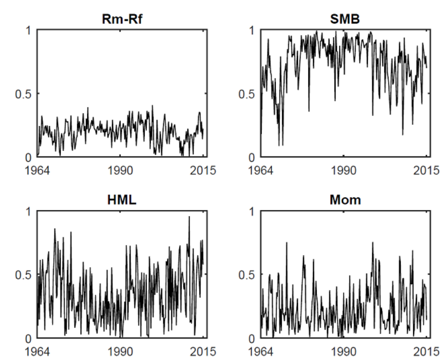
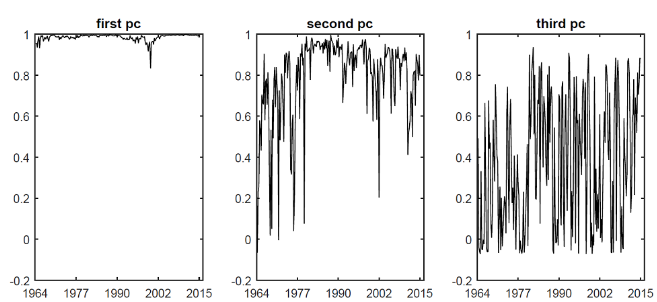


Figure 1.4: Adjusted R-squared from the linear regression of the principal components on the 4 empirical factors of Carhart(1997)



Following the two previous empirical facts, let us examine whether some of the Carhart factors are dynamically useless. We consider again for this purpose the daily returns of the 25 Fama-French portfolios sorted on size and book-to-market, from July 1964 to December 2015. We subdivide the data into several time-periods where the number of days in each period is fixed. Then, we analyze the correlations between the portfolio returns and the factors in each period and determine the number of periods where each factor is useless [4] (see table 1.2). We see that among the four factors, only the Momentum is dynamically useless. Moreover, by varying the number of days in each period, we see that the *probability of being useless* (defined here as the ratio between the number of periods where a factor is useless and the total number of periods) declines as the number of days in each period grows. This result tells us that we have to work with relatively small data periods in order to analyze the "dynamically useless property" of factors. This is precisely the reason why we need high-frequency data.

Table 1.2: Probability of being useless according to the number of days in each period

	20 days (648 periods)	60 days (216 periods)	120 days (108 periods)	252 days (51 periods)	1260 days (10 periods)
Rm-Rf	0	0	0	0	0
SMB	0	0	0	0	0
HML	0	0	0	0	0
Mom	0.64	0.48	0.29	0.27	0

We regress in every period, the daily returns of 25 Fama-French portfolios sorted on size and book-to-market on the four factors. The periods are obtained by grouping the initial data (12965 days) into fixed sized windows. We have a different number of periods following the number of days in each period. For each factor, the probability of being useless is estimated as the number of periods where the factor is useless divided by the total number of periods. A factor is considered as useless in a given period if for all the portfolios,his betas are not significantly different to zero at 5% level of significance. We use the Bonferroni correction to avoid multi-test bias, by taking 0.05/25 as size of the tests.

The illustrations above show that if we consider a linear factor model, it is possible to have a dynamically useless factor among all factors (since all the factors may not be important at the same time). For robustness check, we have added 30 industry portfolios to the initial 25 Fama-French portfolios and have obtained very similar results.

[4] In every period, we estimate the betas of each factor, and we test at the 5% significance level (using a Bonferroni correction to avoid multi-tests bias), if each beta is zero or not. A factor is useless at a given period if all his betas are not significantly different to zero

1.2.2 Formalization and definition

We present here how we formalize the ideas of a useful factor, useless factor and, dynamically useless factor.

Let us consider a factor $f \equiv \{f_t\}_{t=1\dots T}$ with $f_t \in \mathbb{R}$, and let us suppose that the model has n given assets. The beta (sensitivity) of the asset i w.r.t the factor f at time period t is given by b_t^i. Following Kan & Zhang (1999b), the factor f is useless in period t if for all assets i, $b_t^i = 0$. So, if in a given time period the assets' excess returns are uncorrelated with a factor, this factor will be deemed useless at this period. Contrary to the classical (static) definition, we consider in this paper that a factor may be useless at t and not anymore at $t + 1$.

Empirically, the assumption of having a null correlation between returns and risk factors is very strong. Actually, we can always have a very small correlation between at least one return and the factor. Therefore, we extend the definition of a "useless factors" in the sense of Kan & Zhang (1999b) to factors that exhibit small sensitivities to all the assets.[5] Kleibergen (2009) shows that, when such type of factors are included in the factor model, the asymptotic properties of the Fama-MacBeth estimator become unreliable. At each date t, let us assume that the sensitivities are observed through some specific information. If that specific information for date t is collected over D_t periods, then we will say, when the number of periods D_t is large, that a factor has a small correlation with an asset i at date t if $b_t^i = \frac{c_t^i}{\sqrt{D_t}}$ (c_t^i a scalar).

By the formalization presented above, we consider $b_t = \left(b_t^1, \dots, b_t^n\right)'$ the vector of betas at time t, and consider the following definitions:

Definition 1. Assuming that the number of periods D_t is large, a factor f is
(i) *strongly useless at t* if $b_t^i = 0$ for all assets i;
(ii) *weakly useless at t* if $b_t^i = \frac{1}{\sqrt{D_t}} c_t^i$ for all assets i (with c_t^i scalars that are non-null at least for one asset i);
(iii) *useless at t* if it is weakly useless or strongly useless at t.
(iv) *useful at t* if $b_t^i = c_t^i$ for all assets i (with c_t^i scalars that are non-null at least for one asset i)

[5] Following Kleibergen (2009), a loading b is considered here as small if $\hat{b} = o_p(\frac{1}{\sqrt{T}})$, with T the size of the sample from which $\hat{b}$ is estimated.

Following definition 1, it is important to note that useless factors are defined only according to the values of betas. Once the correlation between a factor and one of the priced assets is non-null, that factor is considered as useful regardless of whether its subsequent risk premium is null or not. Hence, useless factors as analyzed in this paper are different from irrelevant factors, which are usually defined as factors with either zero betas or zero risk premia (see for example Gospodinov et al., 2014).

Definition 2. A *dynamically useless factor* is a factor that is useless at least once. So for such a factor, there exists at least one time t_0 such that for all assets i, $b^i_{t_0} = o_p(\frac{1}{\sqrt{D_{t_0}}})$.

By Definition 2 and following the argument presented above (see subsection 1.2.1), some may argue that all the risk factors are eventually dynamically useless. However, we use this definition here assuming that there is at least one factor that is always useful when explaining equity returns (the market return is likely to have this property). Also, Definition 2 also includes factors which are useless in all periods throughout the sample. Therefore, it includes the conventional time-invariant definition of a useless factor.

1.3 The model

The model is presented in this section. Inspired by Kan & Zhang (1999b) and Cosemans, Frehen, Schotman & Bauer (2016), we write the simplest model to explain our theoretical framework. Therefore, we use a two-factor model as we want to consider a model with an always useful factor and a dynamically useless one. However, the results are still valid for the general case with more than 2 factors.

1.3.1 The dynamics

Let $(\Omega, \mathscr{F}^\star, P)$ be a probability space. The flow of information available on the market is represented by the filtration $\mathscr{F}^\star_d, d = 1, 2, \ldots$ (d defined in a high-frequency rate); so we start here with some high-frequency data. We consider that these high-frequency data are grouped in lower frequency samples, let's say at a monthly level. So every month, we have a sub-sample of our high-frequency data. We want to estimate the risk premium every month.

Note that we choose a month-to-month analysis just for simplification. In fact, usefulness of factors are likely to be observed over sub-samples of different sizes. Also, the high-frequency information is assumed to be independent from one month to another. Concretely, we assume that the information used from month to month is not necessarily the same or does not necessarily come from the same data. The only requirement (as we will see below) is the consistency of the high-frequency estimates.

Denote by D_t the number of observations in the sub-sample obtained in month t. For precision, let us suppose that our high-frequency data are sampled at a daily frequency;[6] then D_t is the number of days in the month t. Throughout the paper, the subscript t refers to months and the subscript d to days. For convenience we denote $\mathscr{F}_{t,d}$ the information available at the day d of the month t, and $\mathscr{F}_t$ the information available at the end of the month t. So we have for every month t and for any day d in the chosen month:

$$\begin{aligned}\mathscr{F}_{t,d} &\equiv \mathscr{F}^{\star}_{D_1+\cdots+D_{t-1}+d}\\ \mathscr{F}_t &\equiv \mathscr{F}_{t,D_t}.\end{aligned}$$

There are $K = 2$ factors named us and $\bar{us}$. We assume that the factor us is useful at every period, and the factor $\bar{us}$ is dynamically useless. The objective is to consistently estimate the risk premia of us at every period, and do the same for the factor $\bar{us}$ at periods where it is useful. Throughout the paper, us is considered as factor 1, and $\bar{us}$ as factor 2. There are n assets observed through T months (periods); n and T are both assumed large.

Assumption 1. *The excess return R_t^i of the asset i in month t is defined by:*

$$\begin{aligned}R_t^i &= \phi_t^0 + a_t^i + b_t^{i,us}\cdot f_t^{us} + b_t^{i,\bar{us}}\cdot f_t^{\bar{us}} + \varepsilon_t^i\\ &= \phi_t^0 + a_t^i + (b_t^i)'\cdot f_t + \varepsilon_t^i,\end{aligned} \tag{1.1}$$

where a_t^i, b_t^i, f_t and ε_t^i are random variables; with a_t^i and ε_t^i admitting values in $\mathbb{R}$, whereas $b_t^i = (b_t^{i,1}, b_t^{i,2})'$ and $f_t = (f_t^1, f_t^2)'$ both admit values in $\mathbb{R}^2$. a_t^i and b_t^i are $\mathscr{F}_{t-1}$-measurable. On the other hand, f_t and ε_t^i are $\mathscr{F}_t$-measurable. For any month t, all the processes in model (1.1) are covariance stationary and ergodic. Moreover, $E(\varepsilon_t^i|\mathscr{F}_{t-1}) = 0$ and $Cov(\varepsilon_t^i, f_t|\mathscr{F}_{t-1}) = 0$ for any assets i and for any month t. The pricing error ϕ_t^0 is $\mathscr{F}_{t-1}$-measurable and is included each month in order to consider a potential misspecification of the model.

In our framework, the intercept a_t^i captures measurement errors of the risk-free rate for

[6] We will use daily data in our simulations.

any asset i. We can see it as a difference between the zero-beta rate and the risk free rate. Kan, Robotti & Shanken (2013) consider that this difference appears following a disequilibrium on the risk-free borrowing/lending market in the economy (when the risk-free borrowing rate is different of the risk-free lending rate). Assumption 1 follows the idea of GOS about the dynamics of the excess returns and allows the coefficients to vary over time.

In order to know every month, whether a factor is useless or not and evaluate the corresponding price of risk, we need to estimate the time-varying values of the loadings. In order to do so, we have to fill in the lack of monthly information due to the monthly unobservability.[7] Therefore, we use the daily data to overcome this lack of information and use monthly rolling-window estimations to get the betas. So we build daily frequency models each month, by assuming that the loadings and the intercepts are constant in each of these models. The link between the daily data and the monthly data is set so that the monthly information is captured from the beginning to the last day of the month.

Assumption 2. *Every month t, we have the following daily model for each asset i:*

$$R^i_{t,d} = a^i_t + (b^i_t)' \cdot f_{t,d} + \eta^i_{t,d}; \tag{1.2}$$

where a^i_t, b^i_t, $f_{t,d}$ and $\eta^i_{t,d}$ are random variables; with a^i_t and $\eta^i_{t,d}$ admitting values in $\mathbb{R}$, whereas $b^i_t = (b^{i,1}_t, b^{i,2}_t)'$ and $f_{t,d} = (f^1_{t,d}, f^2_{t,d})'$ both admit values in $\mathbb{R}^2$. a^i_t and b^i_t are assumed constant over any month t and are $\mathscr{F}_{t-1}$-measurable. On the other hand, $f_{t,d}$ and $\eta^i_{t,d}$ are $\mathscr{F}_{t,d}$-measurable. For any month t, $f_{t,d}$ and $\eta^i_{t,d}$ are covariance stationary and ergodic, conditionally to the information available up to the month $t-1$. Moreover, $E(\eta^i_{t,d}|\mathscr{F}_{t-1}) = 0$ and $Cov(\eta^i_{t,d}, f_{t,d}|\mathscr{F}_{t-1}) = 0$ for any assets i and for any month t; also, $E(\eta^i_{t,d}|f_{t,d}) = 0$.

By Assumption 2, we can estimate the betas every month, using a monthly-window estimation. By doing the monthly estimation on daily data, we assume that the daily values of the factors are observed. Now let us consider the following additional assumptions.

Assumption 3. *For any month t, $n^{-1} eig_{max}(\Sigma_{\varepsilon,t}) \overset{L_2}{\to} 0$ when $n \to \infty$; where $\Sigma_{\varepsilon,t} = Var(\varepsilon_t|\mathscr{F}_{t-1})$ is the conditional covariance matrix of the errors $\varepsilon_t = (\varepsilon^1_t, \ldots, \varepsilon^n_t)'$ given $\mathscr{F}_{t-1}$,*

[7] Generally, we can estimate the betas following three procedures: (i) by using rolling sample estimates (Fama & MacBeth, 1973), (ii) by using macroeconomic variables as conditional instruments (Gagliardini et al., 2016), (iii) by mixing the two previous procedures, using prior information on firm characteristics (Cosemans et al., 2016).

and $eig_{max}(\Sigma_{\varepsilon,t})$ its largest eigenvalue.

Assumption 4. *There are no asymptotic arbitrage opportunities in the economy. So there is no portfolio (p_n) such that and $\lim_{n\to\infty} P\left(C(p_n)=0, p_n \geq 0\right)=1$ and $\lim_{n\to\infty} inf P\left(p_n > 0\right) > 0$; (with $C(p_n)$ the cost of the portfolio).*

Assumption 5. *There exist constants $M_a > 0$ and $M_b > 0$ such that for any i, t and d; $\|b_t^i\|_2 \leq M_b$, and $|a_t^i| \leq M_a$*

Assumptions 3 and 4 are Assumptions APR.3 and APR.4 from GOS. Assumption 5 says that the sensitivities and the intercepts are bounded. The boundedness of the sensitivities is a necessary condition for Assumption APR.2(ii) of GOS.

Assumption 6. *Under Assumptions 1-5, if the two factors are useful in month t then there exists a unique $\mathscr{F}_{t-1}$-measurable random variable ν_t in $\mathbb{R}^2$ such that for all assets i,*

$$a_t^i = (b_t^i)' \nu_t.$$

Assumption 6 says that for any month t, for any asset i and under Assumptions 1-4, the difference between the zero-beta rate and the risk free rate (a_t^i) is linearly linked to the sensitivity of assets w.r.t. the useful factors. So with no asymptotic arbitrage opportunities, the non-idiosyncratic parts of the excess returns are fully explained by the systematic risk measured by the betas. Assumption 6 is an asset pricing restriction from which we have the conditional risk premium of GOS. By introducing this restriction into (1.1), we have $E(R_t^i|\mathscr{F}_{t-1}) = \phi_t^0 + (b_t^i)' \left(\nu_t + E(f_t|\mathscr{F}_{t-1})\right)$ in any month t where the two factors are useful. So for those months, the real conditional risk premium is $\lambda_t = \nu_t + E(f_t|\mathscr{F}_{t-1})$.

Assumption 6 is relatively weak, as we show that the equality proposed there is true in the case of a finite number of assets (subsection A3.1 in appendix). Moreover, by Proposition 1 of GOS, we know that this relation is true for a continuum of assets. We make this assumption here as we have a large countable number of assets.

In months where the dynamically useless factor $\bar{us}$ is useless, its risk premium is not identified. In order to estimate the risk premium of the useful factor us in this case, we consider the following assumption.

Assumption 7. *Under assumptions 1-5, if the factor us is the only useful factor in month t, then there exists a unique (up to a $O_p(\frac{1}{\sqrt{T}})$ term) $\mathscr{F}_{t-1}$-measurable random variable ν_t^{us} in $\mathbb{R}$ such that for all assets i:*

$$a_t^i = b_t^{i,us}\nu_t^{us} + O_p(\frac{1}{\sqrt{T}})$$

As for Assumption 6, Assumption 7 gives a linear relation between the intercepts and the loadings of the useful factor. Here because the factor $\bar{us}$ is useless, there is a term which vanished as the number of months T becomes large.

1.3.2 Functional specification

The dynamics of the variables depend on available information. We assume that on the last day of any month t, all the information from the month can be summarized into a unique monthly variable $Z_t \in \mathbb{R}^p$, which is common to all assets. So Z is defined such that for any variable x, $E(x|\mathscr{F}_{t-1,D_{t-1}}) = E(x|\mathscr{F}_{t-1}) = E(x|Z_{t-1})$. Let us now consider the following assumption, which states the link between the dynamics of the factors and the instrument Z.

Assumption 8. *At any period t, $E(f_t|\mathscr{F}_{t-1}) = E(f_t|Z_{t-1}) = FZ_{t-1}$, with $F \in \mathscr{M}(2 \times p)$, the set of matrices with 2 rows and p columns.*

Assumption 8 is Assumption FS.2 of GOS. This assumption gives some restrictions about which kind of instruments we have to choose for Z (see GOS for more details). If the two factors are useful in month t, $\nu_t = \lambda_t - E(f_t|\mathscr{F}_{t-1}) = \lambda_t - FZ_{t-1}$. So we can rewrite the model (1.1) as follows (for the months t where the two factors are useful):

$$R_t^i = \phi_t^0 + (b_t^i)'(\lambda_t - FZ_{t-1} + f_t) + \varepsilon_t^i. \tag{1.3}$$

Let us denote F^j the row j of the matrix F ($j = 1, 2$). On the other hand, if in month t the factor $\bar{us}$ is useless, then $\nu_t^{us} = \lambda_t^{us} - E(f_t^{us}|\mathscr{F}_{t-1}) = \lambda_t^{us} - F^1 Z_{t-1}$; and $\nu_t^{\bar{us}}$ is not identified as T is large. In that particular case as $T \to \infty$, the equation (1.1) becomes:

$$R_t^i = (\phi_t^0 + O_p(\frac{1}{\sqrt{T}})) + b_t^{i,us}(\lambda_t^{us} - F^1 Z_{t-1} + f_t^{us}) + \varepsilon_t^i. \tag{1.4}$$

Let us consider at each period t the variable ϕ_t such that $\phi_t = \lambda_t - FZ_{t-1} + f_t = (\lambda_t^{us} - F^1 Z_{t-1} + f_t^{us},\ \lambda_t^{\bar{us}} - F^2 Z_{t-1} + f_t^{\bar{us}})'$ if the two factors are useful at t, and $\phi_t =$

$(\lambda_t^{us} - F^1 Z_{t-1} + f_t^{us},\ 0)'$ if the factor $\bar{us}$ is useless at t. Then if the factor $\bar{us}$ is useless in month t, the equation (1.1) can take the following formulation (as $T \to \infty$);

$$R_t^i = (\phi_t^0 + O_p(\frac{1}{\sqrt{T}})) + (b_t^i)'\phi_t + \varepsilon_t^i. \tag{1.5}$$

When the two factors are useful at t, the $O_p(\frac{1}{\sqrt{T}})$ term is equal to 0, following Assumption 6. In that particular case, we have as given in equation (1.3): $R_t^i = \phi_t^0 + (b_t^i)'\phi_t + \varepsilon_t^i$.

1.3.3 Risk premium with a dynamically useless factor

We propose a two-step estimation procedure. For that purpose, we define some additional quantities. Let us consider at any month t, the $(n \times 2)$ matrix of betas $B_t = (b_t^1, \cdots, b_t^n)'$ and denote B^j the j^{th} column of B. We also consider the quantities

$$\|B_t^{us}\| = \|B_t^1\| = \underset{i=1,\ldots,n}{sup} |b_t^{i,us}| \qquad \|B_t^{\bar{us}}\| = \|B_t^2\| = \underset{i=1,\ldots,n}{sup} |b_t^{i,\bar{us}}|.$$

As we assume that the number of assets is large, we need to define the spurious properties for the case when n is going to the infinity. For that purpose, we consider the following assumptions.

Assumption 9. *There exist $\alpha \in [0.25; 0.5)$, $\rho_1 > 0$ and $\rho_{2,t} > 1$ at each month t, such that:*
(i) $\frac{n^{2\alpha}}{T} \to \rho_1$ as $T \to \infty$,
(ii) $\frac{D_t}{T} \to \rho_{2,t}$ as $D_t \to \infty$.

Assumption 10. *When the number of assets n is going to infinity,*
(i) the factor $\bar{us}$ is strongly useless at t if $\|B_t^{\bar{us}}\| = \underset{i}{sup}|b_t^{i,\bar{us}}| = 0$,
(ii) the factor $\bar{us}$ is weakly useless at t if $\|B_t^{\bar{us}}\| = \underset{i}{sup}|b_t^{i,\bar{us}}| = \frac{c_t^B}{\sqrt{T}}$ (with c_t^B a non-null scalar),
(iii) The factor $\bar{us}$ is useful at t if $\|B_t^{\bar{us}}\| = c_t^A$ (with c_t^A a non-null scalar).

Assumption 9 links the parameters n, T and D_t. The three parameters become large when n is large. Assumption 10 states that, asymptotically, the spurious properties are defined according to the infinity-norms of the vectors of betas. The idea here is the same as when the number of assets is finite. Now let us present the two steps of our estimation procedure.

First step

We estimate the betas every month from the model (1.2). So we use a discrete-time model sampled at an increasing frequency over the months. This approach is similar to the one proposed by Ang & Kristensen (2012), except that we consider here that the betas are constant in each period.

$\hat{b}_t^i = \left(\frac{1}{D_t}\sum_d \left(f_{t,d} - \bar{f}_t\right)\left(f_{t,d} - \bar{f}_t\right)'\right)^{-1}\left(\frac{1}{D_t}\sum_d \left(f_{t,d} - \bar{f}_t\right)\left(R_{t,d}^i - \bar{R}_t^i\right)\right)$ is the estimated sensitivity of the factor to the asset i (the elements with the bar represent the means on the daily data[8]). We replace the real betas in the model (1.5) by the estimated ones. So we have the following feasible monthly model:

$$R_t^i = \phi_t^0 + (\hat{b}_t^i)'\phi_t + \varepsilon_t^i. \tag{1.6}$$

Second step

We propose a shrinkage estimator for ϕ_t whose second component takes the value 0 in months where the factor $\bar{us}$ is useless. So we introduce from the model (1.6) the following Lasso-modified estimator with $\Phi_t = (\phi_t^0, \phi_t^{us}, \phi_t^{\bar{us}})' = (\phi_t^0, (\phi_t)')'$, $\tau_n > 0$ and $s > 2$ two tuning parameters.

$$\hat{\Phi}_t = \begin{pmatrix}\hat{\phi}_t^0 \\ \hat{\phi}_t\end{pmatrix} = \underset{\Phi}{argmin}\left(\sum_{i=1}^{n}\left(R_t^i - (1\vdots(\hat{b}_t^i)')\Phi\right)^2 + \frac{\tau_n}{T^{s/2}}\sum_{j=1}^{2}\frac{|\phi^j|}{\|\hat{B}_t^j\|^s}\right). \tag{1.7}$$

This estimator uses the shrinkage mechanism proposed by Bryzgalova (2016). The driving force of the penalty term relies on the nature of the factor; so every month t where the factor $\bar{us}$ is useless, we have $\hat{\Phi}_t = (\hat{\phi}_t^0, \hat{\phi}_t^{us}, 0)'$.

Since $E(f_t|\mathscr{F}_{t-1}) = FZ_{t-1}$, we consider the model $f_t = FZ_{t-1} + u_t$, with u_t the idiosyncratic error such that $E(u_t|\mathscr{F}_{t-1}) = 0$. We can then estimate F trough a SURE regression as $\hat{F} = \left(\frac{1}{T}\sum_t f_t Z_{t-1}'\right)\left(\frac{1}{T}\sum_t Z_{t-1}Z_{t-1}'\right)^{-1}$. The same way, by considering the model $f_t^{us} = F^1 Z_{t-1} + u_t^{us}$ (for the months where the factor $\bar{us}$ is useless), we can estimate $\hat{F}^1$ through an OLS regression as $\hat{F}^1 = \left(\frac{1}{T}\sum_t f_t^{us} Z_{t-1}'\right)\left(\frac{1}{T}\sum_t Z_{t-1}Z_{t-1}'\right)^{-1}$.

[8] We have $\bar{f}_t = \frac{1}{D_t}\sum_d f_{t,d}$ and $\bar{R}_t^i = \frac{1}{D_t}\sum_d R_{t,d}^i$.

From the previous and following the definition of ϕ_t, our estimated risk premium in month t is given as following:

$$\begin{cases} \hat{\lambda}_t = \hat{\phi}_t + \hat{F} Z_{t-1} - f_t & \text{if } \hat{\phi}_t^{\bar{us}} \neq 0 \\ \hat{\lambda}_t = \begin{pmatrix} \hat{\lambda}_t^{us} \\ \hat{\lambda}_t^{\bar{us}} \end{pmatrix} = \begin{pmatrix} \hat{\phi}_t^{us} + \hat{F}^1 Z_{t-1} - f_t^{us} \\ 0 \end{pmatrix} & \text{if } \hat{\phi}_t^{\bar{us}} = 0 \end{cases}$$

As the months where the dynamically useless factor $\bar{us}$ is useless are identified through the value of $\hat{\phi}_t^{\bar{us}}$, our estimated risk premium in month t can be summarized as bellow:

$$\begin{cases} \hat{\lambda}_t & = \hat{\phi}_t + \hat{F} Z_{t-1} - f_t & \text{if the two factors are useful at } t \\ \hat{\lambda}_t^{us} & = \hat{\phi}_t^{us} + \hat{F}^1 Z_{t-1} - f_t^{us} & \text{if the factor } \bar{us} \text{ is useless at } t. \\ \hat{\lambda}_t^{\bar{us}} & = 0 & \text{if the factor } \bar{us} \text{ is useless at } t. \end{cases} \tag{1.8}$$

When the factor $\bar{us}$ is useless, its risk premium is not identified. We aim to consistently estimate the risk premium of the factor us at every month t, and to consistently estimate the risk premium of the dynamically useless factor $\bar{us}$ in months where this factor is useful.

Equation (1.8) shows how our estimator is different from the one proposed by Bryzgalova (2016). Indeed, we have an additional term to ϕ_t, which is the difference between the value of the factor and his conditional expected value $\left(E(f_t|\mathscr{F}_{t-1}) - f_t = F Z_{t-1} - f_t\right)$. This additional term is not null, as f_t is $\mathscr{F}_t$-measurable.

1.3.4 Asymptotic properties

We present here the asymptotic properties of the time-varying risk premium estimator. For that purpose, we use some additional assumptions

Assumption 11. *For any given month t, we have as $n \to \infty$*

(i) $\frac{1}{n}\sum_i b_t^i \varepsilon_t^i \xrightarrow{p} 0$, $\frac{1}{n}\sum_i b_t^i (b_t^i)' \xrightarrow{p} Q_t$ *and* $\frac{1}{n}\sum_i b_t^i \xrightarrow{p} q_t$; *with* Q_t *and* $\tilde{Q}_t = \begin{pmatrix} 1 & (q_t)' \\ q_t & Q_t \end{pmatrix}$ *two non singular finite matrices .*

(ii) $\frac{1}{n}\sum_i b_t^{i,us} (b_t^{i,us})' \xrightarrow{p} Q_t^{us}$ *and* $\frac{1}{n}\sum_i b_t^{i,us} \xrightarrow{p} q_t^{us}$; *with* Q_t^{us} *and* $\tilde{Q}_t^{us} = \begin{pmatrix} 1 & q_t^{us} \\ q_t^{us} & Q_t^{us} \end{pmatrix}$ *two non singular finite matrices.*

(iii) $\frac{1}{n}\sum_i \varepsilon_t^i \xrightarrow{p} 0$, *and* $\frac{1}{n}\sum_i \left(\varepsilon_t^i\right)^2 \xrightarrow{p} \sigma_t^2$; *with* σ_t^2 *a finite scalar.*

(iv) $\frac{1}{\sqrt{n}} \sum_i \beta_t^i \varepsilon_t^i \xrightarrow{d} \mathcal{N}(0, \Sigma_{\beta_t})$ *with* $\Sigma_{\beta_t} = \underset{n\to\infty}{plim} \left(\frac{1}{n} \sum_i (\varepsilon_t^i)^2 \beta_t^i (\beta_t^i)' \right)$, *and* $(\beta_t^i)' = (\ 1 \vdots (b_t^i)'\ )$.

Assumption 12. *As* $T \to \infty$,

(i) $Q_Z = E(Z_{t-1} Z_{t-1}')$ *is a full rank matrix*

(ii) $\frac{1}{T} \sum_t (Z_{t-1} \otimes u_t) \xrightarrow{p} 0$

(iii) $\frac{1}{\sqrt{T}} \sum_t (Z_{t-1} \otimes u_t) \xrightarrow{d} \mathcal{N}(0, \Sigma_u)$, *with* $\Sigma_u = E\left(Z_{t-1} Z_{t-1}' \otimes u_t u_t'\right)$.

(iv) $\frac{1}{\sqrt{T}} \sum_t Z_{t-1} u_t^{us} \xrightarrow{d} \mathcal{N}(0, \Sigma_u^{us})$, *with* $\Sigma_u^{us} = E\left((u_t^{us})^2 Z_{t-1} Z_{t-1}'\right)$

Assumption 11 is the central limit theorem (CLT) for the cross-sectional errors each month t. Assumption 12 is Assumption A.2(b) of GOS.

Proposition 1. *Under Assumptions 1-9, if* $\frac{\tau_n}{n} \to \tau_0 > 0$ *as* $n \to \infty$, *then we have at every month* t*:*

$$\hat{\phi}_t = \phi_t + o_p(1). \tag{1.9}$$

From Assumptions 9 and 12 and by applying the LLN, we have $\hat{F} = F + o_p(1)$ as $n \to \infty$. Then by Proposition 1 and by the definition of ϕ_t we have as $n \to \infty$,

$$\begin{cases} \hat{\lambda}_t \xrightarrow{p} \lambda_t & \text{if the two factors are useful at t} \\ \hat{\lambda}_t \xrightarrow{p} \begin{pmatrix} \lambda_t^{us} \\ 0 \end{pmatrix} & \text{if the factor } \bar{us} \text{ is useless at t.} \end{cases} \tag{1.10}$$

At each date t, the consistency holds for the useful factor. When the dynamically useless factor $\bar{us}$ is useless in a given month t its estimated risk premium is shrunk to 0. This value is just a target value, as the risk premium of the factor $\bar{us}$ is not identified in months where this factor is useless.

For the next proposition, remember that $Q_Z \equiv E\left(Z_{t-1} Z_{t-1}'\right)$, as defined as in Assumption 12(i).

Proposition 2. *Under Assumptions 1-12, if* $\frac{\tau_n}{\sqrt{n}} \to \tau_0 > 0$ *as* $n \to \infty$, *then we have at each month* t*:*

(i) if the two factors are useful at t,

$$n^{\alpha}(\hat{\lambda}_t - \lambda_t) \xrightarrow{d} \mathcal{N}\left(0, \rho_1 (Z_{t-1}' \otimes I_2) \Sigma_F (Z_{t-1} \otimes I_2)\right), \tag{1.11}$$

with

$$\Sigma_F = \left(Q_Z^{-1} \otimes I_2\right) \Sigma_u \left(Q_Z^{-1} \otimes I_2\right); \tag{1.12}$$

(ii) if the factor $\bar{us}$ is useless at t,

$$n^\alpha(\hat{\lambda}_t^{us} - \lambda_t^{us}) \xrightarrow{d} \mathcal{N}\left(0, \rho_1 Z_{t-1}' \Sigma_F^{us} Z_{t-1}\right), \tag{1.13}$$

with

$$\Sigma_F^{us} = Q_Z^{-1} \Sigma_u^{us} Q_Z^{-1}. \tag{1.14}$$

The standard error of the risk premium is influenced by the errors-in-variable from the first step, that is why the convergence rate of the second-step estimator is of the order of $\alpha \in [0.25, 0.5)$.

1.4 Simulations

In this section, we evaluate: (i) the robustness of the shrinking procedure to variations in tuning parameters, (ii) the ability for the procedure to detect periods where the factor is useless and (iii) the ability for the estimator to restore reliable asymptotic properties.

In all of our simulations, we consider the 255 portfolios used in our empirical analysis (see the description of the data below) and perform a principal components analysis (pca) on them in order to select the first two principal components (since our theoretical framework relies on two factors). We also work with the convergence rate $\alpha = 0.25$. We choose this value as we show in the appendices that there is an arbitrage to be realized; the asymptotic bias of the estimator increases with its convergence rate. Thus, we achieve precise estimation at the expense of a smaller rate of convergence. We choose $\alpha = 0.25$ as it is a small but yet fair rate of convergence.

1.4.1 Tuning parameters and detection of dynamically useless factors

We perform a Principal Component Analysis (PCA) on the returns and extract the first two principal components as factors. Then, as we have daily and monthly data, we have both factors 1 and 2 at the monthly level and at the daily level.

Spurious replacements. During the simulations, we call a *spurious replacement* the process by which factors 2 in both daily and monthly levels are replaced in certain dates by useless factors. At the monthly level, a spurious replacement at date t will consist of replacing monthly factor 2 at t, by the real consumption per capita at t. At the daily level, a spurious replacement at date t will consist of replacing the D_t values of daily factor 2 in month t, by D_t parameters generated from a normal distribution whose mean and variance are calibrated on the real consumption per capita. Spurious replacements will therefore leave unchanged factor 1, and generate a new factor 2, which will be likely dynamically useless. As the consumption is generally uncorrelated with traded returns,[9] dates where spurious replacements are performed are those where factor 2 will be useless.

We perform 500 Monte-Carlo replications where at each replication, $Nspur$ months are randomly selected, "spurious replacements" are performed in each of the selected months, and risk premia are estimated in all of the T months. At each replication, we also estimate the shrinkage probability as the probability to shrink the risk premium of factor 2 to 0 over the $Nspur$ dates where this factor is useless.

Table 1.3: Shrinkage probabilities when the tuning parameters s and τ vary

	$Nspur = 50$			$Nspur = 100$			$Nspur = 500$		
	$\tau = 1$	$\tau = 50$	$\tau = 200$	$\tau = 1$	$\tau = 50$	$\tau = 200$	$\tau = 1$	$\tau = 50$	$\tau = 200$
$s = 3$	0.998	1	1	0.997	1	1	0.997	1	1
$s = 6$	0.997	0.999	1	0.997	0.999	0.999	0.997	0.999	0.999
$s = 12$	0.996	0.998	0.999	0.996	0.998	0.999	0.997	0.998	0.999

We perform 500 Monte-Carlo replications where at each replication, we randomly select $Nspur$ months, simulate the useless factor to insert in those dates, and estimate the risk premia in all the $T = 617$ dates. The useless factor here is a random variable mimicking the real consumption per capita.
We estimate at each replication the shrinkage probability, as the probability to shrink the risk premium of the second factor to 0, over the $Nspur$ dates where this factor is useless. The probabilities presented in this table are the mean of the shrinkage probabilities over the 500 Monte-Carlo replications.

Table 1.3 presents the shrinkage probabilities across the Monte-Carlo process. We see that the estimator performs very well at detecting the dates where the second factor is useless. For all the selected values of the tuning parameters τ and s, it successfully recognizes the dates where a useless factor has been introduced as second factor, with an accuracy greater than 99.8%. This result confirms the robustness of the shrinking procedure to tuning param-

[9] Bryzgalova (2016) also uses consumption to generate useless factors.

eters variation, as presented by Bryzgalova (2016). Furthermore, the shrinkage ability of the estimator is not reduced when the number of "spurious dates" $Nspur$ increases. In fact, the shrinkage probability is greater than 0.99 when $Nspur = 50$, as much as when $Nspur = 100$ or 500

Now that we know that the estimator will always detect the dates where a factor is useless, we have to measure its probability of performing false detections. A false detection here stands for shrinking the risk premium of a factor to zero when that factor is not useless, and vice versa. Table 1.4 presents the probability for the estimator, to make a false detection during the shrinkage process in the Monte-Carlo. We see that this probability is close to 0 for all of the selected tuning parameters. We can therefore conclude that the estimator performs very well for the detection of dynamically useless factors and is strongly robust to variations in tuning parameters. We will use the values $\tau = 50$ and $s = 3$ in the remaining simulations and in the empirical application.

Table 1.4: Probability of having an erroneous shrinkage

	$Nspur = 50$			$Nspur = 100$			$Nspur = 500$		
	$\tau = 1$	$\tau = 50$	$\tau = 200$	$\tau = 1$	$\tau = 50$	$\tau = 200$	$\tau = 1$	$\tau = 50$	$\tau = 200$
$s = 3$	0	0	0	0	0	0	0.002	0	0
$s = 6$	0	0	0	0	0	0	0.002	0	0
$s = 12$	0	0	0	0	0	0	0.003	0.001	0.001

We perform 500 Monte-Carlo replications where at each replication, we randomly select $Nspur$ months, simulate the useless factor to insert in those dates, and estimate the risk premia in all the $T = 617$ dates. The useless factor here is a random variable mimicking the real consumption per capita.

We estimate at each replication the probability of having an erroneous shrinkage as the probability, over the $Nspur$ dates, to shrink the risk premium when the factor is not useless or to not shrink it when the factor is useless. The probabilities presented in this table are the mean of these probabilities over the 500 Monte-Carlo replications.

1.4.2 Shrinking procedure and volatility of factors

As seen in the previous subsection, the estimator performs very well for the detection of dynamically useless factors. As the shrinking procedure relies on the correlations between the returns and the factors,[10] we verify in this subsection whether noises in the data influence the procedure, besides correlations. We know that a factor is useless when it is uncorrelated with

[10] Following Bryzgalova (2016) the proposed estimator is a penalized-Lasso, with the penalization depending on the values of the betas.

the asset returns. This absence of correlation can come from either the nature of the factor or from noisy variations in the values of the factor. Therefore, we have to verify if the shrinkage ability of the estimator is robust to noisy variations.

We analyze the dynamics of the shrinkage probabilities following the volatility of a simulated factor. As previously, we perform a PCA on the returns and extract the first two principal components (factor 1 and factor 2). We measure the volatility σ_{2d} of the daily factor 2 and consider 500 different values σ_s ($s = 1, \ldots, 500$), between $1/2 \times \sigma_{2d}$ and $100 \times \sigma_{2d}$. Then, we perform 500 replications where for each replication s: (a) we randomly select between 100 and 500 $Nspur$ months; (b) we perform "spurious replacements" in each of the selected $Nspur$ months with the difference that, here for replication s, daily factor 2 is replaced by parameters generated from $\mathcal{N}(0, \sigma_s^2)$ (the other aspects of "spurious replacements" remain as presented above); and (c) we measure the shrinkage probability of the estimator.

Figure 1.5 presents the evolution of the shrinkage probability, following the standard error of the noise. We see that this shrinkage probability is greater than 0.9 as soon as $\frac{std(factor2)}{\sigma} \leq 0.08$. Therefore, our estimator systematically shrinks the risk premia of factor 2 to 0, only at periods where the standard error of this factor is 12.5 times greater than its standard error over the periods where it is useful ($12.5 = 1/0.08$). Since a 12.5-times larger standard error can objectively be considered as huge, we can therefore conclude that the shrinkage ability of the estimator is robust to noisy variations in the factors data.

1.4.3 Ability to restore asymptotic properties

Again, we perform a PCA on the returns and take the first two principal components as factor 1 and factor 2.

Analysis around a randomly selected date

We randomly select a date t_0 and perform a "spurious replacement" at this date. We then consider the model with the new dynamically useless factor as the real DGP. From that, in each month t: (a) we estimate $\hat{b}_t^i$ by an OLS regression of daily returns in month t on daily factors in month t; (b) we estimate $\hat{\Phi}_t = \left(\hat{\phi}_t^0, \hat{\phi}_t'\right)'$ by (1.7); (c) we estimate the covariance matrix (Σ) of the residuals $\hat{\varepsilon}_t^i$ in (1.6); (d) we estimate the risk premium $\hat{\lambda}_{t,\star} = \left(\hat{\lambda}_{t,\star}^{us\ \prime},\ \hat{\lambda}_{t,\star}^{\bar{us}\ \prime}\right)'$ by (1.8).

We perform 1000 Monte-Carlo replications on each of the dates $t_0 - 1$, t_0 and $t_0 + 1$. For each replication s, we estimate the monthly returns following equation (1.6): $\hat{R}^i_{t,s} = \left(1, (\hat{b}^i_{t,s})'\right) \hat{\Phi}_t + \hat{\varepsilon}^i_{t,s}$. Note that $\hat{b}^i_{t,s}$ is generated from a normal distribution with mean and variance calibrated on the empirical betas $\hat{b}^i_t$ obtained above, while $\hat{\varepsilon}^i_{t,s}$ is generated from a normal distribution with mean 0 and variance Σ. Using $\hat{R}^i_{t,s}$ and $\hat{b}^i_{t,s}$, we estimate $\hat{\Phi}_{t,s} = \left(\hat{\phi}^0_{t,s}, \hat{\phi}'_{t,s}\right)'$ and deduce the risk premium $\hat{\lambda}_{t,s} = \left(\hat{\lambda}^{us\,\prime}_{t,s},\ \hat{\lambda}^{\bar{us}\,\prime}_{t,s}\right)'$. In each replication, we test the nulls $H_0:\ \lambda^{us}_t = \lambda^{us}_{t,\star}$ and $H_0:\ \lambda^{\bar{us}}_t = \lambda^{\bar{us}}_{t,\star}$, with $\lambda_{t,\star}$ the empirical value of the risk premium according to the DGP obtained after the spurious replacements described above. The simulations are performed for various values of n (200, 600, 1000, 2500, and 5000) and the rejection probabilities are estimated for each date over the 1000 replications. To simplify notations, the tests will next be denoted without the time-subscripts. Therefore, we will simply write $H_0:\ \lambda^{us} = \lambda^{us}_\star$ and $H_0:\ \lambda^{\bar{us}} = \lambda^{\bar{us}}_\star$ while acknowledging that these tests are performed each month t.

Table 1.5 presents the rejection probabilities at $t_0 - 1$, t_0 and $t_0 + 1$. For the always useful factor (factor 1), we see that the equality between the true risk premium and the estimated one is never rejected at $t_0 - 1$ and $t_0 + 1$. Moreover, the rejection probability for the test $H_0:\ \lambda^{us} = \lambda^{us}_\star$ decreases as n increases. For large values of n (precisely, for n greater than 1000), inference properties are restored for factor 1 at the period t_0 (remember that t_0 is the period where a useless factor has been introduced as factor 2).

For the dynamically useless factor (factor 2), we first see that the null of $H_0:\ \lambda^{\bar{us}} = \lambda^{\bar{us}}_\star$ is never rejected at t_0. This is not surprising, as we have previously shown that the estimator is able to perfectly recognize periods where a factor is useless. Secondly, as for the test on factor 1, we see that the rejection probability also decreases as n increases (remember, we need $n \geq 1000$ for a full restoration of the inference properties related to factor 1). For factor 2, the restoration of inference properties at $t_0 - 1$ and $t_0 + 1$ requires a much higher value of n. These inference properties are fully restored in the simulations for $n > 2500$. Therefore, the more assets we have, the better the estimator will perform.

Figure 1.5: Evolution of the shrinkage probability following the standard error of the noise

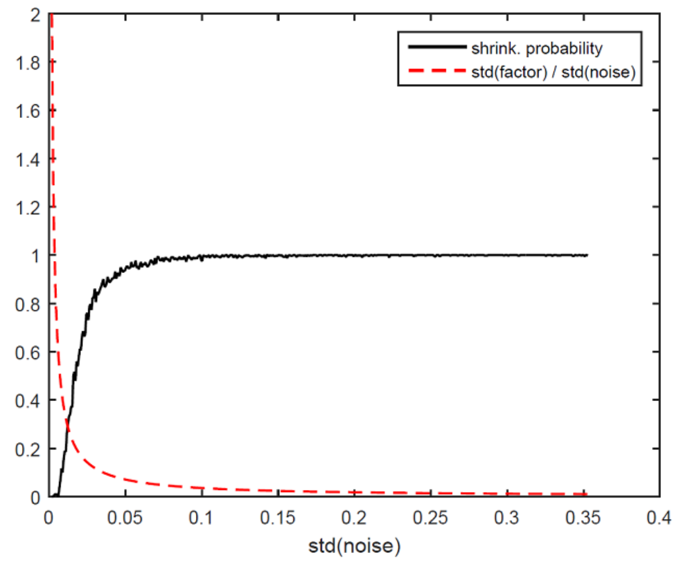
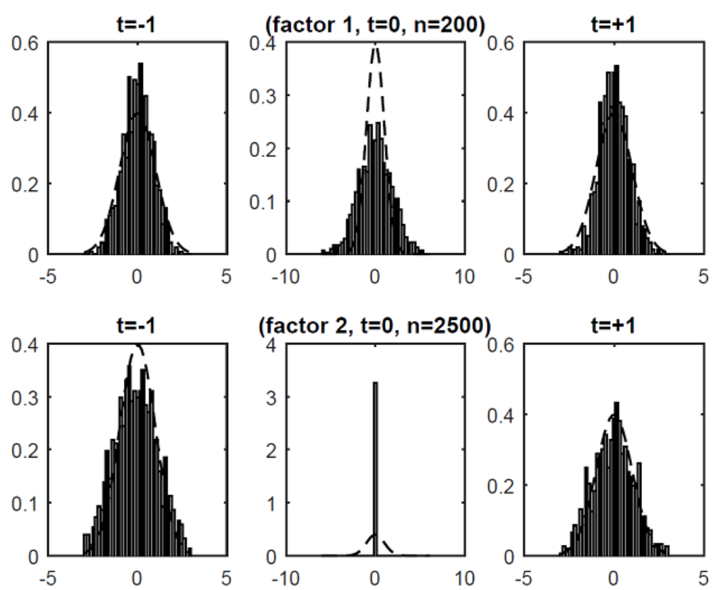


Figure 1.6: Histogram of the t-stats around t_0, and gaussian distribution



The two lines of figures present the histograms from the distribution of respectively the factor 1 and 2, at $t_0 - 1$, t_0 and $t_0 + 1$. $t_0 = 0$ is a random date where the second factor is replaced by a useless factor.
For the factor 1, we present the simulations for $n = 200$, whereas the simulations for the factor 2 are presented for $n = 2500$. The dashed line is the density of the standard normal distribution.

Table 1.5: Rejection probabilities around a random spurious date t_0

	$H_0 : \lambda^{us} = \lambda_\star^{us}$ (panel A)								
	1%			5%			10%		
	-1	0	+1	-1	0	+1	-1	0	+1
$n = 200$	0.001	0.170	0.002	0.014	0.300	0.015	0.041	0.388	0.048
$n = 600$	0	0.018	0	0	0.074	0	0	0.138	0.001
$n = 1000$	0	0.002	0	0	0.015	0	0	0.055	0
$n = 2500$	0	0	0	0	0	0	0	0.001	0
$n = 5000$	0	0	0	0	0	0	0	0	0
	$H_0 : \lambda^{\bar{us}} = \lambda_\star^{\bar{us}}$ (panel B)								
	1%			5%			10%		
	-1	0	+1	-1	0	+1	-1	0	+1
$n = 200$	0.569	0	0.525	0.683	0	0.618	0.726	0	0.659
$n = 600$	0.315	0	0.263	0.424	0	0.402	0.507	0	0.481
$n = 1000$	0.201	0	0.170	0.308	0	0.293	0.398	0	0.383
$n = 2500$	0.040	0	0.026	0.115	0	0.081	0.197	0	0.141
$n = 5000$	0.002	0	0.001	0.029	0	0.016	0.069	0	0.057

We randomly select a date between 1 and T (date 0), and insert in that date, a simulated useless factor as the second factor of the model. The simulated useless factor is a random variable mimicking the real consumption per capita. We also consider the previous date (date -1) and the next one (date +1), and perform 1000 Monte-Carlo experiments at each of the three selected dates.
At each replication, we estimate the risk premium and the t-statistics associated to the tests presented. We compare the t-statistics with the critical values from the different tests sizes, and reject the null if the t-stat is greater than the critical value. Panel A presents the results for the useful factor, while panel B presents those for the dynamically useless one. The results are presented for different values of the number of assets (n).

Analysis on all the sample and usage according to the number of assets n

We repeat the same experiment as previously, except that before the Monte-Carlo, the spurious replacements are performed on $Nspur$ random dates. We also perform estimations at all the T periods in each of the 1000 replications. We then estimate the mean and the median of the rejection probabilities over the T periods for factor 1 (the useful factor), and over $T - Nspur$ periods for factor 2 (the dynamically useless factor).

We see in Tables 1.6 and 1.7 that the means and the medians of the rejection probabilities for the useful factor are almost the same as the means and the medians for the dynamically useless factor (over the dates where this factor is useful). This similarity is particularly true for large values of n. This confirms again that the ability of the estimator to reduce the effects of "useless dates" increases with n.

Table 1.6: Rejection probabilities over time with $Nspur = 50$

	$H_0 : \lambda^{us} = \lambda_\star^{us}$ (panel A)					
	1%		5%		10%	
	mean	median	mean	median	mean	median
$n = 200$	0.117	0.029	0.179	0.095	0.224	0.162
$n = 600$	0.036	0	0.068	0.003	0.097	0.015
$n = 1000$	0.018	0	0.038	0	0.057	0.001
$n = 2500$	0.004	0	0.010	0	0.017	0
$n = 5000$	0.001	0	0.003	0	0.005	0
	$H_0 : \lambda^{\bar{u}s} = \lambda_\star^{\bar{u}s}$ (panel B)					
	1%		5%		10%	
	mean	median	mean	median	mean	median
$n = 200$	0.115	0.026	0.176	0.088	0.222	0.146
$n = 600$	0.036	0	0.068	0.004	0.096	0.013
$n = 1000$	0.018	0	0.038	0	0.057	0.002
$n = 2500$	0.004	0	0.011	0	0.018	0
$n = 5000$	0.001	0	0.003	0	0.006	0

We randomly select 50 dates between 1 and T, and insert in these dates, a simulated useless factor as the second factor of the model. The simulated useless factor is a random variable mimicking the real consumption per capita. We perform 1000 Monte-Carlo experiments.

At each replication, we estimate at all the dates, the risk premium and the t-statistics associated to the tests. We compare the t-statistics with the critical values from the different tests sizes, and reject the null if the t-stat is greater than the critical value. For the useful factor, the mean and the median in each replication are the mean value and the median value of all the rejection probabilities over the T dates. For the dynamically useless factor, the mean and the median are estimated over the dates where this factor is useful ($T - 50$ dates).

The results presented here are the mean and the median over the Monte-Carlo experiments. Panel A presents the results for the useful factor, while panel B presents those for the dynamically useless one. The results are presented for different values of the number of assets (n).

Table 1.7: Rejection probabilities over time with $Nspur = 100$

	$H_0 : \lambda^{us} = \lambda_\star^{us}$ (panel A)					
	1%		5%		10%	
	mean	median	mean	median	mean	median
$n = 200$	0.145	0.043	0.210	0.126	0.257	0.197
$n = 600$	0.053	0	0.091	0.007	0.123	0.025
$n = 1000$	0.029	0	0.055	0	0.078	0.004
$n = 2500$	0.008	0	0.018	0	0.029	0
$n = 5000$	0	0	0.006	0	0.011	0
	$H_0 : \lambda^{\bar{us}} = \lambda_\star^{\bar{us}}$ (panel B)					
	1%		5%		10%	
	mean	median	mean	median	mean	median
$n = 200$	0.087	0.010	0.141	0.048	0.183	0.097
$n = 600$	0.024	0	0.047	0.001	0.070	0.004
$n = 1000$	0.011	0	0.025	0	0.039	0
$n = 2500$	0.002	0	0.006	0	0.011	0
$n = 5000$	0	0	0.001	0	0.003	0

We randomly select 100 dates between 1 and T, and insert in these dates, a simulated useless factor as the second factor of the model. The simulated useless factor is a random variable mimicking the real consumption per capita. We perform 1000 Monte-Carlo experiments.

At each replication, we estimate at all the dates, the risk premium and the t-statistics associated to the tests. We compare the t-statistics with the critical values from the different tests sizes, and reject the null if the t-stat is greater than the critical value. For the useful factor, the mean and the median in each replication are the mean value and the median value of all the rejection probabilities over the T dates. For the dynamically useless factor, the mean and the median are estimated over the dates where this factor is useful ($T - 100$ dates).

The results presented here are the mean and the median over the Monte-Carlo experiments. Panel A presents the results for the useful factor, while panel B presents those for the dynamically useless one. The results are presented for different values of the number of assets (n).

When we consider all the periods (instead of just one period t_0 as previously), we see that the restoration of inference properties in the "useful dates" requires much less assets than what being suggested by the previous analysis around the random date t_0. While considering the mean and median of the rejection probabilities, it appears that the restoration is overall reached for $n \geq 600$, for both factor 1 and factor 2. So even if there are some dates around which the correction requires a very large number of assets, this is not always the case. On average, we do not need a very large number of assets to achieve corrections in most of the "useful dates".

1.5 Empirical analysis

1.5.1 Data

As we assume that n is large, our setting is suitable for large equity data sets. Therefore, we can apply our estimator on both individual assets and portfolios. For our analysis, we consider excess returns over the risk-free rate on 255 portfolios: 25 portfolios sorted by size and investment, 25 portfolios sorted by book-to-market and investment, 25 portfolios sorted by book-to-market and operating profitability, 25 portfolios sorted by operating profitability and investment, 25 portfolios sorted by size and book-to-market, 25 portfolios sorted by size and long-term reversal, 25 portfolios sorted by size and momentum, 25 portfolios sorted by size and operating profitability, 25 portfolios sorted by size and short term reversal, and 30 industry portfolios. Moreover, we consider the five Fama-French factors and the momentum factor. All of these portfolios and factors are from Kenneth French's website; we use both the daily data from 01/07/1964 to 31/12/2015 and the monthly data from July 1964 to December 2015. We also consider for the same frequencies and for the same span, the real consumption per capita (using the consumer price index for all urban consumers as deflator); this variable is used in simulations to calibrate the properties of the dynamically useless factor. On the other hand, we consider as instruments the term spread (proxied by the difference between the yields on 10-year treasury and the yields on 3-month treasury bill), and the default spread (proxied by the difference between the yields on Moody's BAA bonds and the yields on Moody's AAA bonds). The instruments and the consumption variable are all from the Federal Reserve Bank of St-Louis.

As emphasized above, we need factors observed at a high-frequency in order to examine the "dynamically useless" property. So, the factors have to be displayed at least in a daily frequency. Then, for reason of availability, we only analyze three empirical models here: the three-factor model of Fama & French (1993), the four-factor model of Carhart (1997), and the five-factor model of Fama & French (2015). We group the daily data in monthly periods and estimate, for each model, the risk premium of all the factors.

1.5.2 Dynamics of the risk premia and financial implications

Relevance of the factors

Table 1.8 presents the dynamics of the time-varying risk premia in the selected models. In each model, all of the factors have dates with insignificant risk premia (at the 5% level of significance). The market is the factor that is more often rewarded, since its risk premium in all of the three models remains significant more than 95% of the time. The market can therefore be considered as "always useful", as the chance of having that factor rewarded by a null risk premium is negligible. This confirms what we assumed in our model, as we previously stated that the market is likely to be an "always useful" factor.

On the other hand, the size factor SMB is the factor that is more often linked to an insignificant risk premium. It is unrewarded 8.7% of the time in the Fama-French 3 model, 10.2% of time in the Carhart model, and 9.6% in the Fama-French 5 model. Hence, this factor is still rewarded around 90% of the time.

Moreover, we see that the results are very stable across the models. The introduction of new factors on the initial three Fama-French factors does not cause any distortion in the results. This stability in the results tells us that the new factors do not bring any multicollinearity issue, and therefore none of the factors in these models is redundant. This result is particularly interesting for the Fama-French five-factor model of Fama & French (2015), as the main argument used again this model is that some factors would become redundant when we add the operating profitability factor (RMW) and the investment factor (CMA) to the original three factors of Fama & French (1993) (See for example Hou, Xue & Zhang, 2014). While this redundancy issue can serve as an argument for those trying to explain or predict asset returns, it is clearly not an issue here, as the information brought by the additional factors are quite different. It should be noted for the Fama-French five-factor model that, the only month where three factors were unrewarded at the same time is January 1970 (Mkt-Rf, HML, and CMA). Also, the last month for which two factors were unrewarded at the same time is March 1976.

Additional dynamical evidences

Figures 1.7 and 1.8 illustrate the evolution of the instruments (default spread and term spread), and present (in shaded) the US recessions according to the National Bureau of Economic Research (NBER). The figures also show (in red) the months where each factor is unrewarded.

We see that the risk premium of the profitability factor RMW is never insignificant during recessions; this indicates that investors should continue to build strategies on profitability even during recessions. We also see that the risk premium of the investment factor CMA is insignificant mostly at periods where the dynamic of the default spread changes sign. So, when the investment factor becomes unrewarded, the default spread is at its peak or trough, depending on its dynamics on previous dates. Therefore, for agents, the most important consideration regarding investment strategies should not be the gap between the top-rated stocks' returns and the the bottom-rated stocks' returns; rather, it is the dynamics of that gap. The strategies on investment should be performed as long as that gap exhibits a monotonic dynamic in a recent past.

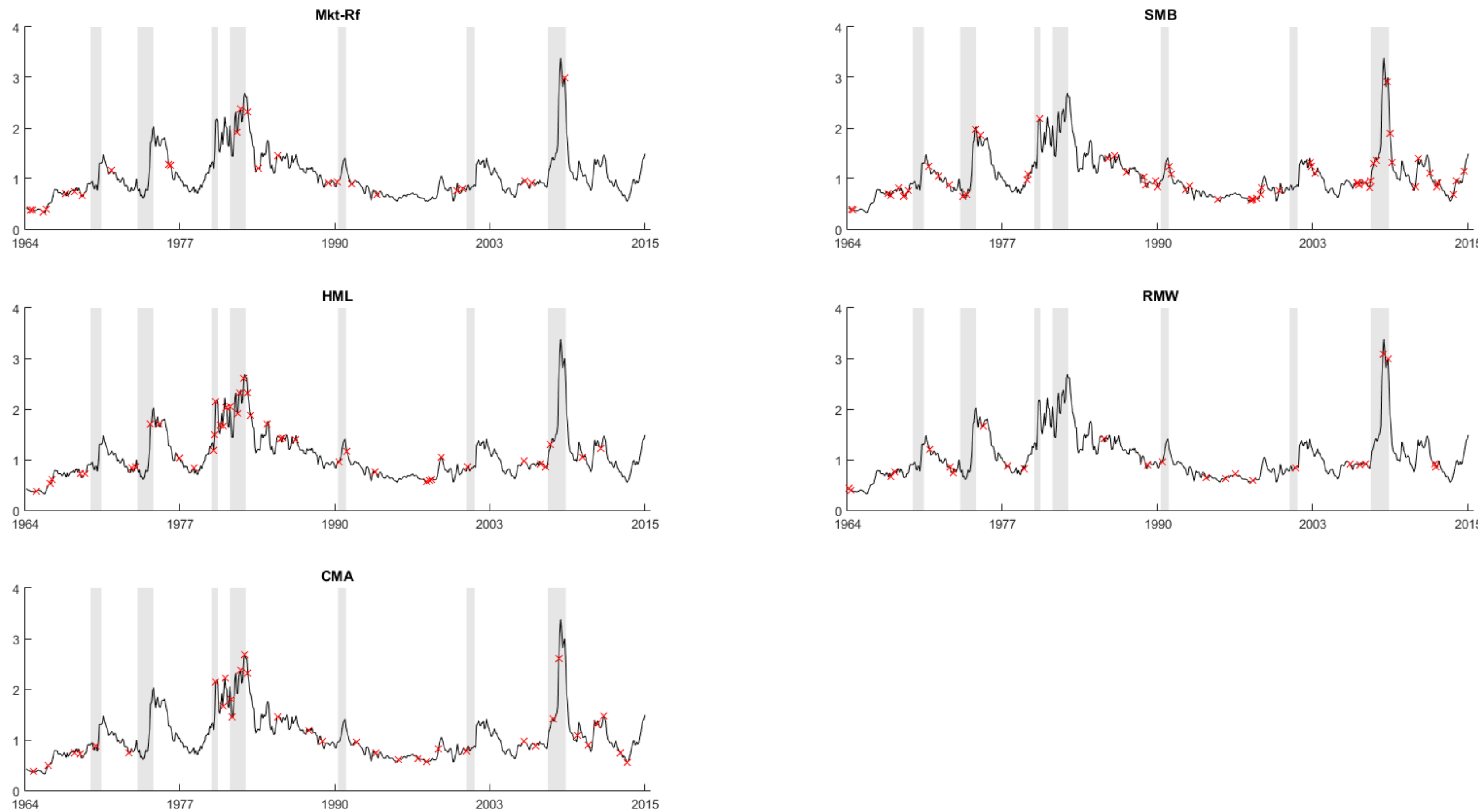
Table 1.8: Empirical risk premia

		risk premium (×100)					std. error (×100)					insignificant risk premia (number of months over 617)
			percentile					percentile				
		mean	20	40	60	80	mean	20	40	60	80	
Fama-French (1993)	Mkt-Rf	-0.16	-2.81	-1.05	0.55	2.49	0.32	0.15	0.22	0.30	0.45	26 (0.042)
	SMB	0.13	-0.61	-0.09	0.34	0.83	0.22	0.13	0.19	0.21	0.27	54 (0.087)
	HML	0.12	-0.93	-0.17	0.43	1.23	0.21	0.09	0.15	0.20	0.28	47 (0.076)
Carhart (1997)	Mkt-Rf	-0.15	-2.70	-0.96	0.46	2.36	0.40	0.22	0.32	0.38	0.51	31 (0.050)
	SMB	0.14	-0.57	-0.08	0.34	0.83	0.24	0.17	0.21	0.25	0.30	63 (0.102)
	HML	0.13	-0.86	-0.17	0.42	1.20	0.24	0.12	0.18	0.24	0.29	48 (0.078)
	Mom	0.11	-1.51	-0.37	0.57	1.64	0.38	0.17	0.25	0.34	0.51	44 (0.071)
Fama-French (2015)	Mkt-Rf	-0.11	-2.71	-1.05	0.56	2.41	0.27	0.15	0.21	0.28	0.37	24 (0.039)
	SMB	0.15	-0.56	-0.04	0.38	0.85	0.21	0.16	0.17	0.22	0.28	59 (0.096)
	HML	0.12	-0.91	-0.17	0.47	1.16	0.23	0.14	0.19	0.22	0.28	41 (0.066)
	RMW	0.06	-0.84	-0.22	0.29	0.95	0.10	0.06	0.07	0.10	0.14	25 (0.040)
	CMA	0.08	-0.76	-0.14	0.35	0.92	0.16	0.10	0.13	0.17	0.23	34 (0.055)

The table gives the dynamics of the risk premia in the chosen models. The means and the percentiles are estimated over the T dates ($T = 617$). The last column gives the number of months over the 617, where the risk premium of the factor is insignificant at the level of significance 5%.

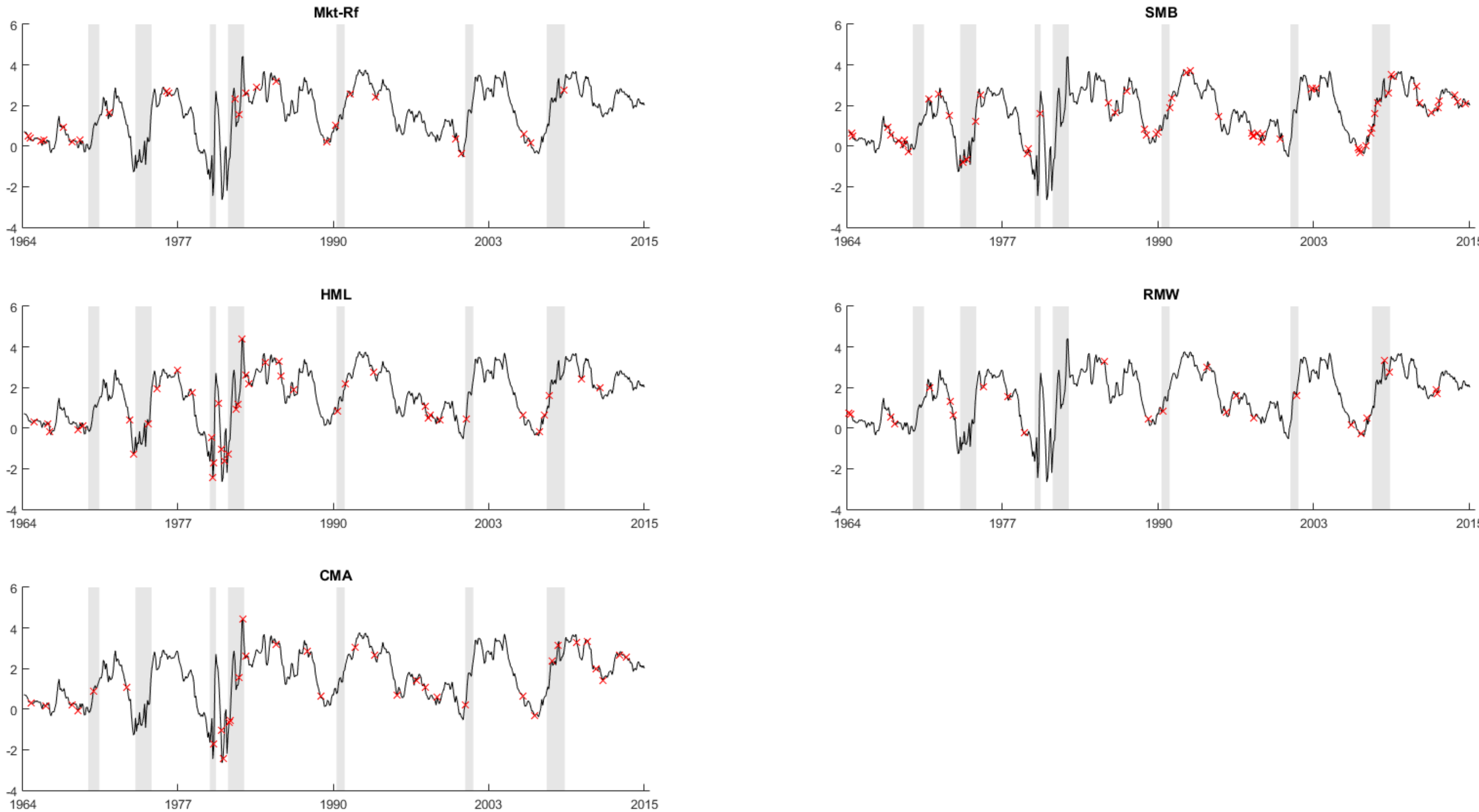
The numbers in brackets are the ratios between the number of periods where the risk premium is insignificant, and the overall number of periods. They give the probability for each factor, of being unrewarded.

Figure 1.7: Default spread and dates where the risk premium of the factor is zero



The figure presents the evolution of the monthly default spread from 1964 to 2015; and for every factor, the dates where its associated risk premium is insignificant (red crosses). The shaded bands represent the US recessions, according to the data of the National Bureau of Economic Research (NBER)

Figure 1.8: Term spread and dates where the risk premium of the factor is zero



The figure presents the evolution of the monthly term spread from 1964 to 2015; and for every factor, the dates where its associated risk premium is insignificant (red crosses). The shaded bands represent the US recessions, according to the data of the National Bureau of Economic Research (NBER)

1.6 Conclusion

There are several factors that have been used in the literature to explain assets returns. Following the dynamics on the markets and due to specific phenomena such as shocks, asymmetries, or momenta, the correlations of these factors with assets returns continuously change through time. Thus, the set of factors that explain assets returns might also change through time.

In this paper, we consider this eventuality and show that given a prior selection of factors, they are not all always important at the same time. Hence, we show that a factor can be useless in a given period and no longer be so in another one. We introduce the notion of a dynamically useless factor as a factor that is useless in at least one time period (assuming that there is at least one factor that is always useful). Then, we propose a method to consistently estimate the conditional risk premium of a factor in all of the periods in which it is useful, assuming that the model includes a dynamically useless factor.

We consider two sets of data in our theoretical framework: a periodic data set as we estimate the risk premium every period, and a high-frequency dataset that helps us overcome the periodic unobservabilities and characterize the factors. Assuming that the number of periods, the number of assets, and the number of high-frequency dates in each period are all large, our estimator is consistent in every period for which the factor is useful, and equal to 0 when the factor is useless.

Empirically, we show through the shrinking mechanism of our estimator, that the principal factors of the Fama-French models are likely dynamically useless, as they all experience times where their risk premia are insignificant. Nevertheless, the market return can be considered as an "always useful" factor, as the times where it has an insignificant risk premium are very negligible. We also show that the profitability factor RMW and the investment factor CMA in the Fama-French five-factor model bring additional and very different information on risk premia. In fact, it appears that: (i) investors should continue to build strategies on profitability during recessions; and (ii) the strategies on investment should be performed as long as the gap between the returns on top rated stocks and on bottom-rated stocks maintains a monotonic dynamic in a recent past.

Appendices for Chapter 1 (A)

A1 Dynamicaly useless factors and risk premium estimation

We present in this section why the presence of a dynamically useless factor can distort the estimation of the time-varying risk premium, when the conventional methods are used. Some papers present in that vein (in the time-invariant framework), how the asymptotic distribution of the risk premium is distorted when a useless factor is included in the model (see for example Kleibergen, 2009; Kleibergen & Zhan, 2015); we do not follow that idea. Here we just present how the presence of a useless factor leads, for the conventional time-varying risk premium estimation, to a failure of the rank condition (singularity of the loadings matrix). For that purpose, we focus our analysis on GOS, who propose a method to test the time-varying properties of the risk premia in an arbitrage pricing framework.[11] We analyze a simplified version of the model presented in that paper.

A1.1 Theoretical framework of GOS

We present here the outline of the theoretical framework used in GOS and show how the time-varying risk premium estimator is defined in that paper.

We have $(\Omega, \mathscr{F}, P)$ a probability space and $S : \Omega \to \Omega$ the measurable function that describes the dynamics. If $\omega \in \Omega$ is the state of the world at time 0, $S^t(\omega)$ is its state at time t with S^t equivalent to t successive applications of S. S is assumed to be measure-preserving and ergodic. The flow of information available in the market is represented by the filtration $\mathscr{F}_t, t = 1, 2, \ldots,$ where $\mathscr{F}_t = \{S^{-t}(A), A \in \mathscr{F}_0\}$ with $\mathscr{F}_0$ a given sub sigma-field of $\mathscr{F}$ and S^{-t} the inverse application of S^t. There is a large number of assets randomly selected in the sub-space $[0, 1]$.[12] The excess return $R_t(\gamma)$ is defined at date $t = 1, 2, \ldots, T$ for a given asset $\gamma \in [0, 1]$ by:

$$R_t(\gamma) = a_t(\gamma) + b_t(\gamma)' f_t + \varepsilon_t(\gamma). \tag{A.1}$$

[11] GOS assume that there is non-degeneracy in the factor sensitivities across assets, so their model does not consider useless factors

[12] GOS show that under a suitable sampling mechanism, the model is robust to reordering of the assets.

Where $(a_t(\gamma), b_t(\gamma)^{'})^{'} \in \mathbb{R} \times \mathbb{R}^K$ with $b_t(\gamma)$ the sensitivity of the factor $f_t \in \mathbb{R}^K$ to the asset γ. $\varepsilon_t(\gamma)$ represent errors. All these variables are defined on the probability space and their conditional specifications are: $a_t(\gamma) = a_t\left(\gamma, S^{t-1}(\omega)\right)$, $b_t(\gamma) = b_t\left(\gamma, S^{t-1}(\omega)\right)$, $\varepsilon_t(\gamma) = \varepsilon_t\left(\gamma, S^{t}(\omega)\right)$, and $f_t(\omega) = f(S^t(\omega))$.

Under assumptions APR.1 to APR.4 from GOS, $a_t(\gamma) = b_t(\gamma)^{'}\nu_t$. So the real risk premia at time t is given by

$$\lambda_t = \nu_t + E\left(f_t|\mathscr{F}_{t-1}\right).$$

Factor sensitivities and risk premia at t are defined conditionally with respect to an instrument Z_{t-1} which may include past observations of the factors and some macroeconomic variables. Stock-specific instruments $Z_{t-1}(\gamma)$ are also used to describe the dynamics of factor sensitivities. The following two assumptions show how the variables are linked to instruments;

Assumption FS 1. *For any $\gamma \in \{\gamma_1, \ldots, \gamma_n\}$ and $t = 1, 2, \ldots,$ the factor loadings are given by $b_t(\gamma) = B(\gamma)Z_{t-1} + C(\gamma)Z_{t-1}(\gamma)$, with $B_t(\gamma) \in \mathbb{R}^{K\times p}$ and $C_t(\gamma) \in \mathbb{R}^{K\times q}$.*

Assumption FS 2. *(i) The risk premia vector is given by $\lambda_t = \Lambda Z_{t-1}$, with $\Lambda \in \mathbb{R}^{K\times p}$. (ii) For any t, $E(f_t|\mathscr{F}_{t-1}) = FZ_{t-1}$, with $F \in \mathbb{R}^{K\times p}$.*

From the previous assumptions and by sub-scripting assets by i (assuming there are n assets), the following notations are defined: $B_i = B(\gamma_i)$, $C_i = C(\gamma_i)$, $Z_{i,t-1} = Z_{t-1}(\gamma_i)$. The initial model is rewritten as follows:

$$R_t^i = x_{it}^{'}\beta_i + \varepsilon_t^i \tag{A.2}$$

with

$$\begin{array}{lllllll}
x_{it}^{'} & = & \left(x_{1,it}^{'}, x_{2,it}^{'}\right) & \beta_i^{'} & = & \left(\beta_{1,i}^{'}, \beta_{2,i}^{'}\right) \\
x_{1,it}^{'} & = & \left([vech(X_t)]^{'}, Z_{t-1}^{'} \otimes Z_{i,t-1}^{'}\right) & \beta_{1,i}^{'} & = & \left([N_p(\nu^{'} \otimes I_p)vec(B_i^{'})]^{'}, [(\nu^{'} \otimes I_q)vec(C_i^{'})]^{'}\right) \\
x_{2,it}^{'} & = & \left(f_t^{'} \otimes Z_{t-1}^{'}, f_t^{'} \otimes Z_{i,t-1}^{'}\right) & \beta_{2,i}^{'} & = & \left([vec(B_i^{'})]^{'}, [vec(C_i^{'})]^{'}\right) \\
\nu & = & vec(\Lambda - F) & N_p & = & \frac{1}{2}D_p^{+}(I_{p^2} + W_p)
\end{array}$$

The duplication matrix D_p is the $p \times p$ matrix such that for every $p \times p$ matrix A, $D_p vech(A) = vec(A)$ (D_p^{+} is its Moore-Penrose inverse). The commutation matrix W_p is the $p^2 \times p^2$ matrix such that for every $p \times p$ matrix A, $vec(A^{'}) = W_p vec(A)$. X_t is a $p \times p$ matrix such that

$$(X_t)_{k,l} = \begin{cases} Z_{t-1,k}^2 & if \quad k = l \\ 2Z_{t-1,k} \cdot Z_{t-1,l} & if \quad k \neq l \end{cases}$$

The estimation procedure presented by GOS is based on a two-step estimation. At the first one, the sensitivity β_i is estimated from the equation (A.2) through an OLS regression. By defining

$\beta_{3,i}^{'} = \left([N_p(I_p \otimes B_i^{'})]^{'}; [(I_p \otimes C_i^{'})]^{'}\right)^{'}$, we have $\beta_{1,i} = \beta_{3,i}\nu$. Then at the second step, the parameter ν is estimated by a WLS regression of $\hat{\beta}_{2,i}$ on $\hat{\beta}_{3,i}\nu$. At the end, we have

$$vec\left[\hat{\Lambda}^{'}\right] = \hat{\nu} + vec\left[\hat{F}^{'}\right],$$

where $\hat{F}$ is obtained by a SURE regression of factors on instruments Z_{t-1}. The estimated risk premia at t is then given by

$$\hat{\lambda}_t = \hat{\Lambda} Z_{t-1}. \tag{A.3}$$

A1.2 GOS and dynamically useless factors

Let us take the simplified case with $k = 1$, $p = 1$ and $q = 0$; where k is the number of factors, p is the number of common instruments, and q is the number of specific instruments. So we just have one factor and there are no asset-specific instruments. We can use the setting $q = 0$ because the model presented by the authors has the classical time-invariant coefficients model as a particular case (this one corresponds to the setting $Z_t = 1$ and $Z_{i,t} = 0$ for all t); therefore, their model is still valid even when $q = 0$. With $k = 1$, $p = 1$ and $q = 0$, the model (A.2) becomes easier to write with $x_{it}^{'} = \left(Z_{t-1}^2; f_t Z_{t-1}\right)$, $\beta_i^{'} = (B_i\nu; B_i)$ and $f_t, Z_{t-1}, B_i \in \mathbb{R}$. Here $\beta_{2i} = \beta_{3i} = B_i$.

Let us suppose firstly that the considered factor is dynamically strongly useless at time period t_0. The sensitivity of the asset i w.r.t this factor at t_0 is $b_{t_0}^i = 0$. From Assumption FS.1, it follows that for any asset i, $B_i = 0$. Let us consider now any time period t, which is not necessarily a period where the factor is useless. From the first step regression, we obtain $\hat{\beta}_i^{'} = (\hat{\beta}_{1i}; \hat{\beta}_{2i})^{'}$, and then perform the second step regression on the model $\hat{\beta}_{1i} = \hat{\beta}_{3i}\nu + \zeta_i$. For simplicity, we estimate ν through an OLS estimation, so we have $\hat{\nu} = \frac{\frac{1}{n}\sum_i \hat{\beta}_{3i}\hat{\beta}_{1i}}{\frac{1}{n}\sum_i \hat{\beta}_{3i}^2}$ (we can use the OLS instead of the WLS as GOS, since we just look at the consistency of the final estimator). From Lemma 3 of GOS, we have as $n, T \to \infty$ (with $n = O(T^{\gamma}), \gamma > 0$),

$$\frac{1}{n}\sum_i \hat{\beta}_{3i}^2 = \frac{1}{n}\sum_i \beta_{3i}^2 + o_p(1) = \frac{1}{n}\sum_i B_i^2 + o_p(1) = o_p(1). \tag{A.4}$$

We obtain the same result if the factor is dynamically weakly useless. Indeed in that case for all assets i, $B_i = \frac{C_i}{\sqrt{T}}$ (with C_i non-null at least for one asset). So when n and T go to infinity, we will have from Lemma 3 of GOS as $n, T \to \infty$ (with $n = O(T^{\gamma}), \gamma > 0$),

$$\frac{1}{n}\sum_i \hat{\beta}_{3i}^2 = \frac{1}{n}\sum_i \beta_{3i}^2 + o_p(1) = \frac{1}{nT}\sum_i C_i^2 + o_p(1) = o_p(1). \tag{A.5}$$

Therefore as $T \to \infty$, the model displays an asymptotic multicollinearity on the dependent variable; so we cannot say anything about the consistency of $\hat{\nu}$. Having a dynamically useless factor leads to singular loadings matrix and therefore to identification issues for the GOS estimator.

A2 Robustness checks

We perform two exercises for robustness checks. In the first one, we add an additional instrument to the two initial instruments. In the second exercise, we keep the instruments as in the paper, but we change the frequency of periods from the month to the quarter.

A2.1 Additional instrument

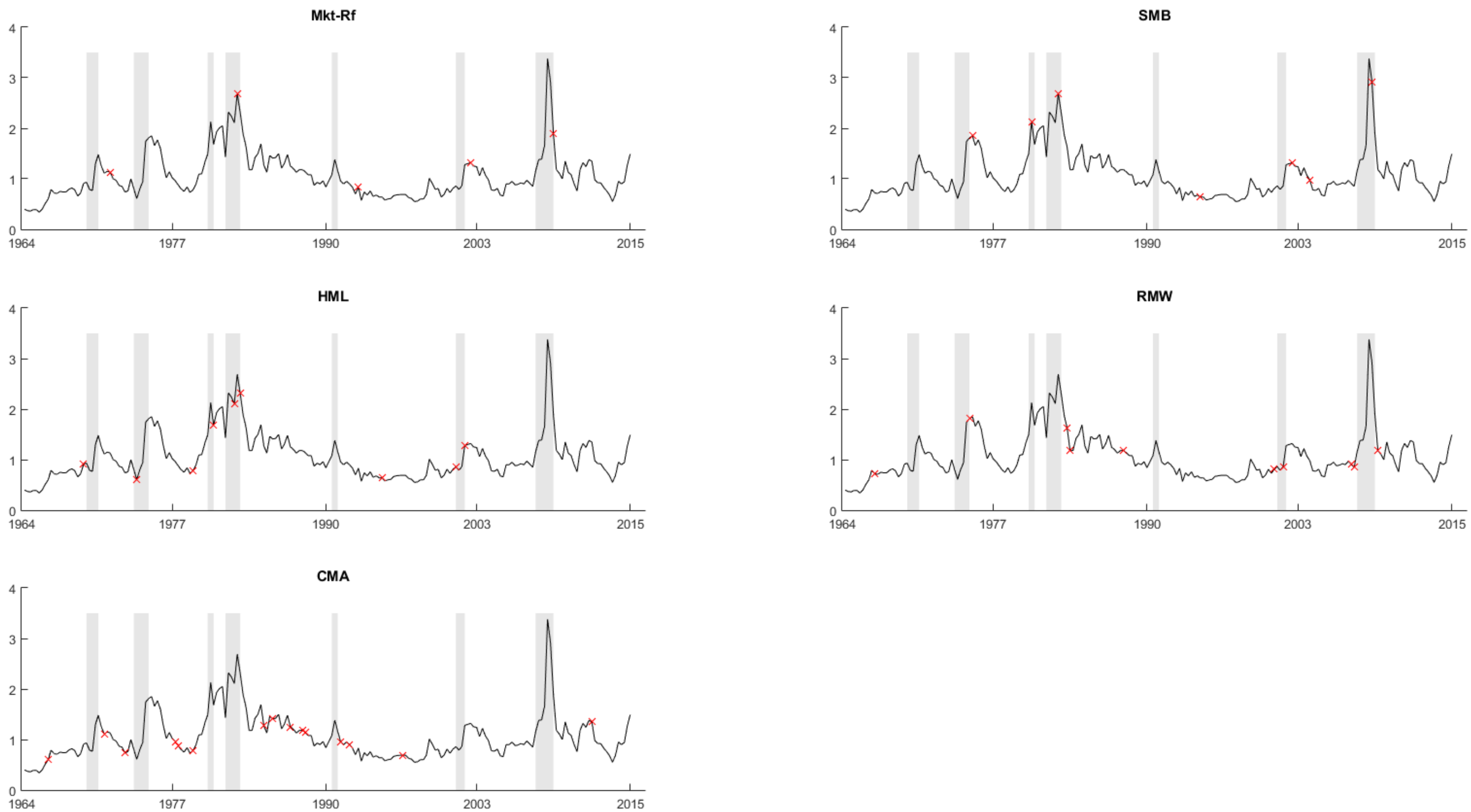
We add the dividend yield[13] as third instrument and repeat the empirical analysis. It appears that the results are almost the same as those we had previously (see table A.1). They are strongly robust to the addition of the third instrument, as the gap between the later and the former results for the risk premia and for the standards errors, is of order of 0.01%.

A2.2 Quarterly analysis

We perform a quarterly analysis, to check if the qualitative results about the profitability and the investment factors are robust to a variation of the period frequency. Figure A.1 shows that the conclusion we had on those factors still hold. The risk premium of the profitability factor RMW (Robust Minus Weak) is never insignificant during crisis, and the risk premium of the investment factor CMA (Conservative Minus Aggressive) is mostly insignificant in quarters where the dynamic of the default spread changes sign.

[13]The dividend yield is also used by Gagliardini et al. (2016) as common instrument for their robustness checks. Our data are from the database of the Federal Reserve Bank of St-Louis

Figure A.1: Default spread and dates where the risk premium of the factor is zero (quarterly analysis)



The figure presents the evolution of the quarterly default spread from 1964 to 2015; and for every factor, the dates where its associated risk premium is insignificant (red crosses). The shaded bands represent the US recessions, according to the data of the National Bureau of Economic Research (NBER)

Table A.1: Empirical risk premia with the dividend yield as additional instrument

		risk premium (×100)					std. error (×100)					insignificant risk premia (number of months over 617)
			percentile					percentile				
		mean	20	40	60	80	mean	20	40	60	80	
Fama-French (1993)	Mkt-Rf	-0.16	-2.94	-1.04	0.59	2.50	0.28	0.15	0.23	0.31	0.39	29 (0.047)
	SMB	0.13	-0.58	-0.09	0.31	0.86	0.23	0.15	0.20	0.25	0.29	62 (0.100)
	HML	0.12	-0.92	-0.18	0.42	1.23	0.20	0.11	0.15	0.20	0.30	42 (0.068)
Carhart (1997)	Mkt-Rf	-0.15	-2.77	-0.97	0.50	2.38	0.33	0.18	0.27	0.34	0.44	27 (0.044)
	SMB	0.14	-0.58	-0.05	0.30	0.86	0.22	0.15	0.20	0.24	0.28	70 (0.113)
	HML	0.13	-0.88	-0.15	0.41	1.21	0.24	0.13	0.18	0.25	0.31	49 (0.079)
	Mom	0.11	-1.51	-0.37	0.55	1.57	0.35	0.18	0.26	0.32	0.47	41 (0.066)
Fama-French (2015)	Mkt-Rf	-0.11	-2.70	-1.06	0.55	2.37	0.25	0.13	0.22	0.27	0.33	27 (0.044)
	SMB	0.16	-0.54	-0.03	0.37	0.87	0.23	0.15	0.20	0.25	0.30	72 (0.117)
	HML	0.12	-0.92	-0.16	0.46	1.19	0.25	0.15	0.20	0.25	0.35	45 (0.073)
	RMW	0.06	-0.86	-0.23	0.27	0.96	0.13	0.07	0.11	0.13	0.19	30 (0.049)
	CMA	0.08	-0.74	-0.17	0.35	0.94	0.21	0.13	0.17	0.21	0.29	46 (0.075)

The table gives the dynamics of the risk premia in the chosen models. The means and the percentiles are estimated over the T dates ($T = 617$). The last column gives the number of months over the 617, where the risk premium of the factor is insignificant at the level of significance 5%.

The numbers in brackets are the ratios between the number of periods where the risk premium is insignificant, and the overall number of periods. They give the probability for each factor, of being unrewarded.

A3 Proofs

A3.1 Motivation for Assumption 6

By Proposition 1 of GOS, the result in Assumption 6 is true when we have a continuum of assets. We assume in the paper a large countable number of assets. So to motivate Assumption 6, we show here that the equation in Assumption 6 is still true when the number of assets is finite. Therefore, we consider here that n is finite.

Let us consider a month t where the two factors are useful. We also consider $B_t = (b_t^1, \cdots, b_t^n)^{'}$ the matrix of betas as defined previously, and $A_t = (a_t^1, \cdots, a_t^n)^{'}$. Following GOS, we define e_t as the residual of the orthogonal projection of A_t on B_t;

$$e_t = A_t - B_t(B_t^{'}B_t)^{-1}B_t^{'}A_t.$$

Let us assume by contradiction that the relation $a_t^i = b_t^i \nu_t$ does not hold for any $\nu_t \in \mathbb{R}^2$. So $\underset{\nu \in \mathbb{R}^2}{inf} \sum_{i=1}^{n} \left(a_t^i - (b_t^i)^{'} \nu \right)^2 > 0$, and then $\sum_{i=1}^{n} \left(a_t^i - (b_t^i)^{'} \nu_{t,\infty} \right)^2 > 0$ with $\nu_{t,\infty} = \left(\sum_{i=1}^{n} b_t^i (b_t^i)^{'} \right)^{-1} \left(\sum_{i=1}^{n} b_t^i a_t^i \right)$. Therefore $e_t^{'} e_t > 0$ and $\|e_t\|^2 > 0$.

Following GOS, we define the portfolio $q_t = \left(\delta_{0,t} \vdots \delta_t^{'} \right)$, with $\delta_t = \dfrac{e_t}{\|e_t\|^2}$, $\delta_{0,t} = -i_n^{'}\delta_t$, and i_n the n-vector of ones (we represent a portfolio here with the vector of the weights on the risk-free asset and on the n risky assets). Then $C(q_t) = 0$ and following their arguments,

$$P(q_t > 0 | \mathscr{F}_0) \geq 1 - \left(n^{-1} eig_{max}(\Sigma_{\varepsilon,t}) \right) n \|e_t\|^{-2}$$

As the two factors are useful, Assumption APR.2 from GOS is satisfied. So following their arguments, $\frac{1}{n}\|e_t\|^2 \overset{L_2}{\to} \gamma$, with $\gamma \neq 0$. Then by Assumption 3, it follows that $P(q_t > 0|\mathscr{F}_0) \overset{L_2}{\to} 1$; so $P(q_t > 0) \to 1$. Therefore we have a contradiction with Assumption 4, since (q_t) displays an asymptotic arbitrage opportunity. So we conclude that $a_t^i = (b_t^i)^{'} \nu_{t,\infty}$, with $\nu_{t,\infty}$ uniquely defined and $\mathscr{F}_{t-1}$-measurable. ■

A3.2 Motivation for Assumption 7

As previously, we assume that the number of assets n is finite and show that the equation in Assumption 7 is true in that case.

If the factor $\bar{us}$ is strongly useless in mouth t, then $b_t^{i,\bar{us}} = 0$ for all assets i. In that case, the initial model (1.1) becomes

$$R_t^i = a_t^i + b_t^{i,us} f_t^{us} + \varepsilon_t^i, \tag{A.6}$$

and then we conclude following the same argument as previously (since the factor us is useful).

If the factor $\bar{us}$ is weakly useless in mouth t then as T is large, there is for any assets i, a scalar c_t^i such that $b_t^{i,\bar{us}} = \frac{c_t^i}{\sqrt{T}}$. The model (1.1) becomes then,

$$\begin{aligned} R_t^i =& \phi_t^0 + a_t^i + b_t^{i,us} f_t^{us} + b_t^{i,\bar{us}} f_t^{\bar{us}} + \varepsilon_t^i \\ =& \phi_t^0 + \left(a_t^i + \frac{c_t^i}{\sqrt{T}} f_t^{\bar{us}} \right) + b_t^{i,us} f_t^{us} + \varepsilon_t^i \end{aligned} \tag{A.7}$$

Let us consider the following model, where for each month t and for any asset i, l_t^i is a given scalar and ξ_t^i is a random variable with the same moments assumptions as ε_t^i (see Assumption 1). a_t^i, $b_t^{i,us}$ and f_t^{us} are also defined as in Assumption 1.

$$R_t^i = \phi_t^0 + \left(a_t^i + \frac{l_t^i}{\sqrt{T}} \right) + b_t^{i,us} f_t^{us} + \xi_t^i \ . \tag{A.8}$$

The conditional covariance matrix of model (A.8) is $\Sigma_{\xi,t} = cov(\xi_t^i, \xi_t^j | \mathscr{F}_{t-1})$. As $T \to \infty$, we have $\Sigma_{\xi,t} \equiv \Sigma_{\varepsilon,t} + o_p(1)$; so $n^{-1} eig_{max}(\Sigma_{\xi,t}) \overset{L_2}{\to} 0$ as $n \to \infty$. So by the same argument as in the motivation for Assumption 6, there is an unique $\mathscr{F}_{t-1}$-measurable random variable $\nu_t^{us} \in \mathbb{R}$ such that $a_t^i + \frac{l_t^i}{\sqrt{T}} = b_t^{i,us} \nu_t^{us}$. Therefore, $a_t^i = b_t^{i,us} \nu_t^{us} + O_p(\frac{1}{\sqrt{T}})$. ■

In the next proofs, we use the following result;

Result 1. *When the number of assets n is going to infinity,*
(i) If the factor $\bar{us}$ is useless at t, then $\|\hat{B}_t^{\bar{us}}\| = O_p(\frac{1}{\sqrt{T}})$, as $T \to \infty$
(ii) If the factor $\bar{us}$ is useful at t, then $\|\hat{B}_t^{\bar{us}}\|$ converges to a positive number different from 0 ($\|\hat{B}_t^{\bar{us}}\| = O_p(1)$).

Proof of Result 1

We follow here the argument of Bryzgalova (2016). As $\hat{B}_t^{\bar{us}}$ is a consistent estimator of $B_t^{\bar{us}}$, we have:

$$\hat{B}_t^{\bar{us}} = B_t^{\bar{us}} + \frac{1}{\sqrt{T}} \Psi_{\hat{B}_t^{\bar{us}}} + o_p(\frac{1}{\sqrt{T}}), \tag{A.9}$$

with $\Psi_{\hat{B}_t^{\bar{us}}}$, a random variable such that $\Psi_{\hat{B}_t^{\bar{us}}} \overset{a}{\sim} \mathcal{N}(0, \Sigma_{\hat{B}_t^{\bar{us}}})$. The result follows by Assumption 10. □

A3.3 Proof of Proposition 1

We look at the properties of $\hat{\phi}_t$ depending on whether the dynamically useless factor $\bar{us}$ is useless or not at t. In the following parts of the proofs, we denote $(\beta_t^i)' = (1 \vdots (b_t^i)')$.

We have $\hat{\Phi}_t = \begin{pmatrix} \hat{\phi}_t^0 \\ \hat{\phi}_t \end{pmatrix} = \underset{\Phi}{argmin} \ L_{T,n,D_t}(\Phi)$, with L_{T,n,D_t} convex for every T, n, D_t in $\mathbb{N}^*$ and given by:

$$L_{T,n,D_t}(\Phi) = \sum_{i=1}^{n} \left(R_t^i - (\beta_t^i)^{'} \Phi \right)^2 + \frac{\tau_n}{T^{s/2}} \sum_{j=1}^{2} \frac{|\phi^j|}{\|\hat{B}_t^j\|^s}.$$

Therefore, $\hat{\Phi}_t = \underset{\Phi}{argmin}\ L_{T,n,D_t}(\Phi) = \underset{\Phi}{argmin}\ \frac{1}{n} L_{T,n,D_t}(\Phi)$. For the next, we consider the following functions:

$$L_{n,D_t}(\Phi) \equiv \frac{1}{n} \sum_{i=1}^{n} \left(R_t^i - (\beta_t^i)^{'} \Phi \right)^2 \qquad L_{T,n}(\Phi) \equiv \frac{\tau_n}{nT^{s/2}} \sum_{j=1}^{2} \frac{|\phi^j|}{\|\hat{B}_t^j\|^s}. \tag{A.10}$$

So we have

$$\hat{\Phi}_t = \underset{\Phi}{argmin} \left(L_{n,D_t}(\Phi) + L_{T,n}(\Phi) \right).$$

If the two factors are useful in month t

As the two factors are useful at t, then following the result 1 and Assumption 9 we have: $\|\hat{B}_t^j\| = O_p(1)$ as $n \to \infty$ $(j = 1, 2)$. So as $n \to \infty$, we have for all Φ in $\mathbb{R}^3$;

$$L_{T,n}(\Phi) = o_p(1). \tag{A.11}$$

Let us now determine the p-limit of L_{n,D_t} as $n \to \infty$. For that purpose, we consider the following matrices: $R_t = \left(R_t^1, \ldots, R_t^n\right)^{'}$, $\beta_t = \left(\beta_t^1, \ldots, \beta_t^1\right)^{'}$ and $\varepsilon_t = \left(\varepsilon_t^1, \ldots, \varepsilon_t^n\right)^{'}$. We have;

$$\begin{aligned} L_{n,D_t}(\Phi) &= \frac{1}{n} \left(R_t - \hat{\beta}_t \Phi \right)^{'} \left(R_t - \hat{\beta}_t \Phi \right) \\ &= \frac{1}{n} \left(R_t - \beta_t \Phi - \hat{e}_t \Phi \right)^{'} \left(R_t - \beta_t \Phi - \hat{e}_t \Phi \right) \qquad (with\ \ \hat{e}_t = \hat{\beta}_t - \beta_t) \\ &= \frac{1}{n} \left[\left(R_t - \beta_t \Phi \right)^{'} \left(R_t - \beta_t \Phi \right) - 2 \left(R_t - \beta_t \Phi \right)^{'} \hat{e}_t \Phi + \Phi^{'} \hat{e}_t^{'} \hat{e}_t \Phi \right] \end{aligned}$$

We have $\hat{e}_t = \left(\hat{e}_t^1, \ldots, \hat{e}_t^n\right)^{'}$, with $\hat{e}_t^i$ a $\mathbb{R}^3$ vector such as $(\hat{e}_t^i)^{'} = (\hat{\beta}_t^i)^{'} - (\beta_t^i)^{'}$ for all asset i. Let us consider the following lemma

Lemma 1. *For each asset i; if we define $\hat{e}_t^i = \hat{\beta}_t^i - \beta_t^i$, with $\beta_t^i = (1 \vdots (b_t^i)^{'})^{'}$ (b_t^i the sensitivity of the asset i at time t), then under Assumptions 1-5 and 9 we have:*

$$\underset{i}{sup} \|\hat{e}_t^i\| = o_p(1) \quad as \quad n \to \infty. \tag{A.12}$$

By using the Cauchy-Schwarz inequality with the previous lemma, we have as $n \to \infty$:

$$\left| \frac{1}{n} \left(R_t - \beta_t \Phi \right)^{'} \hat{e}_t \Phi \right| \leq \frac{1}{n} \sum_i \left| \left(R_t^i - (\beta_t^i)^{'} \Phi \right) \right| \cdot |\hat{e}_t^i \Phi| \leq \underset{i}{sup} \|\hat{e}_t^i\| \left(\frac{1}{n} \sum_i \left| \left(R_t^i - (\beta_t^i)^{'} \Phi \right) \right| \cdot \|\Phi\| \right) = o_p(1).$$

$$\left|\frac{1}{n}\Phi\hat{e}_t^{'}\hat{e}_t\Phi\right| \leq \frac{1}{n}\sum_i |\Phi^{'}(\hat{e}_t^i)^{'}|\cdot|\hat{e}_t^i\Phi| \leq \|\Phi\|\left(\frac{1}{n}\sum_i \|(\hat{e}_t^i)^{'}\|\cdot\|\hat{e}_t^i\|\right)\|\Phi\| \leq \left(\underset{i}{sup}\|\hat{e}_t^i\|\right)^2\|\Phi\|^2 = o_p(1).$$

Moreover from the equation (1.3), $R_t^i = (\beta_t^i)^{'}\Phi_t + \varepsilon_t^i$. So by the assumption 11, we have as $n \to \infty$;

$$\frac{1}{n}\left(R_t - \beta_t\Phi\right)^{'}\left(R_t - \beta_t\Phi\right) = \left(\Phi - \Phi_t\right)^{'}\left(\frac{1}{n}\sum_i \beta_t^i(\beta_t^i)^{'}\right)(\Phi - \Phi_t) + \frac{1}{n}\sum_i(\varepsilon_t^i)^2 - 2(\Phi - \Phi_t)^{'}\left(\frac{1}{n}\sum_i \beta_t^i\varepsilon_t^i\right)$$
$$= (\Phi - \Phi_t)^{'}\tilde{Q}_t(\Phi - \Phi_t) + \sigma_t^2 + o_p(1).$$

Then it follows that as $n \to \infty$,

$$L_{n,D_t}(\Phi) = (\Phi - \Phi_t)^{'}\tilde{Q}_t(\Phi - \Phi_t) + \sigma_t^2 + o_p(1) \tag{A.13}$$

By equations (A.11) and (A.13), we have as $n \to \infty$,

$$L_{T,n,D_t}(\Phi) = (\Phi - \Phi_t)^{'}\tilde{Q}_t(\Phi - \Phi_t) + \sigma_t^2 + o_p(1) \tag{A.14}$$

Let us denote $L(\Phi) = (\Phi - \Phi_t)^{'}\tilde{Q}_t(\Phi - \Phi_t) + \sigma_t^2$. As $L_{T,n,D_t}(\Phi)$ is convex for all Φ in $\mathbb{R}^3$, we have by the convexity lemma of Pollard (1991); $\underset{\Phi}{sup}|L_{T,n,D_t}(\Phi) - L(\Phi)| = o_p(1)$. Therefore, $\hat{\Phi}_t = \underset{\Phi}{argmin}\ L_{T,n,D_t}(\Phi) \xrightarrow{p} \underset{\Phi}{argmin}\ L(\Phi) = \Phi_t$ ∎

If the dynamically useless factor $\bar{us}$ is useless in month t

In the present case $\|\hat{B}_t^1\| = O_p(1)$, $\|\hat{B}_t^2\| = O_p(\frac{1}{\sqrt{T}})$ (as $T \to \infty$) and $\Phi_t = (\phi_t^0, \phi_t^{us}, 0)^{'}$ by definition. Following Assumption 9, we have as $n \to \infty$

$$L_{T,n}(\Phi) = \tilde{\tau}_0|\phi^{\bar{us}}| \qquad (\tilde{\tau}_0 > 0). \tag{A.15}$$

In order to find the p-limit of L_{n,D_t}, we consider the matrices R_t and ε_t as defined previously. We also consider the matrices $\kappa_t = \left(\kappa_t^1, \ldots, \kappa_t^n\right)^{'}$ and $b_t^{\bar{us}} = \left(b_t^{1,\bar{us}}, \ldots, b_t^{n,\bar{us}}\right)^{'}$, with $\kappa_t^i = \left(1, b_t^{i,us}\right)^{'}$ for all i. In the next, we denote $\Phi^{us} = \left(\phi^0, \phi^{us}\right)^{'}$. We have

$$L_{n,D_t}(\Phi) = \frac{1}{n}\sum_{i=1}^n\left(R_t^i - (\hat{\beta}_t^i)^{'}\Phi\right)^2$$
$$= \frac{1}{n}\sum_{i=1}^n\left(R_t^i - \phi^0 - \hat{b}_t^{i,us}\phi^{us} - \hat{b}_t^{i,\bar{us}}\phi^{\bar{us}}\right)^2$$
$$= \frac{1}{n}\left(R_t - \hat{\kappa}_t\Phi^{us} - \hat{b}_t^{\bar{us}}\phi^{\bar{us}}\right)^{'}\left(R_t - \hat{\kappa}_t\Phi^{us} - \hat{b}_t^{\bar{us}}\phi^{\bar{us}}\right)$$

$$= \frac{1}{n}\left(R_t - \kappa_t\Phi^{us} - \tilde{e}_t\Phi^{us} - \hat{b}_t^{\bar{us}}\phi^{\bar{us}}\right)^{'}\left(R_t - \kappa_t\Phi^{us} - \tilde{e}_t\Phi^{us} - \hat{b}_t^{\bar{us}}\phi^{\bar{us}}\right) \qquad (with\ \tilde{e}_t = \hat{\kappa}_t - \kappa_t)$$
$$= \frac{1}{n}\Big[\left(R_t - \kappa_t\Phi^{us}\right)^{'}\left(R_t - \kappa_t\Phi^{us}\right) - 2\left(R_t - \kappa_t\Phi^{us}\right)^{'}\left(\tilde{e}_t\Phi^{us} + \hat{b}_t^{\bar{us}}\phi^{\bar{us}}\right)$$
$$+\left(\tilde{e}_t\Phi^{us} + \hat{b}_t^{\bar{us}}\phi^{\bar{us}}\right)^{'}\left(\tilde{e}_t\Phi^{us} + \hat{b}_t^{\bar{us}}\phi^{\bar{us}}\right)\Big].$$

Following Lemma 1, we have as $n \to \infty$, $\underset{i}{sup}\|\tilde{e}_t^i\| = o_p(1)$. Moreover as the factor $\bar{us}$ is useless at t, $\underset{i}{sup}|\hat{b}_t^{i,\bar{us}}| = o_p(1)$ by Assumption 10. So following the same argument as previously, we have

$$\frac{1}{n}\left[-2\left(R_t - \kappa_t\Phi^{us}\right)^{'}\left(\tilde{e}_t\Phi^{us} + \hat{b}_t^{\bar{us}}\phi^{\bar{us}}\right) + \left(\tilde{e}_t\Phi^{us} + \hat{b}_t^{\bar{us}}\phi^{\bar{us}}\right)^{'}\left(\tilde{e}_t\Phi^{us} + \hat{b}_t^{\bar{us}}\phi^{\bar{us}}\right)\right] = o_p(1).$$

Since $\Phi_t = (\phi_t^0, \phi_t^{us}, 0)^{'}$, we have by the equation (1.5) $R_t^i = \pi_T^i + (\beta_t^i)^{'}\Phi_t + \varepsilon_t^i = \pi_T^i + (\kappa_t^i)^{'}\Phi_t^{us} + \varepsilon_t^i$ (with π_T^i an $O_p(\frac{1}{\sqrt{T}})$-term). Therefore by Assumptions 9 and 11 and following the same argument as previously,

$$L_{n,D_t}(\Phi) = (\Phi^{us} - \Phi_t^{us})^{'}\tilde{Q}_t^{us}(\Phi^{us} - \Phi_t^{us}) + \sigma_t^2 + o_p(1). \tag{A.16}$$

By equations (A.15) and (A.16), we have as $n \to \infty$,

$$L_{T,n,D_t}(\Phi) = (\Phi^{us} - \Phi_t^{us})^{'}\tilde{Q}_t^{us}(\Phi^{us} - \Phi_t^{us}) + \sigma_t^2 + \tilde{\tau}_0|\phi^{\bar{us}}| + o_p(1) \tag{A.17}$$

Let us denote $L(\Phi) = (\Phi^{us} - \Phi_t^{us})^{'}\tilde{Q}_t^{us}(\Phi^{us} - \Phi_t^{us}) + \sigma_t^2 + \tilde{\tau}_0|\phi^{\bar{us}}|$. We have $\underset{\Phi}{argmin}\ L(\Phi) = ((\Phi_t^{us})^{'}, 0)^{'} = (\phi_t^0, \phi_t^{us}, 0)^{'} = \Phi_t$. As L_{T,n,D_t} is convex, we can conclude following Pollard (1991) that $\hat{\Phi}_t \xrightarrow{p} \Phi_t$ as $n \to \infty$. ■

Proof of Lemma 1

As $\hat{e}_t^i$ is a $\mathbb{R}^3$ vector, let us denote for all asset i, the scalar k_i such that $|\hat{e}_t^{i,k_i}| = \underset{k=1,2,3}{sup}|\hat{e}_t^{i,k}|$ (where $\hat{e}_t^{i,k}$ is the k^{th} element of $\hat{e}_t^i$). As all the norms in $\mathbb{R}^3$ are equivalent, there is a scalar $c > 0$ such that $\underset{i}{sup}\|\hat{e}_t^i\| \leq c|\hat{e}_t^{i,k_i}|$ for at least one i.

We have $\hat{e}_t^i = \hat{\beta}_t^i - \beta_t^i$ for all asset i. As $\hat{\beta}_t^i$ is consistent, $\underset{D_t \to \infty}{lim} P\left(\|\hat{e}_t^i\| > \xi\right) = 0$ for all $\xi > 0$. For all $\xi > 0$, we have by the Markov inequality;

$$
\begin{aligned}
P\left(\underset{i}{sup}\|\hat{e}_t^i\| \geq \xi\right) &\leq \sum_{i=1}^{n} P\left(c|\hat{e}_t^{i,k_i}| \geq \xi\right) \\
&\leq \sum_{i=1}^{n} P\left(\hat{e}_t^{i,k_i} \geq \xi/c\right) \mathbb{1}_{\{\hat{e}_t^{i,k_i}>0\}} + \sum_{i=1}^{n} P\left(-\hat{e}_t^{i,k_i} \geq \xi/c\right) \mathbb{1}_{\{\hat{e}_t^{i,k_i}<0\}} \\
&\leq \sum_{i=1}^{n} \frac{E(\hat{e}_t^{i,k_i})}{\xi/c} \mathbb{1}_{\{\hat{e}_t^{i,k_i}>0\}} + \sum_{i=1}^{n} \frac{E(-\hat{e}_t^{i,k_i})}{\xi/c} \mathbb{1}_{\{\hat{e}_t^{i,k_i}<0\}} \\
&\leq \frac{c}{\xi} \sum_{i=1}^{n} \left|E(\hat{e}_t^{i,k_i})\right| \mathbb{1}_{\{\hat{e}_t^{i,k_i}>0\}} + \frac{c}{\xi} \sum_{i=1}^{n} \left|E(-\hat{e}_t^{i,k_i})\right| \mathbb{1}_{\{\hat{e}_t^{i,k_i}<0\}} \\
&\leq \frac{c}{\xi} \sum_{i=1}^{n} \left|E(\hat{e}_t^{i,k_i})\right| \\
&\leq \frac{c}{\xi} \sum_{i=1}^{n} \left|E(E(\hat{e}_t^{i,k_i}|f_{t,d}))\right| .
\end{aligned} \tag{A.18}
$$

By definition, $(\beta_t^i)' = (1 \vdots (b_t^i)')$ for all i; moreover, we have

$$
\hat{b}_t^i - b_t^i = \left[\frac{1}{D_t}\sum_d (f_{t,d} - \bar{f}_t)(f_{t,d} - \bar{f}_t)'\right]^{-1} \left[\frac{1}{D_t}\sum_d (f_{t,d} - \bar{f}_t)\eta_{t,d}^i\right].
$$

$\hat{b}_t^i$ is obtained through an OLS regression on daily data according to the model (1.2). Following Assumption 2, we have $E(\eta_{t,d}^i|f_{t,d}) = 0$. Therefore,

$$
\begin{aligned}
E(\hat{b}_t^i - b_t^i|f_{t,d}) &= \left[\frac{1}{D_t}\sum_d (f_{t,d} - \bar{f}_t)(f_{t,d} - \bar{f}_t)'\right]^{-1} \left[\frac{1}{D_t}\sum_d (f_{t,d} - \bar{f}_t)E(\eta_{t,d}^i|f_{t,d})\right] \\
&= 0
\end{aligned} \tag{A.19}
$$

By the definition of β_t^i, we then have $E(\hat{e}_t^i|f_{t,d}) = E(\hat{\beta}_t^i - \beta_t^i|f_{t,d}) = 0$ for all assets i; therefore, $E(\hat{e}_t^{i,k_i}|f_{t,d}) = 0$ for all i.

We know from Assumption 9 that D_t becomes large as soon as n becomes large. So as $n \to \infty$, we have from equation (A.18) and for all $\xi > 0$,

$$
\underset{n\to\infty}{lim}\ P\left(\underset{i}{sup}\|\hat{e}_t^i\| \geq \xi\right) \leq \underset{n\to\infty}{lim}\ \left(\frac{c}{\xi}\sum_{i=1}^{n}\left|E(E(\hat{e}_t^{i,k_i}|f_{t,d}))\right|\right) = 0\ .
$$

□

A3.4 Proof of Proposition 2

As previously we consider a given month t and proceed following whether the dynamically useless factor $\bar{us}$ is useless or not at t.

If the two factors are useful in month t

We are going first to present the asymptotic properties of the estimator $\hat{F}$. We have $f_t^{'} = Z_{t-1}^{'}F^{'} + u_t^{'}$ (by transposing the initial relation). Then, $\hat{F} - F = \left(\frac{1}{T}\sum_{t=1}^{T} u_t Z_{t-1}^{'}\right)\left(\frac{1}{T}\sum_{t=1}^{T} Z_{t-1}Z_{t-1}^{'}\right)^{-1}$ by a SURE regression. Therefore we have:

$$vec(\hat{F}) - vec(F) = \left(\left[\frac{1}{T}\sum_{t=1}^{T} Z_{t-1}Z_{t-1}^{'}\right]^{-1} \otimes I_2\right)\left(\frac{1}{T}\sum_{t=1}^{T} Z_{t-1} \otimes u_t\right)$$

By applying the CLT with Assumption 12, we have as $T \to \infty$

$$\sqrt{T}\left(vec(\hat{F}) - vec(F)\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_F\right). \tag{A.20}$$

Moreover, we have $\lambda_t = \phi_t + FZ_{t-1} - f_t = \phi_t + (Z_{t-1}^{'} \otimes I_2)vec(F) - f_t$. So now, we consider the following lemma.

Lemma 2. *If the two factors are useful in month t, we have as $n \to \infty$*

$$\sqrt{n}(\hat{\phi}_t - \phi_t) = O_p(1).$$

Using Lemma 2 and Assumption 9, we have as $n \to \infty$,

$$\begin{aligned} n^{\alpha}(\hat{\lambda}_t - \lambda_t) &= \left(\frac{1}{n^{0.5-\alpha}}\right)\sqrt{n}(\hat{\phi}_t - \phi_t) + \left(\frac{n^{\alpha}}{\sqrt{T}}\right)(Z_{t-1}^{'} \otimes I_2)\left(\sqrt{T}\left(vec(\hat{F}) - vec(F)\right)\right) \\ &\overset{a}{\sim} O_p\left(\frac{1}{n^{0.5-\alpha}}\right) + \rho_1^{1/2}\mathcal{N}\left(0, (Z_{t-1}^{'} \otimes I_2)\Sigma_F(Z_{t-1} \otimes I_2)\right) \\ &\xrightarrow{d} \mathcal{N}\left(0, \rho_1(Z_{t-1}^{'} \otimes I_2)\Sigma_F(Z_{t-1} \otimes I_2)\right) \end{aligned} \tag{A.21}$$

This ends the proof. ■

It should be noted that by equation (A.21), it appears that the smaller α is, the more precise the asymptotic convergence will be. So there is a trade-off to make on the estimator, between its convergence rate and its asymptotic bias. The reduction of the asymptotic bias is performed at the expense of the convergence rate. In the paper, we have chosen to reduce the convergence rate as much as possible.

Proof of Lemma 2

Our proof follows the arguments of Knight & Fu (2000), as they propose a framework for asymptotic analysis on lasso-type estimators. We have $\Phi_t = (\phi_t^0, (\phi_t)^{'})^{'}$ and $(\beta_t^i)^{'} = (1\vdots(b_t^i)^{'})$. To prove Lemma 2, it is sufficient to show that as $n \to \infty$, $v = \sqrt{n}(\hat{\Phi}_t - \Phi_t)$ is $O_p(1)$.
We have the following function;

$$L_{T,n,D_t}(\Phi) = \sum_{i=1}^{n} \left(R_t^i - (\hat{\beta}_t^i)^{'}\Phi\right)^2 + \frac{\tau_n}{T^{s/2}} \sum_{j=1}^{2} \frac{|\phi^j|}{\|\hat{B}_t^j\|^s}.$$

Let us consider $v_t = \sqrt{n}(\hat{\Phi}_t - \Phi_t)$. We also consider the vectors R_t, β_t, ε_t and $\hat{e}_t$ as defined previously. We have as equation (1.3), $R_t = \beta_t\Phi_t + \varepsilon_t$; therefore by straightforward calculus,

$$\begin{aligned} L_{T,n,D_t}(\hat{\Phi}_t) &= \left(\varepsilon_t - \beta_t\frac{v_t}{\sqrt{n}} - \hat{e}_t(\Phi_t - \frac{v_t}{\sqrt{n}})\right)^{'} \left(\varepsilon_t - \beta_t\frac{v_t}{\sqrt{n}} - \hat{e}_t(\Phi_t - \frac{v_t}{\sqrt{n}})\right) + \frac{\tau_n}{T^{s/2}} \sum_{j=1}^{2} \frac{|\phi_t^j + \frac{v_t^j}{\sqrt{n}}|}{\|\hat{B}_t^j\|^s} \\ &= A_n(v_t) + B_n(v_t) \\ &\equiv H_n(v_t). \end{aligned}$$

$v_t = \sqrt{n}(\hat{\Phi}_t - \Phi_t)$ minimizes $H_n(v)$; so it minimizes also $H_n(v) - H_n(0) = (A(v) - A(0)) + (B(v) - B(0))$. As for all i, $\hat{e}_t = o_p(1)$ when $n \to \infty$;

$$\begin{aligned} A_n(v) - A_n(0) &= \left(\varepsilon_t - \beta_t\frac{v}{\sqrt{n}} - \hat{e}_t(\Phi_t - \frac{v}{\sqrt{n}})\right)^{'} \left(\varepsilon_t - \beta_t\frac{v}{\sqrt{n}} - \hat{e}_t(\Phi_t - \frac{v}{\sqrt{n}})\right) - \varepsilon_t^{'}\varepsilon_t \\ &= \left(\varepsilon_t - \beta_t\frac{v}{\sqrt{n}}\right)^{'} \left(\varepsilon_t - \beta_t\frac{v}{\sqrt{n}}\right) - \varepsilon_t^{'}\varepsilon_t + o_p(1) \\ &= -\frac{2}{\sqrt{n}}v^{'}\beta_t^{'}\varepsilon_t + \frac{1}{n}v^{'}\beta_t^{'}\beta_t v + o_p(1) \\ &= -2v^{'}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta_t^i\varepsilon_t^i\right) + v^{'}\left(\frac{1}{n}\sum_{i=1}^{n}\beta_t^i(\beta_t^i)^{'}\right)v + o_p(1) \\ &\xrightarrow{d} -2v^{'}\tilde{A} + v^{'}\tilde{Q}_t v, \end{aligned} \tag{A.22}$$

where $\tilde{A}$ is a variable such as $\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta_t^i\varepsilon_t^i\right) \xrightarrow{d} \tilde{A}$, and $\tilde{Q}_t$ is defined as in Assumption 11. Moreover because the two factors are useful, we have as $n \to \infty$;

$$B_n(v) - B_n(0) = \frac{\tau_n}{T^{s/2}} \sum_{j=1}^{2} \frac{1}{\|\hat{B}_t^j\|^s} \left(|\frac{v^j}{\sqrt{n}} + \phi_t^j| - |\phi_t^j| \right)$$
$$= \frac{\tau_n}{\sqrt{n}} \frac{1}{T^{s/2}} \sum_{j=1}^{2} \frac{1}{\|\hat{B}_t^j\|^s} \left(v^j sign(\phi_t^j) \mathbb{1}_{\phi_t^j \neq 0} + |v^j| \mathbb{1}_{\phi_t^j} \right) \tag{A.23}$$
$$\xrightarrow{p} 0$$

We have $H_n(v) - H_n(0) \xrightarrow{d} -2v^{'}\tilde{A} + v^{'}\tilde{Q}_t v \equiv H(v)$. As H_n and H are convex, then we conclude following the argument of Knight & Fu (2000), that $\sqrt{n}(\hat{\Phi}_t - \Phi_t) = \underset{v}{argmin}\ (H_n(v) - H_n(0)) \xrightarrow{d} \underset{v}{argmin}\ H(v) = \tilde{Q}_t^{-1}\tilde{A}$. □

If the dynamically useless factor $\bar{us}$ is useless in month t

Since for all t, $f_t = FZ_{t-1} + u_t$, we also have $f_t^{us} = F^1 Z_{t-1} + u_t^{us}$ (with F^1 the first row of the matrix F). So $(\hat{F}^1 - F^1)^{'} = \left(\frac{1}{T} \sum_{t=1}^{T} Z_{t-1} Z_{t-1}^{'} \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} Z_{t-1} u_t^{us} \right)$. As previously by the CLT and Assumption 12, we have

$$\sqrt{T}(\hat{F}^1 - F^1)^{'} \xrightarrow{d} \mathcal{N}(0, \Sigma_F^{us}) \tag{A.24}$$

Now let us consider the following lemma;

Lemma 3. *If the dynamically useless factor $\bar{us}$ is useless in month t, we have as $n \to \infty$*

$$\sqrt{n}(\hat{\phi}_t - \phi_t) = O_p(1).$$

Following the same argument as previously (see the case A3.4), we conclude using the previous lemma. ■

Proof of Lemma 3

Let us consider the vectors R_t, Φ^{us}, κ_t, $\beta_t^{\bar{us}}$, $\tilde{e}_t$ and ε_t defined as previously. We also consider the vectors $\pi_T = \left(\pi_T^1, \ldots, \pi_T^n\right)^{'}$ [14] and $v_t = \sqrt{n}(\hat{\Phi}_t - \Phi_t)$. Following the equation (1.5), v_t is the minimum of $H_n(v) = A_n(v) + B_n(v)$ with;

$$A_n(v) = \left(\pi_T + \varepsilon_t - \kappa_t \frac{v^{us}}{\sqrt{n}} - \tilde{e}_t(\Phi_t^{us} + \frac{v^{us}}{\sqrt{n}}) - \hat{b}_t^{\bar{us}}(\phi_t^{\bar{us}} + \frac{v^2}{\sqrt{n}}) \right)^{'} \left(\pi_T + \varepsilon_t - \kappa_t \frac{v^{us}}{\sqrt{n}} - \tilde{e}_t(\Phi_t^{us} + \frac{v^{us}}{\sqrt{n}}) - \hat{b}_t^{\bar{us}}(\phi_t^{\bar{us}} + \frac{v^2}{\sqrt{n}}) \right)$$

[14] with π_T^i an $O_p(\frac{1}{\sqrt{T}})$-term for each i

$$B_n(v) = \frac{\tau_n}{T^{s/2}} \sum_{j=1}^{2} \frac{|\phi_t^j + \frac{v^j}{\sqrt{n}}|}{\|\hat{B}_t^j\|^s}.$$

We have here $v = (v^0, v^1, v^2)' = (v^0, v^{us}, v^{\bar{us}})'$ and $V^{us} = (v^0, v^{us}) = \sqrt{n}(\hat{\Phi}^{us} - \Phi_t^{us})$. Moreover by Assumption 9 we have as $n \to \infty$, $\hat{b}_t^{\bar{us}} = o_p(1)$, $\tilde{e}_t = o_p(1)$, and $\pi_T = o_p(1)$. Therefore,

$$A_n(v) - A_n(0) \xrightarrow{d} -2(V^{us})'\tilde{A}_2 + (V^{us})'\tilde{Q}_t^{us} V^{us},$$

where $\tilde{A}_2$ is a variable such as $\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \kappa_t^i \varepsilon_t^i\right) \xrightarrow{d} \tilde{A}_2$, and $\tilde{Q}_t^{us}$ is defined as in Assumption 11. Moreover because the factor $\bar{us}$ is useless at t, we have as $n \to \infty$;

$$\begin{aligned} B_n(v) - B_n(0) &= \frac{\tau_n}{T^{s/2}} \sum_{j=1}^{2} \frac{1}{\|\hat{B}_t^j\|^s} \left(|\frac{v^j}{\sqrt{n}} + \phi_t^j| - |\phi_t^j| \right) \\ &= \frac{\tau_n}{\sqrt{n}} \frac{1}{T^{s/2}} \sum_{j=1}^{2} \frac{1}{\|\hat{B}_t^j\|^s} \left(v^j sign(\phi_t^j) \mathbb{1}_{\phi_t^j \neq 0} + |v^j| \mathbb{1}_{\phi_t^j} \right) \\ &\xrightarrow{p} \tilde{\tau}_0 |v^{\bar{us}}|. \qquad (\tilde{\tau}_0 > 0) \end{aligned} \tag{A.25}$$

Let us denote $H(v) = -2(V^{us})'\tilde{A}_2 + (V^{us})'\tilde{Q}_t^{us} V^{us} + \tau_0 |v^2|$. As $v_t = \sqrt{n}(\hat{\Phi}_t - \Phi_t)$ minimizes $H_n(v) - H_n(0)$, we have following the same argument as previously,

$$\sqrt{n}(\hat{\Phi}_t - \Phi_t) = \underset{v}{argmin} \ (H_n(v) - H_n(0)) \xrightarrow{d} \underset{v}{argmin} \ H(v)$$

. □

Chapter 2

Dynamical relevance of factors in misspecified asset pricing models*

2.1 Introduction

In the asset pricing literature, linear factor models have been used extensively to explain asset returns. Since the initial CAPM,[1] various factor models have been developed to overcome the drawbacks from the empirical failures observed in some models. However, the debate regarding which model is the best remains inconclusive; as the empirical performance of an asset pricing model depends on the specificities of the market from which the data are collected, and on the properties of the assets to which the model is applied. The multiplicity of proposed factors (and therefore of models) is another problem in this regard, as there are more than 300 factors that have been declared significant over the years (see Harvey, Liu & Zhu, 2016).

Nevertheless, the literature has established some principles on which we can rely. From the works of Bollerslev et al. (1988), Ferson & Harvey (1991) or Jagannathan & Wang (1996), we know that alphas and betas are time-varying. This time variation can be analyzed through conditional models or approximated through moving rolling-window estimations. As alphas and betas measure systematic risk of assets in comparison to risk factors, having time-varying

*I am grateful to René Garcia for his invaluable guidance. For useful suggestions and comments, I would like to thank Bertille Antoine, Prosper Dovonon, Jean-Marie Dufour, and the participants of the CEA 2018.

[1]See Sharpe (1964) and Lintner (1965).

alphas and betas is a sufficient motivation for a dynamical analysis of the parameters of stochastic discount factor models (hereafter SDF models).[2] Nagel & Singleton (2011) perfectly address this issue by proposing an optimal GMM estimator for conditional SDF models. However, they do not correct for misspecifications, induced by discrepancies between the priced SDFs and the admissible ones.

Kan & Robotti (2009) and Kan et al. (2013) have recently shown that inference on SDF models should be performed through misspecification-robust processes. Indeed, as models are approximations of reality, there are always some differences between real and empirical SDFs. These differences have to be taken into account, otherwise some irrelevant factors might be mistakenly identified as priced (see Gospodinov et al., 2014). Note that the three papers just mentioned above are built on frameworks in which the SDF parameters are not time-varying. This is clearly a limitation since strategies on markets are conditioned by available information and vary through time and over business cycles. Therefore, having a conditional SDF with time-varying parameters is a more realistic specification. However, this implies a more challenging misspecification issue, as pricing errors are more severe for conditional models than for unconditional models (see Ghysels, 1998). This paper intends to develop a misspecification-robust inference method for conditional SDF parameters.

More specifically, we infer the time-varying parameters of potentially misspecified SDF models. The misspecification here refers to issues that lead to a divergence between the actual SDF and the chosen one. We extend the model of Gospodinov et al. (2014) to the framework of conditional models, as the parameters and the covariances are allowed to vary over time. Misspecifications are measured with the Hansen-Jagannathan distance (hereafter HJ-distance; see Hansen & Jagannathan, 1997), which results from the minimization of a quadratic function of pricing errors. We do not look at nonlinear sources of misspecification.[3] In practice, researchers generally analyze misspecification by introducing some differences between the true DGPs and some simulated misspecified DGPs (for example addition or omission of a variable, introduction of multicollinearity, etc...). This approach is clearly

[2]The SDF parameters are the parameters that allow to express the stochastic discount factor as a linear function of risk factors

[3]The reader can refer to Almeida & Garcia (2012) for a more general measure, effective on both linear and nonlinear sources of misspecification. The authors use the theory of minimum discrepancy estimators, and present asymptotic properties for the Cressie-Read family of discrepancies. Note that unlike here, they use an unconditional model.

limited, as it is very difficult to consider all differences. That is why a general SDF approach is better, since misspecifications are summarized by a distance that is a function of pricing errors.

In the proposed conditional model, the SDF parameters are time-varying and linked to a set of observable common instruments. While defining the useless factors the same way as Kan & Zhang (1999b) (factors that are uncorrelated with all the priced asset returns), we show that when inferences are performed without correction for misspecifications, the inclusion of a useless factor in the SDF model leads to unreliable inference results. However, our misspecification-robust inference is able to eliminate the negative effects of potential useless factors, while maintaining the relevance of the useful ones.

As stressed by Pondi (2017), a risk factor can be dynamically useless meaning useless in one period, and useful in another one. We pursue this idea by analyzing the time-varying importance of the factors in several linear asset pricing models from 1963 to 2016. The idea is to see how the significance of these factors evolves through time, assuming that the chosen SDFs are potentially (and probably) misspecified. Seven SDFs are chosen as benchmarks: (i) the CAPM; (ii) the three-factor model of Fama & French (1993), FF3; (iii) the four factor model of Carhart (1997), CARH; (iv) the five-factor model of Fama & French (2015), FF5; (v) a human capital model inspired by Lettau & Ludvigson (2001) and Gospodinov et al. (2014), C-LAB; (vi) the Consumption - CAPM; (vii) the Durable Consumption - CAPM of Yogo (2006). The data are monthly for the first four SDFs and quarterly for the last three. We measure the linear pricing ability of each of these SDFs for a set of 43 Fama-French portfolios from 1963 to 2016.[4]

Our results indicate that SDFs with just traded factors are obviously more relevant for the pricing of asset returns. However, as for the size factor *SMB* in FF3 and CARH, or for the profitability factor *RMW* in FF5, a traded factor can also be totally irrelevant (in the mentioned models, these two factors are not relevant at any date). Moreover, we show that additional factors do not systematically lead to better SDFs. A comparison of the SDFs is performed using as criterion the *relevance frequency of the SDF*, defined as the ratio between the number of periods where at least one factor other than the constant factor in the SDF is priced, and the total number of periods. It shows that FF3 and FF5 have been the overall

[4]Following Lewellen et al. (2010), the 43 portfolios include 17 Fama-French industry portfolios in addition to the 25 Fama-French size and book-to-market portfolios, and the one month T-bill rate.

best SDFs in the last 50 years. However, since 2000, the best SDF is CARH, followed by FF5 as the second best.

As the results confirm the advantage of using SDFs with traded factors, they also bring a nuance on non-traded factors. In fact, it appears that the C-LAB and the D-CCAPM have some relevance for the linear pricing of our portfolios. However, the CCAPM is not able to provide any linear risk-adjustment in connection with the selected portfolios. Since any SDF is theoretically linked to the consumption (by the Euler identity), this may call for further analysis of its relevance for pricing nonlinear risk-adjustments.

Besides the aforementioned articles, this paper is linked to papers that analyze the effects of useless factors on inference in asset pricing models and solve the ensuing problems. Gospodinov et al. (2017) address that issue by characterizing the asymptotic behavior of a GMM estimator under the failure of the rank conditions. Kleibergen & Zhan (2018) propose a test for the risk premia on mimicking portfolios of non-traded factors. Kleibergen & Zhan (2015) analyze the effects of a weak correlation between the observed proxy factors and the true unobserved ones, and build a sample distribution for the R^2 following the estimation of risk premia. Bryzgalova (2016) develops a penalized estimator that detects the useless factors in the model and restores the inference properties for the risk premia associated with the useful factors. Unlike these papers, our analysis relies on a conditional model and our inference process is designed for time-varying parameters.

On papers that build inference methods for time-varying parameters, Gagliardini et al. (2016) propose to infer the dynamics of the risk premia associated to risk factors when the number of assets is large. Kelly, Pruitt & Su (2017) use an Instrumented Principal Component Analysis to estimate the factors when the betas are time-varying. These two papers rely on regression models, where the returns are expressed as linear functions of the factors. This paper is different from that perspective, since our analysis is built on an SDF model and since we infer on SDF parameters instead of betas or risk premia. Unlike Burnside (2016) or Kan & Zhang (1999a), who use unconditional SDF models, the SDF model here is conditional.

After Nagel & Singleton (2011), inference on conditional SDF models has also been analyzed by Gagliardini & Ronchetti (2014), who introduce a conditional version of the HJ-distance as a kernel-based GMM estimator. We do not use that distance here since our framework includes potential useless factors, doing so can therefore create identification issues. Moreover, we do not follow Antoine, Proulx & Renault (2018), who use that conditional HJ-distance to estimate pseudo-true SDFs. For inference purposes, the pseudo-true SDFs in this paper are selected based on theoretical considerations.

Finally, this paper is linked to Feng et al. (2017), which evaluates the contribution of new factors to asset pricing models, in comparison to prior factors. That paper proposes a way to assess the real contribution of each new factor in the literature in comparison to those already found. The main difference here is that we analyze the dynamical relevance of the factors in several well-known and widely used asset pricing models.

The rest of the paper is organized as follows. Section 2 presents the formalization of the model. Section 3 reports results from the Monte Carlo simulations. Section 4 presents the empirical analysis. Section 5 concludes.

2.2 Time-varying SDF parameters: formalization

2.2.1 The model

The model below is a conditional version of the unconditional model proposed by Kan & Robotti (2009) and Gospodinov et al. (2014). Throughout the paper, the terms "conditional SDF parameters" and "time-varying SDF parameters" are used interchangeably.

Let us consider the following SDF:

$$y_t = \tilde{f}_t' \lambda_t, \tag{2.1}$$

where $\tilde{f}_t = (1, f_t')'$ a $(k+1)$-vector with f_t a k-vector of risk factors at time t, and $\lambda_t = (\lambda_t^0, \ldots, \lambda_t^k)'$ the vector of SDF parameters at time t. Following the idea of Gagliardini et al. (2016), we assume that $\tilde{f}_t$ is known at time t while λ_t is known at time $t-1$.[5] A factor is priced at a given time if its SDF parameter is statistically significant at that time. Also, the pricing ability of the SDF is measured at each date conditionally to all past available information.

Moreover, let us consider the gross returns on n portfolios at time t, R_t. If the SDF

[5] Formally, if $\{\mathscr{F}_t\}_{t=1,2,\ldots}$ is the filtration that characterizes the flow of information, then $\tilde{f}_t$ is $\mathscr{F}_t$-measurable and λ_t is $\mathscr{F}_{t-1}$-measurable.

correctly prices the n portfolios at time t, then by denoting 1_n the n-vector of 1, we have:

$$E_{t-1}(y_t R_t) = 1_n \ . \tag{2.2}$$

To define the dynamics, we assume that the SDF parameters are linked to p observed common instruments that are summarized in a p-dimensional vector Z. Then we consider the following assumptions.

Assumption 1. *There exists a $((k+1) \times p)$ - matrix Λ, such that the conditional SDF parameters at time t are given by the following relation;*

$$\lambda_t = \Lambda Z_{t-1}. \tag{2.3}$$

Assumption 2. *The matrix $U = E(R_t R_t^{'})$ is nonsingular.*

Assumption 1 is the assumption FS1 of Gagliardini et al. (2016). The variables in Z_t are financial or macroeconomic variables that are common to all assets. We assume that the SDF parameters are linear functions of lagged common instruments.[6] Assumption 2 is satisfied as soon as the selected assets are well diversified. This assumption is likely to be always verified for portfolios, as the classification criteria of individual assets into portfolios are different from one portfolio to another.

By assumption 1, the pricing error at time t depends on the estimation of Λ and is given by (with *vec* the vectorization operator):

$$e_t(\Lambda) \equiv E_{t-1}(y_t R_t) - 1_n = E_{t-1}\left(R_t(\tilde{f}_t^{'} \otimes Z_{t-1}^{'})\right) vec(\Lambda^{'}) - 1_n. \tag{2.4}$$

Because the matrix Λ is constant, we define a constant pricing error $e(\Lambda)$, as the expectation of the pricing errors through time. Denoting $D = E\left(R_t(\tilde{f}_t^{'} \otimes Z_{t-1}^{'})\right)$, this aggregate pricing error is given by:

$$e(\Lambda) \equiv E\left(e_t(\Lambda)\right) = E\left(R_t(\tilde{f}_t^{'} \otimes Z_{t-1}^{'})\right) vec(\Lambda^{'}) - 1_n = Dvec(\Lambda^{'}) - 1_n \ . \tag{2.5}$$

[6] Unlike Gagliardini et al. (2016), there is no specific instrument here. Specific instruments take into account that a particular instrument can impact some parameters but not all of them. When the number of assets is not large, we can just consider any specific instrument as a common instrument.

The HJ-distance (δ_{HJ}) is built from the pricing error and gives the degree of misspecification of the model. The model is correctly specified when $\delta_{HJ} = 0$ and misspecified when $\delta_{HJ} > 0$. Following Hansen & Jagannathan (1997) and by assumption 2, the HJ-distance of the model is given by $\delta_{HJ} = \left(e(\Lambda_*)^{'} U^{-1} e(\Lambda_*)\right)^{\frac{1}{2}}$; with the (pseudo-true) optimal matrix Λ_* chosen such that:

$$vec(\Lambda_*^{'}) = \underset{\Lambda}{argmin}\ e(\Lambda)^{'} U^{-1} e(\Lambda). \tag{2.6}$$

Let us consider now the $(p(k+1)+n)$ - vector $Y_t = (Z_{t-1}^{'} \vdots\ f_t^{'} \otimes Z_{t-1}^{'} \vdots\ R_t^{'})^{'}$ and its covariance matrix, $V = (V_{ij})_{1 \leq i,j \leq 3}$ (every V_{ij} is a block matrix). Following the argument of Kan & Robotti (2009), we can derive $vec(\hat{\Lambda}^{'})$ by replacing U^{-1} in (2.6) by V_{33}^{-1}. By denoting $\hat{V}$ and $\hat{D}$ respectively the empirical counterparts of V and D, the optimal matrix $\hat{\Lambda}$ and the HJ-distance from the model are estimated by:

$$vec(\hat{\Lambda}^{'}) = (\hat{D}^{'} \hat{V}_{33}^{-1} \hat{D})^{-1} (\hat{D}^{'} \hat{V}_{33}^{-1} 1_n), \tag{2.7}$$

$$\hat{\delta}_{HJ} = \left(1_n^{'} \hat{V}_{33}^{-1} 1_n - 1_n^{'} \hat{V}_{33}^{-1} \hat{D} vec(\hat{\Lambda}^{'})\right)^{1/2}. \tag{2.8}$$

Therefore, the sample estimate of the SDF vector at each date t is given by $\hat{\lambda}_t = \hat{\Lambda} Z_{t-1}$.

2.2.2 Choice of the instruments in Z

The choice of the instruments raises some issues. By equation (2.7), we see that $vec(\hat{\Lambda}^{'})$ is obtained by an OLS regression of $\hat{V}_{33}^{-1/2} 1_n$ on $\hat{V}_{33}^{-1/2} \hat{D}$. As $\hat{D} = \frac{1}{T-1} \sum_{t=2}^{T} R_t (\tilde{f}_t^{'} \otimes Z_{t-1}^{'})$. $vec(\hat{\Lambda}^{'})$ is not invariant to a change on Z and would vary according to the selected instruments. To solve the issue, an idea would be to choose the instruments which lead to the smallest HJ-distance. However, as shown by Kan & Robotti (2009), we cannot compare the relevance of two different models by comparing their HJ-distance (especially when the models are not nested).

As solution, we follow Gagliardini et al. (2016) and select all the instruments they use for their analysis and for their robustness checks: the term spread, the default spread and the dividend yield. The three are used at the same time here, in order to have the complete information we could get from potential instruments. The importance of the three selected variables with regard to forecasting stock returns (for the dividend yield) or capturing the variation in expected returns (for the term spread and the default spread) has been proven in the literature (see for example Fama & French, 1989; Jagannathan & Wang, 1996; Jensen, Mercer & Johnson, 1996; Petkova, 2006).

In this paper (and according to the literature), the term spread is measured as the difference of interest rates between 10-year Treasury bill and 3-month Treasury bill, and the default spread is measured as the difference of yields between Moody's Baa-rated and Aaa-rated corporate bonds. In the empirical analysis, the data on the three instruments are from the Federal Reserve Economic Data.

2.2.3 Asymptotic properties

The asymptotic properties of the estimated conditional SDF are presented, assuming that the model is potentially misspecified and without considering the particular case of having a useless factor in the model. A case with a useless factor, which can be considered as an extreme case of misspecification, will be considered in the next subsection.

In the following, remember that $Y_t = (Z'_{t-1} \vdots\, f'_t \otimes Z'_{t-1} \vdots\, R'_t)'$ and $D = E\left(R_t(\tilde{f}'_t \otimes Z'_{t-1})\right)$. We note $V = (V_{ij})_{1\leq i,j\leq 3} = Var(Y_t)$, and $\mu = (\mu'_1, \mu'_2, \mu'_3)' = E(Y_t)$. We also consider the vector ϕ and the random vector $r_t(\phi)$ defined as:

$$\phi = \begin{pmatrix} \mu \\ vech(V) \end{pmatrix} \quad \text{and} \quad r_t(\phi) = \begin{pmatrix} Y_t - \mu \\ vech\left((Y_t-\mu)(Y_t-\mu)' - V\right) \end{pmatrix},$$

with *vech* the half-vectorization operator.[7]

Proposition 1. *If the model is potentially misspecified,*

$$\sqrt{T}(vec(\hat{\Lambda}') - vec(\Lambda'_*)) \to \mathcal{N}\left(0_{p(k+1)}, \Sigma\right), \tag{2.9}$$

where Σ *is a* $p(k+1) \times p(k+1)$ *- matrix such as*

$$\Sigma = \left(\frac{\partial vec(\Lambda'_*)}{\partial \mu'} \vdots \frac{\partial vec(\Lambda'_*)}{\partial vec(V)'} \cdot D_{p(k+1)+n}\right) S_0 \left(\frac{\partial vec(\Lambda'_*)}{\partial \mu'} \vdots \frac{\partial vec(\Lambda'_*)}{\partial vec(V)'} \cdot D_{p(k+1)+n}\right)'; \tag{2.10}$$

[7] The half-vectorization operator is defined for symmetric matrices, and is equivalent to a vectorization applied on the lower triangular part of the initial matrix

with $D_{p(k+1)+n}$ *a duplication matrix, and* $S_0 = \sum_{j=-\infty}^{\infty} E\left(r_t(\phi) r_{t+j}(\phi)'\right)$.

The expressions of the derivatives $\frac{\partial vec(\Lambda')}{\partial \mu'} \equiv (\frac{\partial vec(\Lambda')}{\partial \mu_1'} \vdots \frac{\partial vec(\Lambda')}{\partial \mu_2'} \vdots \frac{\partial vec(\Lambda')}{\partial \mu_3'})$ and $\frac{\partial vec(\Lambda')}{\partial vec(V)'}$ in the previous proposition are given by the following lemma.

Lemma 1. *If we denote* $e \equiv e(\Lambda)$ *and* $H \equiv D' V_{33}^{-1} D$*; the derivatives in Proposition 1 are given by the following expressions:*

$$\frac{\partial vec(\Lambda')}{\partial \mu_1'} = -H \begin{pmatrix} I_p \\ 0_{kp\times p} \end{pmatrix} \otimes e' V_{33}^{-1} \mu_3 - \left(vec(\Lambda')\right)' \begin{pmatrix} I_p \\ 0_{kp\times p} \end{pmatrix} \otimes H D' V_{33}^{-1} \mu_3 \tag{2.11}$$

$$\frac{\partial vec(\Lambda')}{\partial \mu_2'} = -H \begin{pmatrix} 0_{p\times kp} \\ I_{kp} \end{pmatrix} \otimes e' V_{33}^{-1} \mu_3 - \left(vec(\Lambda')\right)' \begin{pmatrix} 0_{p\times kp} \\ I_{kp} \end{pmatrix} \otimes H D' V_{33}^{-1} \mu_3 \tag{2.12}$$

$$\frac{\partial vec(\Lambda')}{\partial \mu_3'} = -H \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \otimes e' V_{33}^{-1} - \left(vec(\Lambda')\right)' \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \otimes H D' V_{33}^{-1} \tag{2.13}$$

$$\begin{aligned} \frac{\partial vec(\Lambda')}{\partial vec(V)'} = & \left[H \begin{pmatrix} I_p & 0_{p\times kp} \\ 0_{kp\times p} & I_{kp} \end{pmatrix} \vdots 0_{p(k+1)\times n} \right] \otimes \left(0_{1\times p(k+1)} \vdots -e' V_{33}^{-1} \right) \\ & - \left[\left(vec(\Lambda')\right)' \begin{pmatrix} I_p & 0_{p\times kp} \\ 0_{kp\times p} & I_{kp} \end{pmatrix} \vdots e' V_{33}^{-1} \right] \otimes \left(0_{p(k+1)\times p(k+1)} \vdots H D' V_{33}^{-1} \right) \end{aligned} \tag{2.14}$$

Proposition 1 (with lemma 1) gives the asymptotic variance associated with the estimation of the matrix Λ. This variance is different, depending on whether $e = 0$ or not (meaning on whether the model is correctly specified or not). It is therefore important to consider any potential missecification while testing for the relevance of the factors in the model. Note that we do not know here, whether the asymptotic variance increases or decreases following a misspecification.[8] However, having an asymptotic variance which depends on the pricing errors is a sufficient motivation for a misspecification-robust inference.

Under a potential model misspecification, the asymptotic distribution of the time-varying SDF parameters is given by the following proposition (which is a corollary of Proposition 1).

[8] Kan & Robotti (2009) show for example that when factors and returns are multivariate elliptically distributed, misspecifications increase the asymptotic variance of the parameters

Proposition 2. *If the SDF is potentially misspecified, then at each date t,*

$$\sqrt{T}(\hat{\lambda}_t - \lambda_t^*) \to \mathcal{N}\left(0_{k+1}, (I_{k+1} \otimes Z_{t-1}')\Sigma(I_{k+1} \otimes Z_{t-1})\right), \tag{2.15}$$

with Σ defined as in Proposition 1.

2.2.4 Asymptotic properties in the presence of a useless factor

Let us assume that the SDF includes a useless factor along with some useful ones. Under that assumption, we present here the asymptotic behaviors of the SDF parameters associated with the useful factors. The idea is to: (i) present the inference issues arising from the presence of the useless factor and (ii) present how a misspecification-robust inference procedure can eliminate these issues.

The estimation of the conditional SDF parameters in our model relies on the estimation of the matrix Λ, as the instruments are observable. So, the misspecification issues should first be assessed with regard to the estimation of that matrix. As Λ is constant, the covariance matrix of $vec(\hat{\Lambda}')$ is defined with unconditional expectations. Therefore, the useless factors need to be analyzed here according to the classical definition of Kan & Zhang (1999b).

Definition 1. Let us consider a factor $g = \{g_t\}$, where for all t, g_t belongs to $\mathbb{R}$. A **useless factor** is a factor which is uncorrelated with asset returns. So if g is useless, then $E(R_t g_t) = 0_{n\times 1}$.

The useless factors are defined here the same way as in Gospodinov et al. (2014). Now suppose that a useless factor g is added to k initial factors, which are assumed to be all useful. Therefore, the SDF takes the following form:

$$y_t = \tilde{f}_t' \lambda_t^{us} + g_t \lambda_t^{\bar{us}} \ . \tag{2.16}$$

Following the theoretical framework, we assume that at each date t, $\lambda_t^{us} = \Lambda_{us} Z_{t-1}$ and $\lambda_t^{\bar{us}} = \Lambda_{\bar{us}} Z_{t-1}$ (with Λ_{us} a $(k+1)\times p$ - matrix, and $\Lambda_{\bar{us}}$ a $1 \times p$ - matrix). By denoting $\Lambda = (\Lambda_{us}' \vdots \Lambda_{\bar{us}}')'$, the aggregate pricing error is therefore given by

$$e(\Lambda) = E(e_t(\Lambda)) = E\left[E_{t-1}\left(R_t(\tilde{f}_t' \otimes Z_{t-1}')\right) vec(\Lambda_{us}') + E_{t-1}\left(R_t g_t\right) Z_{t-1}' vec(\Lambda_{\bar{us}}') - 1_n\right]$$

$$
\begin{aligned}
&= E\left(R_t(\tilde{f}_t' \otimes Z_{t-1}')\right) vec(\Lambda_{us}') - 1_n \\
&= e(\Lambda_{us}) \ . \qquad (2.17)
\end{aligned}
$$

The matrix Λ is obtained through a minimization of a quadratic function of the aggregate pricing error $e(\Lambda)$. As the matrix $\Lambda_{\bar{us}}$ associated with the useless factor is not identified, we follow Gospodinov et al. (2014) and consider the following assumptions.

Assumption 3. *For a useless factor g, the pseudo-true value of the matrix $\Lambda_{\bar{us}}$ is set such that $\Lambda_{\bar{us}}^* = 0_{1\times p}$.*

Assumption 4. *(i) The number of assets n is such that $n > p(k+2)$; (ii) The processes (R_t', f_t', g_t) are jointly stationary and ergodic, with finite fourth moments; (iii) $e_t(\Lambda_{us}) - E(e_t(\Lambda_{us}))$ defines a martingale difference sequence; (iv) If we denote $D_{us} \equiv E(R_t(\tilde{f}_t' \otimes Z_{t-1}'))$ and $D_{\bar{us}} \equiv E(R_t g_t Z_{t-1}')$, the column rank of both the matrices D_{us} and $(D_{us} \vdots D_{\bar{us}})$ is $p(k+1)$.*

Assumption 3 is consistent with Gospodinov et al. (2014) as they show that the SDF parameter of a useless factor is symmetrically distributed around 0. As the true value of $\Lambda_{\bar{us}}^*$ is actually not identifiable, this seems to be a "natural choice" according to their words. There is another motivation for this assumption; since useless factors are not correlated to priced returns, they should not be rewarded (by being useless, they do not characterize any risk). So the risk premium linked to a useless factor should be null. One way to achieve that condition is to set at 0, the SDF parameters associated with useless factors.

Assumption 4 is the assumption 1 from Gospodinov et al. (2014). Note that D_{us} is a $n \times p(k+1)$ matrix, and $(D_{us} \vdots D_{\bar{us}})$ is a $n \times p(k+2)$ matrix.

Proposition 3. *Under assumptions 3 and 4, we have as $T \to \infty$:*
(i) if the SDF is correctly specified, $vec(\hat{\Lambda}_{us}' - \Lambda_{us}') = O_p(T^{-1/2})$; but its asymptotic distribution is not normal;
(ii) if the SDF is misspecified, $vec(\hat{\Lambda}_{us}' - \Lambda_{us}') = O_p(1)$; so $vec(\hat{\Lambda}_{us}')$ is not consistent.

The previous proposition shows that when the SDF includes a useless factor, a conventional estimation procedure would lead to inference issues on the matrix associated with the useful factors (Λ_{us}'). These issues would be even worse when the model is misspecified. Proposition 3 is a multivariate version of Proposition 1 of Gospodinov et al. (2014).[9]

[9] We present in appendices, some asymptotic behaviors of the vector $vec(\hat{\Lambda}_{\bar{us}}')$.

The presence of a useless factor in an asset pricing model is generally presented as an issue only for the estimation of the risk premium, this one being following Fama & MacBeth (1973), a second-step estimator which depends on the estimation of the betas.[10] Proposition 3 shows that under the general SDF approach, useless factors generate inference issues even for first-step estimators like SDF parameters.

Note that although the presence of a useless factor does not necessarily mean the SDF is misspecified, a useless factor leads here to a failure of conditions for consistency and asymptotic normality. This may increase the pricing errors, and therefore increase the HJ-distance. So, having a useless factor in the SDF model have similar consequences on the HJ-distance than any kind of misspecification. Therefore, by solving misspecification issues, our estimator should also be able to correct issues from having useless factors in the SDF.

In the next subsection, we show that in the presence of a useless factor, our misspecification-robust conditional model can actually provide standard statistics for inference on Λ.

Wald tests in the presence of a useless factor

Gospodinov et al. (2014) show that the t-statistics built with misspecification-robust standard errors are normal, even when the asymptotic behaviors of the estimators are not standard (following the inclusion of a useless factor in the SDF). Likewise, we analyze the asymptotic behavior of the Wald tests on Λ, with misspecification-robust variance matrices.

Starting from the model (2.16), we consider the empirical counterpart of the asymptotic variance Σ, as presented in Proposition 1. We can write $\hat{\Sigma}$ as follows (with $\hat{\Sigma}_{11}$ a $p(k+1) \times p(k+1)$ - matrix, and $\hat{\Sigma}_{22}$ a $p \times p$ - matrix):

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix} , \tag{2.18}$$

with $\hat{\Sigma}_{11}$ and $\hat{\Sigma}_{22}$ respectively the covariance matrices associated with $\sqrt{T}vec(\hat{\Lambda}'_{us})$ and $\sqrt{T}vec(\hat{\Lambda}'_{\bar{us}})$.

[10] The beta-estimator of useless factors have non-standard limit behaviors; this compromises inferences on risk premia

Let us consider the following Wald statistics:

$$W(vec(\hat{\Lambda}'_{us})) = T\left(vec(\hat{\Lambda}'_{us}) - vec(\Lambda'_{us})\right)'(\hat{\Sigma}_{11})^{-1}\left(vec(\hat{\Lambda}'_{us}) - vec(\Lambda'_{us})\right) \tag{2.19}$$

$$W(vec(\hat{\Lambda}'_{\bar{us}})) = T\left(vec(\hat{\Lambda}'_{\bar{us}})\right)'(\hat{\Sigma}_{22})^{-1}\left(vec(\hat{\Lambda}'_{\bar{us}})\right) \tag{2.20}$$

$W(vec(\hat{\Lambda}'_{us}))$ and $W(vec(\hat{\Lambda}'_{\bar{us}}))$ are respectively the statistics associated with the Wald tests on the useful factors and on the useless factor.

Let us also consider $D_{us} \equiv E(R_t(\tilde{f}'_t \otimes Z'_{t-1}))$ and $D_{\bar{us}} \equiv E(R_t g_t Z'_{t-1})$ as defined in Assumption 4; $\hat{D}_{us}$ and $\hat{D}_{\bar{us}}$ their empirical counterparts (which are assumed to be consistent). As saw previously, $vec(\hat{\Lambda}') = \left(vec(\hat{\Lambda}_{us})' \; vec(\hat{\Lambda}_{\bar{us}})'\right)'$ can be obtained through an OLS regression on the following model:

$$\hat{U}^{-\frac{1}{2}}1_n = \hat{U}^{-\frac{1}{2}}\hat{D}_{us}vec(\Lambda'_{us}) + \hat{U}^{-\frac{1}{2}}\hat{D}_{\bar{us}}vec(\Lambda'_{\bar{us}}) + \varepsilon \; . \tag{2.21}$$

Assumption 5. *If we denote* $\tilde{D}_{\bar{us}} \equiv \hat{U}^{-\frac{1}{2}}\hat{D}_{\bar{us}}$ *and* $\hat{M}_{\bar{us}} \equiv I_n - \tilde{D}_{\bar{us}}(\tilde{D}'_{\bar{us}}\tilde{D}_{\bar{us}})^{-1}\tilde{D}'_{\bar{us}}$, *then*

$$\sqrt{T}\tilde{D}'_{us}\hat{M}_{\bar{us}}\varepsilon \xrightarrow{p} r_\varepsilon \sim \mathcal{N}(0_{p(k+1)}, V_{r_\varepsilon}) \; . \tag{2.22}$$

As a motivation, note that from (2.21), we have $\hat{M}_{\bar{us}}\hat{U}^{-\frac{1}{2}}1_n = \hat{M}_{\bar{us}}\tilde{D}_{us}vec(\Lambda'_{us}) + \hat{M}_{\bar{us}}\varepsilon$. Assumption 5 is a regularity condition on the errors of this model.[11]

The following proposition gives the asymptotic behaviors of the two Wald statistics $W(vec(\hat{\Lambda}'_{us}))$ and $W(vec(\hat{\Lambda}'_{\bar{us}}))$.

Proposition 4. *If the model is misspecified, then under assumptions 3, 4 and 5, we have as* $T \to \infty$:

$$W(vec(\hat{\Lambda}'_{us})) \xrightarrow{d} \chi^2(p(k+1)) \;\; \text{and} \;\; W(vec(\hat{\Lambda}'_{\bar{us}})) \xrightarrow{d} \chi^2(p) \; , \tag{2.23}$$

with $W(vec(\hat{\Lambda}'_{us}))$ *and* $W(vec(\hat{\Lambda}'_{\bar{us}}))$ *respectively defined as in (2.19) and (2.20).*

Proposition 4 follows the conclusions of Gospodinov et al. (2014). It shows that although the inclusion of a useless factor leads to asymptotically non-standard estimators, we can still perform reliable inferences by using misspecification-robust covariance matrices.

[11] Since the model includes a useless factor we know, from the results of Gospodinov et al. (2014), that $r'_\varepsilon V_{r_\varepsilon}^{-1} r_\varepsilon$ is dominated by a chi-squared distribution. Assumption 5 is a sufficient condition for that result.

2.3 Monte Carlo simulations

Two exercises are performed in this section. The first is to measure the rejection probabilities of the tests on $vec(\Lambda^{'})$[12]. For that exercise, we follow the idea of Gospodinov et al. (2014). However, we consider that all the models are potentially misspecified, so the pricing errors in the simulations are set to their empirical values.[13] The second exercise is to compare the mean of the conditional SDF parameters to the unconditional SDF parameter from Kan & Robotti (2009).

Misspecification and pseudo-true values of estimates

Since the model is misspecified, we do not have the real value of $vec(\Lambda^{'})$.[14] For our subsequent analysis, we will choose its empirical counterparts as pseudo-true value.

The mean vector and the covariance matrix associated with the vector $Y_t = (Z^{'}_{t-1} \vdots\ f^{'}_t \otimes Z^{'}_{t-1} \vdots\ R^{'}_t)^{'}$ are respectively given by $\mu = (\mu^{'}_1 \vdots\ \mu^{'}_2 \vdots\ \mu^{'}_3)^{'}$ and $V = (V_{ij})_{1\leq i,j\leq 3}$ (recall that V_{ij} is the covariance matrix between elements i and j from the vector Y_t). Therefore, $D = E\left(R_t(\tilde{f}^{'}_t \otimes Z^{'}_{t-1})\right) = \left(V_{31} + \mu_3\mu^{'}_1\ ,\ V_{32} + \mu_3\mu^{'}_2\right)$. By denoting $X = \left(\hat{V}_{31} + \hat{\mu}_3\hat{\mu}^{'}_1\ ,\ \hat{V}_{32} + \hat{\mu}_3\hat{\mu}^{'}_2\right)$, the pseudo-true values of $vec(\Lambda^{'})$ is defined by:

$$vec(\Lambda^{'}_*) = (X^{'}\hat{V}^{-1}_{33}X)^{-1}(X^{'}\hat{V}^{-1}_{33}1_n), \tag{2.24}$$

and the pseudo-true value of the SDF vector at each date t is defined by $\lambda^*_t = \Lambda_* Z_{t-1}$.

2.3.1 Rejection probabilities of the Wald tests on vec($\Lambda^{'}$)

Under assumption 3, we show that there is no size distortion when testing if the row vectors of Λ are respectively equal to their real counterparts (even when the model includes a useless factor). Testing for each row of Λ is equivalent to testing for each p consecutive elements of $vec(\Lambda^{'})$. The reason for doing so is that, as $\lambda_t = \Lambda Z_{t-1}$, the i^{th} parameter of λ_t is

[12] Any potential size distortion of the t-tests on the conditional parameter λ_t would be a consequence of a size distortion of the t-tests on $\vec{(\Lambda^{'})}$; as at each date t, $\hat{\lambda}_t = \hat{\Lambda}Z_{t-1}$

[13] Gospodinov et al. (2014) set the pricing error to zero to simulate a correctly specified model (and accordingly adjust the mean of the returns R during the simulations). They show that a misspecification-robust test is efficient on both a correctly specified and a misspecified model, while a classic test is efficient only when the model is correctly specified. When the model is correctly specified, the misspecification-robust statistic and the classical one are identical

[14] Because of the misspecification, some parameters are not fully identified.

obtained by a scalar product of the i^{th} row of Λ with the instruments. Therefore, by testing for the row vectors of Λ, we are able to obtain a first information about the unconditional relevance of each factor. This unconditional relevance will be later compared with the results of Gospodinov et al. (2014) to see if our conditional model is coherent with the literature. Note that testing for the time-varying relevance of the factors is the second step of our analysis. As $\lambda_t = \Lambda Z_{t-1}$, we can deduce the time-varying distribution of λ_t from the distribution of $vec(\Lambda^{'})$.

For the simulations, we consider three different specifications as linear factor models: (i) a model with a constant and a useful factor (model 1), (ii) a model with a constant and a useless factor (model 2), (iii) a model with a constant, a useful and a useless factor (model 3). As data, we use the monthly gross returns on: the one-month T-bill rate, the 25 Fama-French portfolios sorted on size and book-to-market, and the 17 Fama-French industry portfolios. The data are from Jul 1963 to Dec 2016, so the number of periods is $T = 642$ and the number of assets is $n = 43$. We also consider the market return as the factor from which the characteristics of the normal random variable used as useful factor during the simulations are calibrated. The market factor and all the returns are from Kenneth French's website. As usual, the useless factor is generated as a random variable uncorrelated with the returns.[15] The simulation design is the same as in Gospodinov et al. (2014), with the exception that the empirical value of Λ is considered as the pseudo-true value of that matrix (denoted Λ_* or Λ^*) during the simulations. We perform 10 000 Monte Carlo replications. For each replication, we simulate a normal random variable whose characteristics are calibrated on the mean and the variance of $Y_t = (Z^{'}_{t-1} \vdots\, f^{'}_t \otimes Z^{'}_{t-1} \vdots\, R^{'}_t)^{'}$, and we estimate the matrix $\hat{\Lambda}$. As Λ is a $(k+1) \times p$ -matrix, we perform $k+1$ Wald tests to assess the equality between the row vectors of $\hat{\Lambda}$ and those of Λ_*. Four sample sizes are considered: $T = 200$, $T = 600$, $T = 1000$, and $T = 2500$.

Table 2.1 shows the rejection probabilities from the Wald tests on $H_0 : \Lambda_i = \Lambda^*_i$ (with Λ_i the row vector of Λ associated with the i^{th} factor). We see that when the misspecified SDF does not include a useless factor (model 1), there is no size distortion, as the empirical sizes from the tests are very close to the nominal sizes. This result holds even for relatively small sample sizes. When the misspecified SDF includes a useless factor (models 2 and 3), there is almost no size distortion for the tests on vectors associated with the useless factor;

[15] We will use a normal distribution with mean 0 and variance 1 in our simulations, to generate a useless factor

on the other hand, there is an under-rejection for the tests on vectors associated with the useful factors. This under-rejection is not necessarily bad, as it just shows that the model is conservative on the useful factors when there is a potential useless factor in the SDF. The most important observation here is that the model is able to recognize the useless factors,[16] and that the estimated row vectors of Λ are equal to the pseudo-true values, even when the misspecified SDF contains useless factors. However, we need to verify if the estimated row vectors associated with the useful factors are different from the null vector, as this would mean they are effectively significant.

The previous analysis shows that the conditional model is able to detect the useless factors and shrink the value of the row vectors associated with them to $0_{1\times p}$. Next, we need to measure the ability of the conditional model to properly detect useful factors. Indeed, the real values of the row vectors of Λ associated with the useful factors are different from $0_{1\times p}$. So, we need to assess how often in the simulations these row vectors are different from the null vector.

Table 2.2 presents the rejection probabilities from the Wald tests on $H_0 : \Lambda_i = 0_p$. The objective here is to see how the power of this test evolves, relatively to the useful factors. We see that when the misspecified SDF does not contain a useless factor, the power of the test is 1. Therefore, without any useless factor, the useful factors are always correctly identified by the conditional model. On the other hand, when there is a useless factor in the SDF, the power of the test decreases with the nominal size. For the useful factors, the test rejects the null $H_0 : \Lambda_i = 0_p$ with a probability greater than 0.8 only when the nominal size is 0.1. So, in order for the model to detect the useful factors with a success rate of at least 80%, the type-1 error could be increased to 0.1 (instead of the usual 0.05 in empirical analysis).[17]

Note that as the real value of the row vector associated with a useless factor is $0_{1\times p}$, the results for the tests on the useless factors in Table 2.2 are the same as in Table 2.1 (the two tests are identical for useless factors).

[16] Remember that according to assumption 3, the real value of the row vector associated with a useless factor is set to $0_{1\times p}$. So, the proximity between the rejection probabilities and the nominal sizes actually means that the model performs well at detecting the useless factors.

[17] Here, increasing the ability of the model to detect properly the useful factors is equivalent to decreasing the type-2 error. One way to achieve this goal is to increase the type-1 error; this is what we will do here.

Table 2.1: Rejection probabilities from the Wald tests on $H_0 : \Lambda_i = \Lambda_i^*$

		$size = 10\%$				$size = 5\%$				$size = 1\%$			
		200	600	1000	2500	200	600	1000	2500	200	600	1000	2500
Model 1	constant (Λ_0)	0.098	0.098	0.098	0.099	0048	0.049	0.047	0.049	0.012	0.011	0.009	0.010
	useful (Λ_1)	0.109	0.102	0.105	0.104	0.057	0.053	0.053	0.050	0.013	0.013	0.012	0.009
Model 2	constant (Λ_0)	0.001	0	0	0	0	0	0	0	0	0	0	0
	useless (Λ_2)	0.155	0.110	0.105	0.100	0.083	0.055	0.051	0.047	0.020	0.009	0.009	0.009
Model 3	constant (Λ_0)	0.007	0.001	0	0	0.002	0	0	0	0	0	0	0
	useful (Λ_1)	0.010	0.001	0	0	0.003	0	0	0	0	0	0	0
	useless (Λ_2)	0.153	0.109	0.104	0.094	0.080	0.055	0.051	0.046	0.017	0.010	0.009	0.009

We perform 10 000 Monte-Carlo replications. At each replication, we generate returns, factors and instruments from a multivariate normal distribution with characteristics calibrated to the empirical mean and the empirical variance of $Y_t = (Z_{t-1}^{'} \ , \ f_t^{'} \otimes Z_{t-1}^{'} \ , \ R_t^{'})'$. The mean of a useless factor is set to 1, and the covariance between a useless factor and the returns (or the instruments) is set to zero. We then estimate Λ and test for each row i of that matrix, the null $H_0 : \Lambda_i = \Lambda_i^*$. Λ_0 is the row vector associated with the constant factor parameter, while Λ_1 and Λ_2 are associated respectively to the useful factor parameter and the useless factor parameter. The Wald statistics are compared to the critical values from the χ^2 distribution with p levels of freedom. The table gives the rejection probabilities across the replications

Table 2.2: Rejection probabilities from the Wald tests on $H_0 : \Lambda_i = 0_{1\times p}$

		size = 10%				*size* = 5%				*size* = 1%			
		200	600	1000	2500	200	600	1000	2500	200	600	1000	2500
Model 1	constant (Λ_0)	1	1	1	1	1	1	1	1	1	1	1	1
	useful (Λ_1)	1	1	1	1	1	1	1	1	1	1	1	1
Model 2	constant (Λ_0)	0.691	0.771	0.787	0.800	0.564	0.666	0.691	0.707	0.329	0.450	0.480	0.494
	useless (Λ_2)	0.155	0.110	0.105	0.100	0.083	0.055	0.051	0.047	0.020	0.009	0.009	0.009
Model 3	constant (Λ_0)	0.820	0.882	0.893	0.905	0.724	0.815	0.836	0.847	0.496	0.640	0.672	0.700
	useful (Λ_1)	0.687	0.812	0.836	0.855	0.560	0.717	0.751	0.782	0.303	0.498	0.554	0.591
	useless (Λ_2)	0.153	0.109	0.104	0.094	0.080	0.055	0.051	0.046	0.017	0.010	0.009	0.009

We perform 10 000 Monte-Carlo replications. At each replication, we generate returns, factors and instruments from a multivariate normal distribution with characteristics calibrated to the empirical mean and the empirical variance of $Y_t = (Z'_{t-1}\ ,\ f'_t \otimes Z'_{t-1}\ ,\ R'_t)'$. The mean of a useless factor is set to 1, and the covariance between a useless factor and the returns (or the instruments) is set to zero. We then estimate Λ and test for each row i of that matrix, the null $H_0 : \Lambda_i = 0_{1\times p}$. Λ_0 is the row vector associated with the constant factor parameter, while Λ_1 and Λ_2 are associated respectively to the useful factor parameter and the useless factor parameter. The Wald statistics are compared to the critical values from the χ^2 distribution with p levels of freedom. The table gives the rejection probabilities across the replications

2.3.2 Importance of using misspecification-robust tests

As shown by Gospodinov et al. (2014), when the SDF is misspecified and the tests on the SDF parameters are performed without taking into account pricing errors, researchers could mistakenly conclude that a useless factor is priced. So, by using SDF models, they confirm the results obtained with linear factor models, about inference in the presence of useless factors.[18]

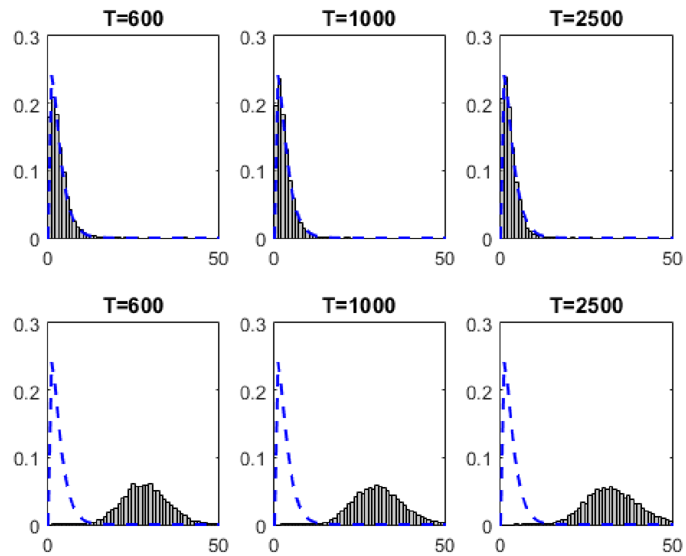
In this paper, we do not present the comparison between the results from a conventional test (built under the assumption that the model is correctly specified and the HJ-distance is 0), and the results from our misspecification-robust test. These results are similar to those of Gospodinov et al. (2014) and are available upon request.[19] Instead, we present here an illustration to show how the correction proposed improves inferences. For this purpose, we consider again model 3 as presented in the previous subsection (a constant, a useful and a useless factor). Following the simulations made in that model and presented in Table 2.1, we compare the distributions of each row vector of Λ, with the $\chi^2(3)$ distribution (as we have $p = 3$ instruments in Z, the row vectors of Λ are of dimension 3). We also perform the same simulation process with conventional standard errors, assuming that the model is correctly specified and that the pricing error is 0.

Figures 2.1 and 2.2 present the distributions of the row vectors associated with the useful factor and the useless factor, when using respectively no correction and the misspecification correction. We see that the misspecification-robust Wald test is able to correct the distortions, particularly for the useless factor. With this correction, a useless factor is much less likely to be declared priced. However, we can observe that the misspecification-robust Wald test create a little divergence in the result of the row vector associated with the useful factor. But this little divergence does not generate a big issue, as we have seen in the previous subsection (we can still have a reasonable statistical power for the test $H_0 : \Lambda_1 = 0_{1\times p}$, where Λ_1 is the row vector associated with the useful factor).

[18] See Kan & Zhang (1999b); Kleibergen (2009); Bryzgalova (2016); Pondi (2017).

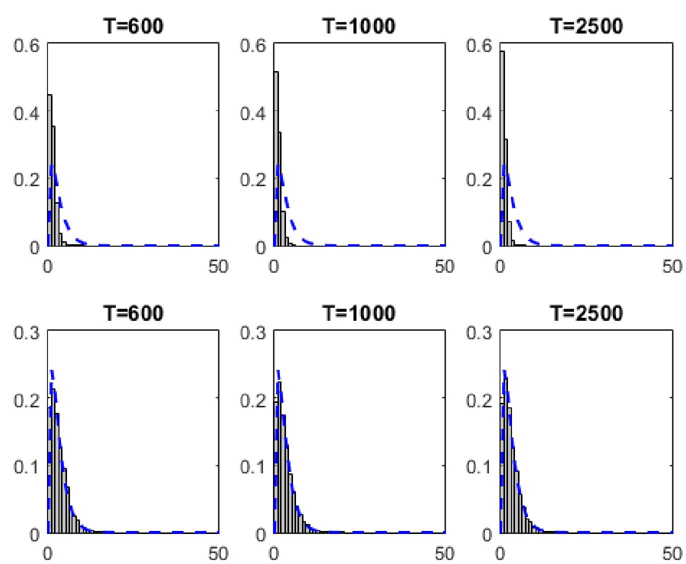
[19] When the model is correctly specified, the misspecification-robust test and the conventional test have similar results. But when the model is misspecified, the conventional test performs poorly.

Figure 2.1: Wald statistics (with no correction) and $\chi^2(3)$ distribution



We analyze the Wald statistics from the test $H_0 : \Lambda^i = \Lambda^i_*$, where Λ^i is the i^{th} row vector of Λ. The first line and the second line figures respectively present the results for the useful factor, and the results for the useless factor (for different values of T during the simulations). $\Lambda^2_* = 0_{1\times 3}$ is taken as the real value of the row vector associated with the useless factor.
The dashed line is the $\chi^2(3)$ distribution; we consider that distribution because we have $p = 3$ instruments in Z, so the row vectors are of dimension 3.

Figure 2.2: Misspecification-robust Wald statistics and $\chi^2(3)$ distribution



2.4 Empirical analysis

In this section, we perform two exercises. In the first, we analyze the coherence of our model with the literature. In the second, we analyze the dynamical relevance of the factors in some popular models, which are used in this paper as benchmarks.

2.4.1 Benchmarks and data

Seven asset pricing models are used as benchmarks:[20] (i) the **CAPM** (the market excess return *Mkt-Rf* as unique factor); (ii) the three-factor model of Fama & French (1993), **FF3** (CAPM factor plus the size factor *SMB* and the value factor *HML*); (iii) the four factor model of Carhart (1997), **CARH** (FF3 factors plus the momentum factor *Mom*); (iv) the five-factor model of Fama & French (2015), **FF5** (FF3 factors plus the profitability factor *RMW* and the investment factor *CMA*); (v) a human capital model inspired by Lettau & Ludvigson (2001) and Gospodinov et al. (2014), **C-LAB** (CAPM factor plus the consumption-wealth ratio *Cay* and the growth rate in per capita labor income *Lab*); (vi) the consumption CAPM, **CCAPM** (the growth rate in real per capita non-durable consumption *Cnd* as unique factor); (vii) the durable CCAPM of Yogo (2006), **D-CCAPM** (CAPM and CCAPM factors plus the growth rate in real per capita durable consumption *Cd*).[21]

Furthermore, we consider as returns the gross return on the one-month T-bill rate, the gross returns of 25 Fama-French portfolios sorted on size and book-to-market, and the gross returns of 17 Fama-French industry portfolios; so $n = 43$ (all the returns are from Kenneth French's website). As indicated in subsection 2.2.2, three financial variables are considered as instruments: the default spread, the term spread and the dividend yields. For CAPM, FF3, CARH, and FF5, we have monthly data from Jul 1963 to Dec 2016, whereas for C-LAB, CCAPM, and D-CCAPM, we have quarterly data from Q3 1963 to Q4 2016. As lagged variables, the instruments go from Jun 1963 to Nov 2016 in monthly models, and from Q2 1963 to Q3 2016 in quarterly models. The monthly sample has $T = 642$ months, and the quarterly sample has $T = 214$ quarters.

[20] For a complete description of each SDF, see Gospodinov et al. (2014).

[21] The Fama-French factors and the momentum are from K. French's website. The consumption wealth ratio is from Martin Lettau's website. The labor income, the durable and the non-durable consumption are from the Bureau of Economic Analysis.

2.4.2 Coherence of the model with the literature

As emphasized above, the SDF parameter estimator presented in this paper is a conditional version of the unconditional SDF parameter estimator proposed by Kan & Robotti (2009) and Gospodinov et al. (2014) (hereafter GKR). Similar to this paper, GKR analyze the relevance of the factors in some popular linear asset pricing models. As our model estimates firstly an unconditional row vector for each factor in the SDF, it would be interesting to compare the coherence between the significance of these row vectors and the significance of their associated factors, following the selection process of GKR. Precisely, for a given SDF, we want to see if the row vector of Λ associated with a factor is significant only when this factor is significant according to GKR. For that comparison, we simply consider the SDFs used by GKR that are: the CAPM, the FF3, the C-LAB, the CCAPM, and the D-CCAPM.

Table 2.3 presents, for the selected SDFs, the significance of the row vectors of Λ for each factor, as well as the significance of their associated factors in GKR. We observe that, for each factor in the selected SDFs, there is a coherence between the significance of the associated row vector according to our model and its significance according to GKR. In all the SDFs, the row vector of Λ associated with a given factor is: (i) significant for the factors with a significant SDF parameter in GKR; (ii) insignificant for those with an insignificant SDF parameter in GKR.

Note that if the row vector of Λ associated with a factor is significant, this does not mean that the conditional SDF parameters associated with that factor will all be significant through time. In fact, the significance of a row vector as analyzed above does not tell us anything about each component of that vector; for example, some may be significant if we perform single t-tests on them, and some others may not (recall that a vector can be significant with only one of its components being significant). Since $\lambda_t = \Lambda Z_{t-1}$, the conditional SDF parameter (at time t) associated with a factor is therefore, as a scalar product, the sum of terms whose significance is not known. Then it can be significant or not, no matter the results we have on the significance of the row vectors of Λ.

Table 2.3: Significance of the row vectors of $\hat{\Lambda}$ and coherence with GKR

		Monthly models			
		significance GKR	Wald stat ($H_0 : \Lambda_i = 0_{1\times p}$)	significance $\hat{\Lambda}_i$	coherence with GKR
CAPM	Mkt-Rf	yes	12.52 (***)	yes	yes
FF3	Mkt-Rf	yes	20.32 (***)	yes	yes
	SMB	no	0.99	no	yes
	HML	yes	24.00 (***)	yes	yes
		Quarterly models			
		significance GKR	Wald stat ($H_0 : \Lambda_i = 0_{1\times p}$)	significance $\hat{\Lambda}_i$	coherence with GKR
CCAPM	Cnd	no	1.65	no	yes
D-CCAPM	Mkt-Rf	no	2.02	no	yes
	Cnd	no	2.20	no	yes
	Cd	no	3.93	no	yes
C-LAB	Mkt-Rf	no	2.52	no	yes
	Cay	no	4.20	no	yes
	Lab	no	2.50	no	yes

For each model, we estimate the matrix $\hat{\Lambda}$ and we test whether the row vectors of this matrix are significant or not. Then we compare if the result on the significance of the row vector associated with a given factor, is coherent with the selection process of GKR. The Wald tests here are performed at 10%, 5% and 1% level of significance (the results are summarized respectively by $*$, $**$ and $***$). The critical values of the Wald tests are estimated from the distribution $\chi^2(p)$ (with $p = 3$ the number of instruments), and are respectively equal to 6.25, 7.81 and 11.34.

2.4.3 Comparisons with the unconditional estimator of GKR

We compare the SDF parameters of the conditional model (λ_t) with those of GKR unconditional model (λ). Two exercises are performed here. First, we compare the mean of the conditional parameters with the unconditional one; second, we compare at every month (or every quarter) for both types of estimators.

Average conditional parameter vs unconditional parameter

Our aim here is to see if the conditional parameters can be presented as a result of a temporal disaggregation of the information brought by the unconditional parameter. For that purpose, we test if the mean of the conditional estimators, is equal through time, to the value of the unconditional estimator.

Table 2.4 presents the p-values from the test of $H_0 : E(\lambda_t) = \lambda$. The asymptotic variance used for the test is estimated through bootstrap; the table provides the results for different number of replications in the bootstrap process (599, 999, and 9999). It appears that, except for the market factor *Mkt-Rf* in CAPM, FF3, and FF5, the null is globally accepted. So, on average and for most factors, the unconditional SDF parameter is the mean of the conditional SDF parameters. This means that, for some factors, the information brought by the conditional SDF parameters proposed in this paper is a temporal disaggregation of the information brought by the unconditional parameter of Kan & Robotti (2009). Nonetheless, as the test is rejected for the market factor in three different SDFs, this result cannot be generalized. This indicates that, even if there is a link between the unconditional and the conditional SDF parameters, the information brought by both types of parameters is, not surprisingly, very different. Note in this regard that the unconditional model can be viewed as a particular case of the conditional model, with one instrument ($p = 1$) and $Z_t = 1$ at each period t.

Conditional parameters vs unconditional parameters

For each SDF benchmark, we aim to observe here how often the conditional parameter is equal to the unconditional parameter. The idea is to make sure that the difference between the two is not caused by noise.

Table 2.5 summarizes the information on the p-values from the tests $H_0 : \lambda_t = \lambda$ and gives the frequency at which the null is rejected (a test is performed every month or quarter, so we have T decisions at the end). Like previously, the asymptotic variances used for the tests are estimated through bootstrap. We observe that in all the SDFs, the *probability of rejection*[22] is non-null at least for one factor other than the constant factor. This confirms that there is a real difference between our conditional parameter and the non-conditional parameter from GKR. Note that in each SDF with only traded factors (CAPM, FF3, CARH, and FF5), the market factor *Mkt-Rf* is the factor with the highest probability of rejection (respectively 0.74, 0.62 0.52, 0.62). On the other hand, factors for which the difference between both estimators is never significant are: the size factor *SMB* in FF3 and CARH, the profitability factor *RMW* in FF5, the non-durable consumption factor *Cnd* in CCAPM and D-CCAPM, the labor income factor *Lab* in C-LAB, and the market factor *Mkt-Rf* in D-CCAPM and C-LAB.

[22]What we call here probability of rejection is the ratio between the number of rejection through time, and the overall number of periods (months or quarters)

Table 2.4: p-values from the test of $H_0 : E(\lambda_t) = \lambda$

		Monthly models		
		$nrep = 599$	$nrep = 999$	$nrep = 9999$
CAPM	Int	0.114	0.114	0.106
	Mkt-Rf	0.007	0.007	0.007
FF3	Int	0.084	0.083	0.086
	Mkt-Rf	0.010	0.011	0.011
	SMB	0.543	0.543	0.558
	HML	0.174	0.168	0.161
CARH	Int	0.211	0.206	0.200
	Mkt-Rf	0.118	0.109	0.105
	SMB	0.461	0.457	0.475
	HML	0.681	0.673	0.666
	Mom	0.495	0.498	0.491
FF5	Int	0.308	0.313	0.310
	Mkt-Rf	0.004	0.005	0.005
	SMB	0.987	0.987	0.987
	HML	0.216	0.212	0.212
	RMW	0.600	0.599	0.595
	CMA	0.523	0.528	0.526
		Quarterly models		
		$nrep = 200$	$nrep = 600$	$nrep = 1000$
CCAPM	Int	0.487	0.495	0.492
	Cnd	0.503	0.510	0.509
D-CCAPM	Int	0.049	0.055	0.052
	Mkt-Rf	0.326	0.326	0.337
	Cnd	0.527	0.531	0.535
	Cd	0.077	0.079	0.090
C-LAB	Int	0.465	0.455	0.459
	Mkt-Rf	0.701	0.693	0.699
	Cay	0.208	0.206	0.204
	Lab	0.770	0.767	0.768

The table presents the results of the test of the equality between the mean of the conditional SDF parameters and the unconditional parameter of GKR. The asymptotic variances are estimated through bootstrap; *nrep* bootstrap replications are performed. By defining $M = E(\lambda_t) - \lambda$, the test is equivalent to $H_0 : M = 0$.

At each replication s, a bootstrap sample is generated and the matrix $\Lambda^{*,s}$ is estimated. The conditional parameter at t is estimated through the relation $\lambda_t^{*,s} = \Lambda^{*,s} Z_{t-1}$. The unconditional parameter $\lambda^{*,s}$ is estimated by replicating GKR. Thereafter, we define $M^{*,s} = \frac{1}{T}\sum_{t=1}^{T}\lambda_t^{*,s} - \lambda^{*,s}$. The asymptotic variance of M is therefore given by $\frac{1}{nrep-1}\sum_{s=1}^{nrep}(M^{*,s} - \bar{M}^*)(M^{*,s} - \bar{M}^*)'$, with $\bar{M}^* = \frac{1}{nrep}\sum_{s=1}^{nrep} M^{*,s}$

Table 2.5: Summary of the results from the tests $H_0 : \lambda_t = \lambda$

		Monthly models					
		p-values					rejection of the null
			percentiles				(number of periods
		mean	20	40	60	80	over 642)
CAPM	Int	0.247	0.002	0.037	0.140	0.593	299 (0.47)
	Mkt-Rf	0.092	0	0.001	0.008	0.120	475 (0.74)
FF3	Int	0.258	0.005	0.051	0.192	0.551	254 (0.40)
	Mkt-Rf	0.168	0.001	0.005	0.041	0.353	397 (0.62)
	SMB	0.627	0.466	0.523	0.650	0.834	0 (0)
	HML	0.480	0.220	0.355	0.556	0.764	20 (0.03)
CARH	Int	0.261	0.019	0.099	0.243	0.477	211 (0.33)
	Mkt-Rf	0.209	0.004	0.015	0.108	0.481	335 (0.52)
	SMB	0.499	0.261	0.405	0.533	0.724	0 (0)
	HML	0.529	0.181	0.445	0.649	0.851	19 (0.03)
	Mom	0.338	0.098	0.177	0.339	0.623	70 (0.11)
FF5	Int	0.347	0.053	0.175	0.356	0.668	123 (0.19)
	Mkt-Rf	0.149	0.001	0.004	0.042	0.313	396 (0.62)
	SMB	0.137	0.013	0.023	0.052	0.195	380 (0.59)
	HML	0.182	0.013	0.030	0.096	0.338	319 (0.50)
	RMW	0.438	0.187	0.285	0.480	0.723	0 (0)
	CMA	0.243	0.028	0.065	0.192	0.486	233 (0.36)

		Quarterly models					
		p-values					rejection of the null
			percentiles				(number of periods
		mean	20	40	60	80	over 214)
CCAPM	Int	0.523	0.273	0.436	0.605	0.833	10 (0.05)
	Cnd	0.514	0.336	0.408	0.531	0.707	0 (0)
D-CCAPM	Int	0.344	0.095	0.183	0.328	0.644	25 (0.12)
	Mkt-Rf	0.511	0.244	0.336	0.606	0.805	0 (0)
	Cnd	0.633	0.389	0.540	0.733	0.866	0 (0)
	Cd	0.298	0.031	0.095	0.288	0.600	62 (0.29)
C-LAB	Int	0.358	0.076	0.178	0.389	0.675	18 (0.08)
	Mkt-Rf	0.423	0.173	0.289	0.438	0.679	0 (0)
	Cay	0.403	0.073	0.214	0.520	0.736	31 (0.14)
	Lab	0.568	0.407	0.511	0.593	0.739	0 (0)

The table presents the results from the tests of the equality between the conditional SDF parameters and the unconditional parameter by GKR; note that there are T tests, as we have T periods. The asymptotic variances are estimated through bootstrap; *nrep* bootstrap replications are performed. By defining $D_t = \lambda_t - \lambda$, the test at period t is equivalent to $H_0 : D_t = 0$.

At each replication s, a bootstrap sample is generated and the matrix $\Lambda^{*,s}$ is estimated. The conditional parameter at t is estimated through the relation $\lambda_t^{*,s} = \Lambda^{*,s} Z_{t-1}$. The unconditional parameter $\lambda^{*,s}$ is estimated by replicating GKR. Thereafter, we define at every period t, $D_t^{*,s} = \lambda_t^{*,s} - \lambda^{*,s}$. The asymptotic variance of D_t is therefore given by $\frac{1}{nrep-1} \sum_{s=1}^{nrep} (D_t^{*,s} - \bar{D_t}^*)(D_t^{*,s} - \bar{D_t}^*)'$, with $\bar{D_t}^* = \frac{1}{nrep} \sum_{s=1}^{nrep} D_t^{*,s}$

The numbers in brackets are the probabilities for each factor, to have a significant difference between the two types of estimators. They are obtained by the ratios between the number of periods where the null is rejected, and the overall number of periods. The tests are performed at 5% level of significance.

2.4.4 Dynamical relevance of the factors

We analyze the relevance of each factor in the selected benchmark models. This relevance is assessed through the significance of the SDF parameter, which is tested at each period t of the samples (642 periods for monthly SDFs and 214 periods for quarterly SDFs). For each factor, the *relevance frequency* is the ratio between the number of periods where its SDF parameter is significant and the overall number of periods. This relevance frequency is presented below either as a probability or as a percentage.

Traded and non-traded factors

Table 2.6 presents the results for the dynamical relevance of the factors in each of the seven SDFs used as benchmark. We see that except for the CCAPM, all of the benchmarks are relevant for the linear pricing of the selected portfolios. In fact, all of the factors are significant during some periods between 1963 and 2016. With the exception of factors with relevance frequencies equal to 0, the least significant factors are the consumption-wealth factor *Cay* in C-LAB and the durable consumption factor *Cd* in D-CCAPM; both are relevant 28% of the time.

Not surprisingly, SDFs with only traded factors are those with the highest dynamical significance. However, we can see that in some of these SDFs, not all the traded factors are significant through time. This is particularly the case for the size factor *SMB* in FF3 and CARH, and for the profitability factor *RMW* in FF5; the relevance frequency of the two factors is equal to 0 in each of these SDFs. Moreover, we see that adding a supplementary factor to a prior SDF does not always have the same consequences on the relevance of the prior factors. For example, the transition CAPM $\rightarrow$ FF3 leads to a decrease of the relevance frequency of the market factor *Mkt-Rf*, from 0.64 to 0.58, whereas the transition FF3 $\rightarrow$ FF5 leads to an increase of the relevance frequency of the same factor, from 0.58 to 0.67.

Table 2.6: Dynamical relevance of the factors

		Monthly models						
		p-values					significance at 5% (over 642 periods)	significance at 10% (over 642 periods)
			percentiles					
		mean	20	40	60	80		
CAPM	Int	0.267	0.016	0.078	0.264	0.556	199 (0.31)	290 (0.45)
	Mkt-Rf	0.181	0.001	0.009	0.062	0.374	371 (0.58)	414 (0.64)
FF3	Int	0.259	0.012	0.088	0.185	0.519	192 (0.30)	277 (0.43)
	Mkt-Rf	0.204	0.002	0.019	0.123	0.435	306 (0.48)	370 (0.58)
	SMB	0.578	0.388	0.465	0.611	0.770	0 (0)	0 (0)
	HML	0.198	0.011	0.067	0.147	0.399	222 (0.35)	307 (0.48)
CARH	Int	0.212	0.003	0.045	0.144	0.447	268 (0.42)	345 (0.54)
	Mkt-Rf	0.234	0.006	0.038	0.196	0.502	269 (0.42)	328 (0.51)
	SMB	0.598	0.408	0.545	0.637	0.783	0 (0)	0 (0)
	HML	0.207	0.022	0.086	0.169	0.309	190 (0.30)	272 (0.42)
	Mom	0.142	0.008	0.018	0.035	0.243	416 (0.65)	459 (0.71)
FF5	Int	0.312	0.059	0.208	0.328	0.553	123 (0.19)	168 (0.26)
	Mkt-Rf	0.157	0.001	0.008	0.061	0.285	363 (0.57)	433 (0.67)
	SMB	0.251	0.098	0.146	0.216	0.358	0 (0)	148 (0.23)
	HML	0.288	0.064	0.135	0.257	0.495	91 (0.14)	195 (0.30)
	RMW	0.511	0.320	0.396	0.525	0.696	0 (0)	0 (0)
	CMA	0.319	0.026	0.126	0.341	0.645	197 (0.31)	234 (0.36)
		Quarterly models						
		p-values					significance at 5% (over 214 periods)	significance at 10% (over 214 periods)
			percentiles					
		mean	20	40	60	80		
CCAPM	Int	0.729	0.553	0.669	0.834	0.941	0 (0)	0 (0)
	Cnd	0.452	0.254	0.315	0.450	0.655	0 (0)	0 (0)
D-CCAPM	Int	0.745	0.586	0.694	0.804	0.891	0 (0)	0 (0)
	Mkt-Rf	0.693	0.512	0.677	0.782	0.897	0 (0)	0 (0)
	Cnd	0.559	0.335	0.471	0.614	0.794	0 (0)	0 (0)
	Cd	0.375	0.074	0.196	0.401	0.695	5 (0.02)	60 (0.28)
C-LAB	Int	0.537	0.239	0.456	0.611	0.813	0 (0)	0 (0)
	Mkt-Rf	0.520	0.206	0.421	0.607	0.820	0 (0)	0 (0)
	Cay	0.293	0.075	0.149	0.294	0.460	23 (0.11)	61 (0.28)
	Lab	0.516	0.200	0.380	0.572	0.859	0 (0)	0 (0)

The table presents the dynamics of the relevance of the factors in the asset-pricing models used as benchmark. Every period t, we perform a bilateral test on $H_0 : \lambda_t = 0$; the "p-values" columns give the dynamics of the p-values of these tests.

The numbers in brackets are the probabilities for each factor, to have a significant SDF parameter. They are obtained by the ratios between the number of periods where the SDF parameters are significant, and the overall number of periods.

As the results confirm the advantage of using SDFs with traded factors, they bring a nuance on non-traded factors, as it appears that the C-LAB and the D-CCAPM carry some relevance for linear pricing. Although none of the factors of these two SDFs survives during the selection process of GKR, our conditional model shows that these SDFs possess some linear pricing abilities. The consumption-wealth ratio *Cay* of C-LAB and the durable consumption *Cd* of D-CCAPM are priced 28% of the time.

The results on the CCAPM are striking, as they show that this SDF is not able to provide any linear risk-adjustment in connection with the selected portfolios. So, unlike C-LAB and D-CCAPM, the results from GKR regarding the irrelevance of CCAPM for capturing linear risk-adjustment are completely confirmed by our conditional analysis. Since any SDF is theoretically linked to consumption through the Euler identity, this result calls for further analysis of the relevance of the CCAPM for pricing non-linear risk-adjustments.

We know that the relevance of a given factor depends not only on that factor, but also on the other factors in the SDF. We may have a factor which is relevant in a given SDF but not in another one. This result is confirmed by our model, as we observe that the market excess return is not relevant in D-CCAPM and in C-LAB, even if it is relevant in CAPM, FF3, CARH and FF5. In order to see in each of our benchmarks, how the relevance of prior factors evolves after an addition of new factors in the SDF, we perform a case by case analysis.

Case by case analysis: transitions in monthly SDFs

In this part, we present an analysis of the four SDFs with only traded factors (CAPM, FF3, CARH and FF5). For each factor in these SDFs, we want to see how the dynamics (and the relevance) of the SDF parameters change when we move from one SDF to another. We analyze the following transitions: *CAPM to FF3*, *FF3 to CARH*, and *FF3 to FF5*. We only consider SDFs with traded factors here, because they share several factors.

CAPM to FF3. The addition of the size and value factors (*SMB* and *HML*) to the initial market factor *Mkt-Rf* does not change the dynamic of the SDF parameters associated with the latter. Figures 2.3 and 2.4 show that *Mkt-Rf* has the same dynamics in both CAPM and FF3. However, its relevance frequency decreases during the transition, as presented in Table 2.6 (64% to 58%).

GKR show though their unconditional model that, in FF3, the size factor *SMB* is irrelevant, while the market factor *Mkt-Rf* and the value factor *HML* are relevant; our results concur with their conclusion. Indeed, in FF3, the relevance frequencies of *Mkt-Rf* and *HML* are respectively 0.58 and 0.48; while the relevance frequency of *SMB* is 0. Note that since the late 1990s the value factor is mainly irrelevant, unlike the market factor. It should also be noted that during the last US recession, the market factor has been relevant in some months whereas the value factor has been mainly irrelevant.

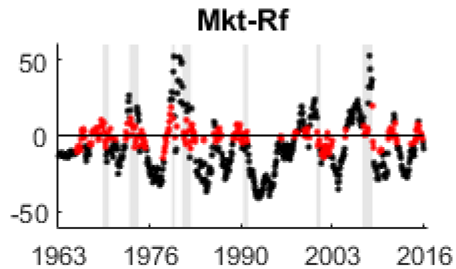
FF3 to CARH. Adding the momentum factor *Mom* to the three factors of FF3 leads to an important modification of the dynamics of the SDF parameters associated with the size factor *SMB* and the value factor *HML*. It also results in a slight modification of the dynamics of the SDF parameters associated with the market factor *Mkt-Rf* (see Figures 2.4 and 2.5). These changes suggest that a part of the information explained by the additional momentum factor was already explained by the three factors of FF3. In other words, there is a correlation between the momentum factor and the FF3 factors.

The relevance frequency of the market factor (from 0.58 to 0.51) and the value factor (from 0.48 to 0.42) both decrease after the transition, whereas the relevance of the size factor remains the same at 0%. So, although the additional momentum factor has an important relevance frequency (0.71), we cannot conclude here (at least not yet), which of the SDF between FF3 and CARH is better. A global comparison between the SDFs will be presented later.

FF3 to FF5. As previously, the addition of the profitability factor *RMW* and the investment factor *CMA* to the three factors of FF3 leads to changes in the dynamics of the SDF parameters associated with the FF3 factors. However, in this case the changes are more pronounced, as the volatility of the SDF parameters increases drastically (Figures 2.4 and 2.6).

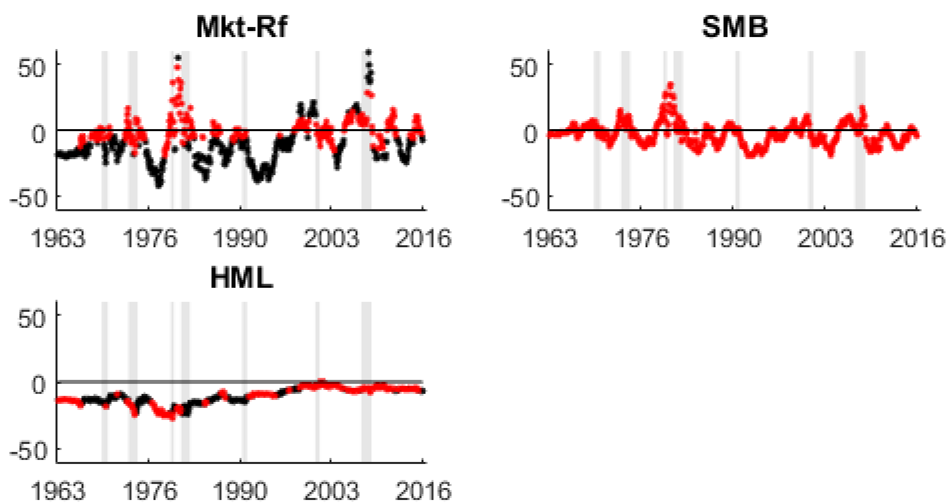
Again, similar to the previous transition, we cannot determine yet whether FF5 is better than FF3, as Table 2.6 shows that the relevance frequency of the value factor *HML* decreases after the transition (0.42 to 0.30). However, it should be noted that the relevance frequency of the size factor *SMB*, with a value of 0.23, is no longer null in FF5.

Figure 2.3: Evolution of the SDF-parameters (CAPM)



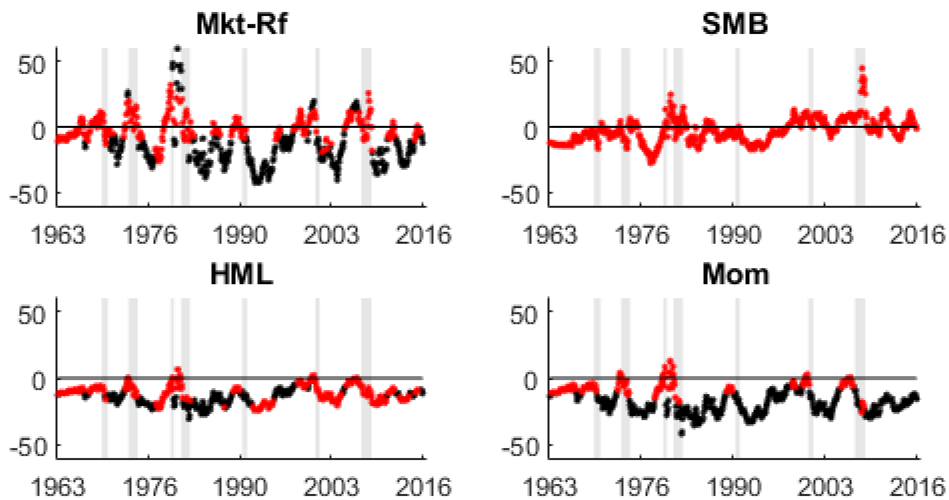
The figure presents the evolution of the conditional SDF-parameters associated with the market factor *Mkt-Rf* in CAPM. The red dots are the insignificant SDF-parameters, according to the tests performed every period. The straight horizontal line is the line $y = 0$; and the vertical shaded bands represent US recessions, according to the National Bureau of Economic Research (NBER).

Figure 2.4: Evolution of the SDF-parameters (FF3)



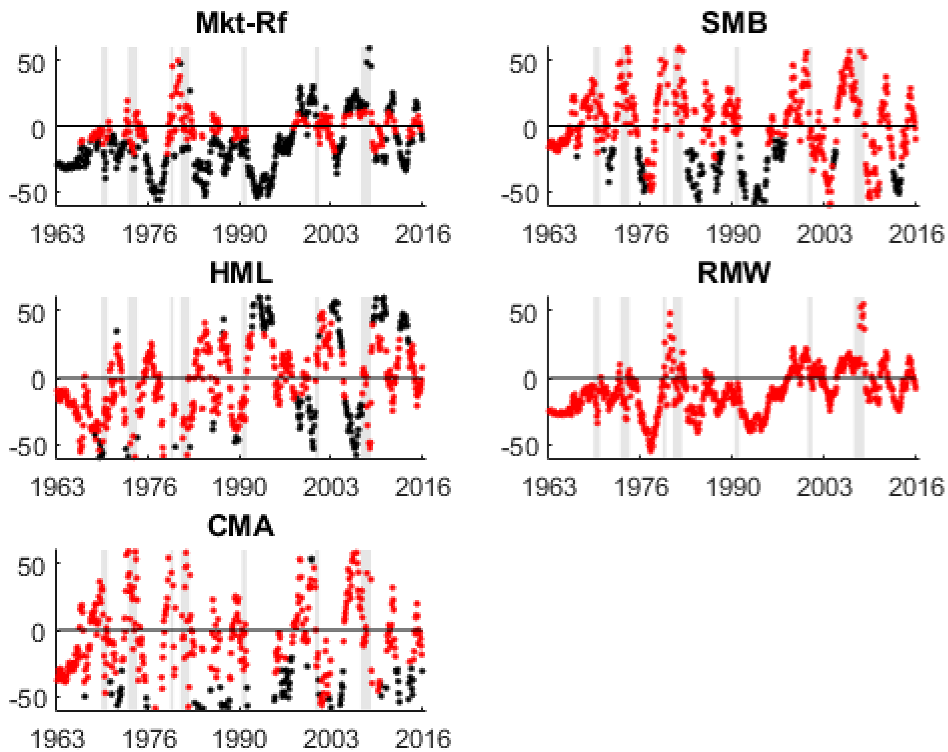
The figure presents the evolution of the conditional SDF-parameters associated with the factors of FF3. The red dots are the insignificant SDF-parameters, according to the tests performed every period. The straight horizontal line is the line $y = 0$; and the vertical shaded bands represent US recessions, according to the National Bureau of Economic Research (NBER).

Figure 2.5: Evolution of the SDF-parameters (CARH)



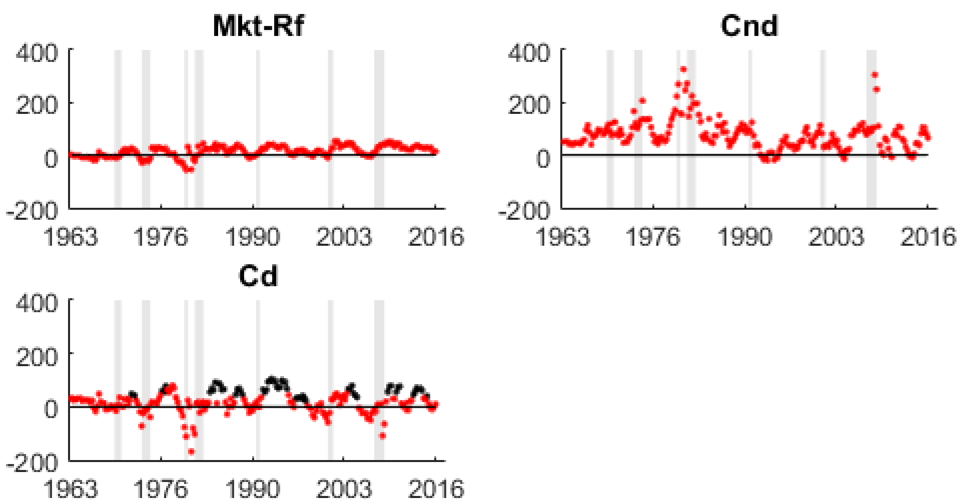
The figure presents the evolution of the conditional SDF-parameters associated with the factors of CARH. The red dots are the insignificant SDF-parameters, according to the tests performed every period. The straight horizontal line is the line $y = 0$; and the vertical shaded bands represent US recessions, according to the National Bureau of Economic Research (NBER).

Figure 2.6: Evolution of the SDF-parameters (FF5)



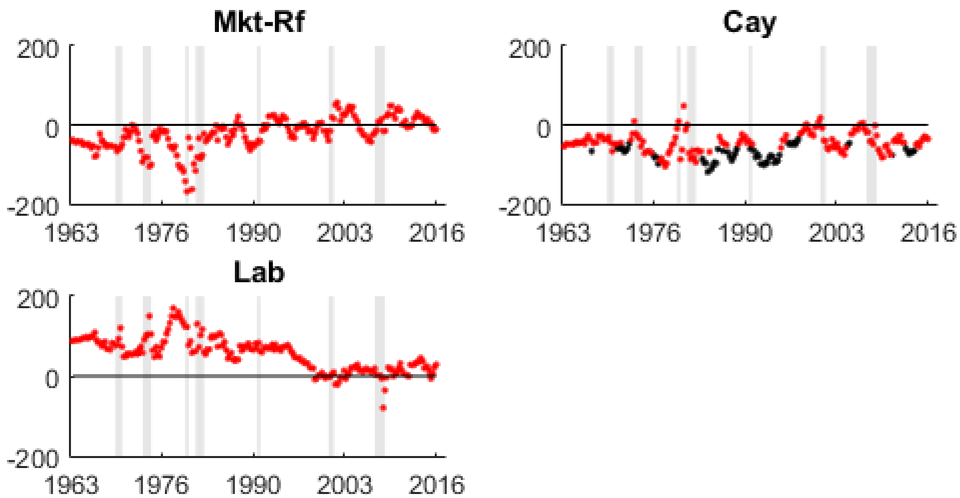
The figure presents the evolution of the conditional SDF-parameters associated with the factors of FF5. The red dots are the insignificant SDF-parameters, according to the tests performed every period. The straight horizontal line is the line $y = 0$; and the vertical shaded bands represent US recessions, according to the National Bureau of Economic Research (NBER).

Figure 2.7: Evolution of the SDF-parameters (D-CCAPM)



The figure presents the evolution of the conditional SDF-parameters associated with the factors of D-CCAPM. The red dots are the insignificant SDF-parameters, according to the tests performed every period. The straight horizontal line is the line $y = 0$; and the vertical shaded bands represent US recessions, according to the National Bureau of Economic Research (NBER).

Figure 2.8: Evolution of the SDF-parameters (C-LAB)



The figure presents the evolution of the conditional SDF-parameters associated with the factors of C-LAB. The red dots are the insignificant SDF-parameters, according to the tests performed every period. The straight horizontal line is the line $y = 0$; and the vertical shaded bands represent US recessions, according to the National Bureau of Economic Research (NBER).

2.4.5 Dynamical relevance of the SDFs: general comparisons

A general comparison of the benchmarks is performed here. As a comparison criterion, we use the number of periods where at least one factor other than the constant factor in the SDF is priced (relevant months or quarters). We consider here that the higher this number, the better the SDF. We will subsequently consider the *relevance frequency of the SDF* as the ratio between the number of periods where at least one factor other than the constant factor in the SDF is priced, and the total number of periods. The selected criteria give the frequency at which each SDF is able to price at least a part of the linear adjustment induced by the selected portfolios.

We will use for the current analysis Figures 2.9 and 2.10, which provide respectively, for all the monthly SDFs and quarterly SDFs, dates where no factor in the SDF other than the constant factor is relevant (irrelevant months or quarters). The figures combined summarize information about the relevance of each SDF and provide the dates for which each of them can be considered as irrelevant. These two figures are completed for the purposes of our analysis by Table 2.7, which presents the relevance frequency of the SDFs in the last decades.

SDFs with only traded factors

On Figure 2.9, we observe that FF3 and FF5 are the two best SDFs, as these are the two SDFs with the smallest number of irrelevant months (respectively 128 and 125 months). Therefore, FF3 is not surprisingly an overall better SDF than CAPM. Moreover, we see that adding a new traded factor to FF3 does not guarantee a better model; CARH has more irrelevant months than FF3 while FF5 and FF3 have almost the same number of irrelevant months. Nonetheless, it should be noted that the relevance of the benchmarks follows various dynamics.

Table 2.7 shows that, until the 1990s, FF3 was the best SDF. Since the 2000s, FF5 and CARH have better relevance frequencies. FF5 was better than CARH until the 1970s, but since then it is no longer the case. Thus, although FF3 and FF5 have been overall the best SDFs in the last 50 years, the dynamical analysis shows that CARH is the best SDF since the year 2000, followed by FF5, which is the second best in that period.

SDFs with non-traded factors.

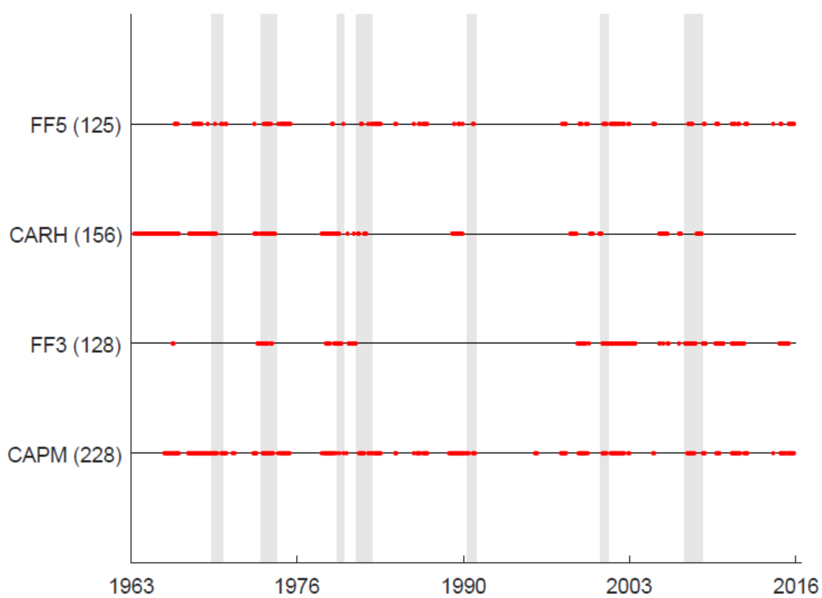
Figure 2.10 shows that the number of irrelevant quarters is quite the same between C-LAB and D-CCAPM (respectively 153 and 154). Hence, the overall relevance of these two SDFs is the same. Moreover, the dynamics of the relevance of the two SDFs do not differ much from each other. However, from Table 2.7, we see that C-LAB has been a better SDF until the 1990s. Since the 2000s, the relevance frequency of D-CCAPM is higher between the two.

Table 2.7: Relevance frequencies of the benchmarks over decades

	Monthly models				
	overall	60s-70s	80s	90s	00s-10s
CAPM	0.645	0.540	0.633	0.783	0.672
FF3	0.801	0.889	0.917	0.933	0.569
CARH	0.757	0.480	0.842	0.925	0.877
FF5	0.805	0.803	0.783	0.917	0.755
	Quarterly models				
	overall	60s-70s	80s	90s	00s-10s
CCAPM	0	0	0	0	0
D-CCAPM	0.280	0.091	0.325	0.550	0.279
C-LAB	0.285	0.182	0.400	0.600	0.132

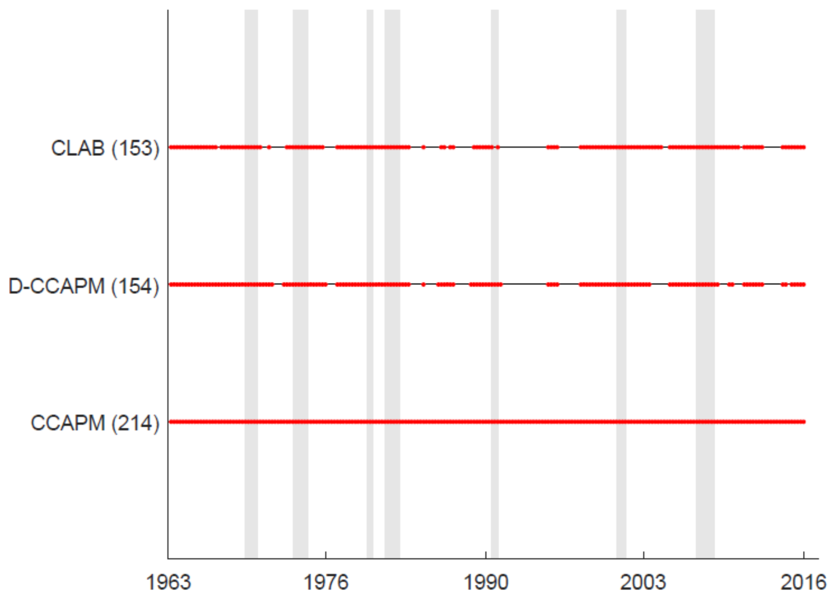
The table gives the relevance frequencies of the SDFs in the last decades. For each decade, the relevance frequency is estimated as the ratio between the number of relevant months (or quarters) in the decade, and the total number of months (or quarters) in the same decade. The overall relevance frequencies are those estimated on all the sample. For each SDF, the relevant periods (months or quarters) are the periods where at least one factor other than the constant factor in the SDF is priced.

Figure 2.9: Relevance of the monthly SDFs through time



The figure presents the evolution of the relevance of the monthly SDFs through time. The red dots are the months where no factor in the SDF other than the constant factor is significant, and the numbers in brackets are the total number of these months (there are 642 months in the sample). The vertical shaded bands represent US recessions, according to the National Bureau of Economic Research (NBER)

Figure 2.10: Relevance of the quarterly SDFs through time



The figure presents the evolution of the relevance of the quarterly SDFs through time. The red dots are the quarters where no factor in the SDF other than the constant factor is significant, and the numbers in brackets are the total number of these quarters (there are 214 quarters in the sample). The vertical shaded bands represent US recessions, according to the National Bureau of Economic Research (NBER)

2.5 Conclusion

There are always some empirical differences between the priced SDF models and the admissible ones. These differences generate pricing errors, which must be taken into account during inferences processes on SDF parameters, otherwise some irrelevant factors may be mistakenly presented as priced. Moreover, as markets strategies are dynamics, the SDFs parameters must be analyzed under a dynamical perspective (with conditional SDF models), and subsequent inference strategies must be developed.

In this paper, we develop a misspecification-robust inference method for conditional SDF models. Our inference method is designed for time-varying SDF parameters. Simulations show that even in an extreme case where a useless factor is included in the model, our method is able to eliminate the negative effects of the useless factor and restore the relevance of the useful factors.

Empirically, we apply our method on seven popular SDF models, and analyze their dynamical relevance from 1963 to 2016. For models with only traded factors, the results show that FF3 and FF5 have been the overall best SDFs in the last 50 years. However, since 2000, the best SDF is CARH. These results confirm that having more factors does not systematically lead to better SDFs; an SDF with three factors can outperform another with four or five factors. For models with non-traded factors, the results show that C-LAB and D-CCAPM carry some relevance for linear pricing, while CCAPM is not able to provide any linear-risk adjustment throughout the period of the analysis.

Appendices for Chapter 2 (B)

B1 Robustness checks; bootstrapped variances

The time-varying SDF parameters presented in this paper rely on estimation of the covariance matrix $V = Cov(Y_t)$, with $Y_t = (Z'_{t-1} \vdots\ f'_t \otimes Z'_{t-1} \vdots\ R'_t)'$ (see equation (2.7)). We need to robustly estimate this covariance, so that the model will produce reliable empirical results.

The results presented in the paper are obtained by estimating V as a robust order 2 matrix. Instead of using a conventional covariance matrix, this choice is the one which give the better results for the data used and according to the asymptotic variances presented in propositions 1 and 2. In this section we test the robustness of this strategy. Specifically, using non-parametric methods, we estimate V here as a conventional covariance matrix and perform all the *relevance tests*[23] presented in the paper. For each test here, asymptotic variances are estimated by bootstrap.

Table B.1 shows the relevance frequencies of factors in each of the benchmarks. The results obtained in the paper are confirmed on several aspects. The size factor *SMB* is never relevant in FF3 and CARH, while the profitability factor *RMW* is never relevant in FF5. On SDFs with non-tradable factors, the results about factors with a 0% relevance frequency are also confirmed. However, it should be noted in all these SDFs the constant factor carry some relevance (0.44% in CCAPM, 8% in D-CCAPM and 49% in C-LAB). Nonetheless, as the constant factor is not considered as a factor of interest in our analysis[24], this is not an issue. Therefore, we can also conclude here that D-CCAPM and C-LAB are able to provide linear risk-adjustments while CCAPM cannot.

Results from Table B.2 confirm that since the 2000s CARH is the best SDF, followed by FF5. However, FF5 is still a better overall SDF. On the other hand, results on SDFs with non-tradable factors show that C-LAB is a better SDF than D-CCAPM (overall and over the last decades). This contradicts a bit the results obtained in the paper, as we concluded that the two SDFs have a similar overall

[23]What we call relevance tests here are the tests about the relevance frequencies of factors in the selected benchmarks

[24]Recall that the relevance frequency of a SDF is defined in the paper as: the ratio between the number of periods where at least one factor other than the constant factor in the SDF is priced and the total number of periods

relevance and that D-CCAPM is better since the 2000s (C-LAB being better until the 1990s). This difference can be explained by the fact that while the relevance frequency of the durable consumption factor *Cd* in D-CCAPM is the same in the paper as in here (0.28), the relevance frequency of the consumption-wealth ratio *Cay* in C-LAB has increased from 0.28 in the paper, to 0.56 in the current analysis.

Table B.1: Dynamical relevance of the factors (bootstrapped variances)

		Monthly models					
		p-values					rejection of the null
			percentiles				(number of periods
		mean	20	40	60	80	over 642)
CAPM	Int	0.085	0	0	0	0.020	537 (0.84)
	Mkt-Rf	0.115	0	0	0.005	0.155	465 (0.72)
FF3	Int	0.062	0	0	0.001	0.022	527 (0.82)
	Mkt-Rf	0.145	0	0.003	0.033	0.262	406 (0.63)
	SMB	0.566	0.339	0.436	0.658	0.803	0 (0)
	HML	0.201	0.016	0.067	0.131	0.386	224 (0.35)
CARH	Int	0.029	0	0	0	0.026	551 (0.86)
	Mkt-Rf	0.173	0	0.005	0.079	0.366	351 (0.55)
	SMB	0.486	0.266	0.384	0.515	0.728	0 (0)
	HML	0.189	0.005	0.045	0.118	0.339	265 (0.41)
	Mom	0.127	0.001	0.004	0.015	0.226	429 (0.67)
FF5	Int	0.028	0	0	0.002	0.037	539 (0.84)
	Mkt-Rf	0.130	0	0.002	0.026	0.191	420 (0.65)
	SMB	0.135	0.010	0.020	0.068	0.217	356 (0.55)
	HML	0.194	0.015	0.039	0.113	0.351	296 (0.46)
	RMW	0.373	0.139	0.233	0.388	0.593	0 (0)
	CMA	0.238	0.009	0.050	0.188	0.492	257 (0.40)
		Quarterly models					
		p-values					rejection of the null
			percentiles				(number of periods
		mean	20	40	60	80	over 214)
CCAPM	Int	0.149	0.006	0.043	0.121	0.268	95 (0.44)
	Cnd	0.334	0.134	0.176	0.275	0.580	0 (0)
D-CCAPM	Int	0.437	0.098	0.225	0.608	0.799	17 (0.08)
	Mkt-Rf	0.569	0.321	0.430	0.642	0.832	0 (0)
	Cnd	0.406	0.196	0.241	0.387	0.656	0 (0)
	Cd	0.309	0.034	0.107	0.303	0.591	59 (0.28)
C-LAB	Int	0.191	0.003	0.015	0.103	0.396	105 (0.49)
	Mkt-Rf	0.403	0.105	0.255	0.438	0.687	0 (0)
	Cay	0.166	0.002	0.014	0.079	0.282	119 (0.56)
	Lab	0.329	0.083	0.166	0.296	0.613	0 (0)

The table presents the dynamics of the relevance of the factors in the asset-pricing models used as benchmark. Every period t, we perform a bilateral test on $H_0 : \lambda_t = 0$; the "p-values" columns give the dynamics of the p-values of these tests. The asymptotic variances are estimated through bootstrap; *nrep* bootstrap replications are performed.

At each replication s, a bootstrap sample is generated and the matrix $\Lambda^{*,s}$ is estimated. Then the conditional parameters are estimated through the relation $\lambda_t^{*,s} = \Lambda^{*,s} Z_{t-1}$. The asymptotic variance of λ_t is therefore given by $\frac{1}{nrep-1}\sum_{s=1}^{nrep}(\lambda_t^{*,s} - \bar{\lambda}_t^{*})(\lambda_t^{*,s} - \bar{\lambda}_t^{*})'$, with $\bar{\lambda}_t^{*} = \frac{1}{nrep}\sum_{s=1}^{nrep}\lambda_t^{*,s}$.

The numbers in brackets are the probabilities for each factor, to have a significant SDF parameter. They are obtained by the ratios between the number of periods where the SDF parameters are significant, and the overall number of periods. The tests are performed at 5% level of significance.

Table B.2: Relevance frequencies of the benchmarks over decades (bootstrapped variances)

	Monthly models				
	overall	60s-70s	80s	90s	00s-10s
CAPM	0.724	0.626	0.675	0.817	0.794
FF3	0.765	0.707	0.833	0.942	0.676
CARH	0.723	0.389	0.833	0.900	0.877
FF5	0.902	0.965	0.933	0.958	0.789
	Quarterly models				
	overall	60s-70s	80s	90s	00s-10s
CCAPM	0	0	0	0	0
D-CCAPM	0.275	0.091	0.325	0.550	0.265
C-LAB	0.556	0.454	0.700	0.775	0.441

The table gives the relevance frequencies of the SDFs in the last decades. For each decade, the relevance frequency is estimated as the ratio between the number of relevant months (or quarters) in the decade, and the total number of months (or quarters) in the same decade. The overall relevance frequencies are those estimated on all the sample. For each SDF, the relevant periods (months or quarters) are the periods where at least one factor other than the constant factor in the SDF is priced.

B2 Proofs

B2.1 Proof of Equation (2.4)

As the SDF y_t is a scalar, we have:

$$y_t R_t = R_t y_t = R_t vec(y_t^{'}) = R_t vec(\lambda_t^{'} \tilde{f}_t) = R_t vec(Z_{t-1}^{'} \Lambda^{'} \tilde{f}_t) = R_t \left((\tilde{f}_t^{'} \otimes Z_{t-1}^{'}) vec(\Lambda^{'}) \right)$$
$$= R_t \left((\tilde{f}_t^{'} \otimes Z_{t-1}^{'}) \right) vec(\Lambda^{'}) \qquad \text{(B.1)}$$

The result follows. □

B2.2 Proof of Proposition 1

We follow for this proof the argument of Kan & Robotti (2009). With the definitions of ϕ and $r_t(\phi)$ in section 2.2.3 , we have by the argument of Hansen & Singleton (1982),

$$\sqrt{T}(\hat{\phi} - \phi) \rightarrow \mathcal{N}\left(0_{\frac{(s+n)(s+n+3)}{2}}, S_0\right), \qquad \text{(B.2)}$$

with

$$s = p(k+1) \ \text{ and } \ S_0 = \sum_{j=-\infty}^{\infty} E\left(r_t(\phi) r_{t+j}(\phi)^{'}\right). \qquad \text{(B.3)}$$

We have $Y_t = \begin{pmatrix} Z_{t-1} \\ f_t \otimes Z_{t-1} \\ R_t \end{pmatrix}$, $\mu = E(Y_t) = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$, and $V = var(Y_t) = \begin{pmatrix} V_{11} & V_{22} & V_{33} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{pmatrix}$. The matrix D can be written as following:

$$\begin{aligned} D &= E\left(R_t(\tilde{f}_t^{'} \otimes Z_{t-1}^{'})\right) = E\left(R_t\left[(1,\ f_t^{'}) \otimes Z_{t-1}^{'}\right]\right) \\ &= \left(E\left(R_t Z_{t-1}^{'}\right)\ ,\ E\left(R_t[f_t^{'} \otimes Z_{t-1}^{'}]\right)\right) \\ &= \left(cov(R_t, Z_{t-1}^{'}) + E(R_t)E(Z_{t-1}^{'})\ ,\ cov(R_t, f_t^{'} \otimes Z_{t-1}^{'}) + E(R_t)E(f_t^{'} \otimes Z_{t-1}^{'})\right) \\ &= \left(V_{31} + \mu_3 \mu_1^{'}\ ,\ V_{32} + \mu_3 \mu_2^{'}\right). \qquad \text{(B.4)} \end{aligned}$$

D is a function of ϕ. Therefore, $vec(\Lambda^{'})$ is a function of ϕ. Then, by the delta method, we have:

$$\sqrt{T}(vec(\hat{\Lambda}^{'}) - vec(\Lambda_*^{'})) \rightarrow \mathcal{N}\left(0_s, \left(\frac{\partial vec(\Lambda_*^{'})}{\partial \phi^{'}}\right) S_0 \left(\frac{\partial vec(\Lambda_*^{'})}{\partial \phi^{'}}\right)^{'}\right), \qquad \text{(B.5)}$$

with $\frac{\partial vec(\Lambda_*^{'})}{\partial \phi'} = \left(\frac{\partial vec(\Lambda_*^{'})}{\partial \mu_1^{'}} \vdots \frac{\partial vec(\Lambda_*^{'})}{\partial \mu_2^{'}} \vdots \frac{\partial vec(\Lambda_*^{'})}{\partial \mu_3^{'}} \vdots \frac{\partial vec(\Lambda_*^{'})}{\partial vech(V)'} \right)$.

We know that the duplication matrix D_{s+n} is a $\left((s+n)^2 \times \frac{1}{2}(s+n)(s+n+1)\right)$ - matrix such that $vec(V) = D_{s+n} vech(V)$. Therefore,

$$\frac{\partial vec(\Lambda_*^{'})}{\partial vech(V)'} = \frac{\partial vec(\Lambda_*^{'})}{\partial vec(V)'} \cdot \frac{\partial vec(V)}{\partial vech(V)'} = \frac{\partial vec(\Lambda_*^{'})}{\partial vec(V)'} \cdot D_{s+n}. \tag{B.6}$$

It follows that $\Sigma = \left(\frac{\partial vec(\Lambda_*^{'})}{\partial \mu'} \vdots \frac{\partial vec(\Lambda_*^{'})}{\partial vec(V)'} \cdot D_{p(k+1)+n} \right) S_0 \left(\frac{\partial vec(\Lambda_*^{'})}{\partial \mu'} \vdots \frac{\partial vec(\Lambda_*^{'})}{\partial vec(V)'} \cdot D_{p(k+1)+n} \right)'$ is the asymptotic variance of $\sqrt{T}(vec(\hat{\Lambda}^{'}) - vec(\Lambda_*^{'})$. This completes the proof. ■

B2.3 Proof of Lemma 1

We have $\frac{\partial vec(\Lambda^{'})}{\partial \phi'} = \left(\frac{\partial vec(\Lambda^{'})}{\partial \mu_1^{'}} \vdots \frac{\partial vec(\Lambda^{'})}{\partial \mu_2^{'}} \vdots \frac{\partial vec(\Lambda^{'})}{\partial \mu_3^{'}} \vdots \frac{\partial vec(\Lambda^{'})}{\partial vec(V)'} \cdot D_{p(k+1)+n} \right)$. Now let us determine the expressions of the derivatives in brackets.

Expression of $\frac{\partial \mathbf{vec}(\mathbf{\Lambda}^{'})}{\partial \mu'}$

First of all let us note that

$$\begin{aligned} \frac{\partial vec(\Lambda^{'})}{\partial \mu_1^{'}} &= \frac{\partial vec(\Lambda^{'})}{\partial vec(D)'} \cdot \frac{\partial vec(D)}{\partial \mu_1^{'}} \\ \frac{\partial vec(\Lambda^{'})}{\partial \mu_2^{'}} &= \frac{\partial vec(\Lambda^{'})}{\partial vec(D)'} \cdot \frac{\partial vec(D)}{\partial \mu_2^{'}} \\ \frac{\partial vec(\Lambda^{'})}{\partial \mu_3^{'}} &= \frac{\partial vec(\Lambda^{'})}{\partial vec(D)'} \cdot \frac{\partial vec(D)}{\partial \mu_3^{'}}. \end{aligned} \tag{B.7}$$

By equation (B.4), we have

$$dD = \left(dV_{31} + (d\mu_3)\mu_1^{'} + \mu_3(d\mu_1^{'})\ ,\ dV_{32} + (d\mu_3)\mu_2^{'} + \mu_3(d\mu_2^{'}) \right) \tag{B.8}$$

$$dvec(D) = \begin{pmatrix} dvec(V_{31}) + (\mu_1 \otimes I_n) dvec(\mu_3) + (I_p \otimes \mu_3) dvec(\mu_1) \\ dvec(V_{32}) + (\mu_2 \otimes I_n) dvec(\mu_3) + (I_{kp} \otimes \mu_3) dvec(\mu_2) \end{pmatrix}\ , \tag{B.9}$$

with d, the total differential as defined by Magnus & Neudecker (2007). Therefore,

$$\begin{aligned}
\frac{\partial vec(D)}{\partial \mu_1^{'}} &= \begin{pmatrix} I_p \otimes \mu_3 \\ 0_{kpn\times p} \end{pmatrix} = \begin{pmatrix} I_p \\ 0_{kp\times p} \end{pmatrix} \otimes \mu_3 \\
\frac{\partial vec(D)}{\partial \mu_2^{'}} &= \begin{pmatrix} 0_{np\times kp} \\ I_{kp} \otimes \mu_3 \end{pmatrix} = \begin{pmatrix} 0_{p\times kp} \\ I_{kp} \end{pmatrix} \otimes \mu_3 \\
\frac{\partial vec(D)}{\partial \mu_3^{'}} &= \begin{pmatrix} \mu_1 \otimes I_n \\ \mu_2 \otimes I_n \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \otimes I_n \ .
\end{aligned} \tag{B.10}$$

Now we consider the following result;

Result 1. *If we differentiate with respect to μ to get the Jacobian matrix, then for $H = (D^{'}V_{33}^{-1}D)^{-1}$,*

$$\frac{\partial vec(\Lambda^{'})}{\partial vec(D)^{'}} = -H \otimes e^{'}V_{33}^{-1} - \left(vec(\Lambda^{'})\right)^{'} \otimes HD^{'}V_{33}^{-1} \ . \tag{B.11}$$

We deduce the expressions of $\frac{\partial vec(\Lambda^{'})}{\partial \mu_1^{'}}$, $\frac{\partial vec(\Lambda^{'})}{\partial \mu_2^{'}}$ and $\frac{\partial vec(\Lambda^{'})}{\partial \mu_3^{'}}$ through straightforward developments, using (B.7), (B.10) and (B.11).

Expression of $\frac{\partial \mathbf{vec}(\mathbf{\Lambda}^{'})}{\partial \mathbf{vec}(\mathbf{V})^{'}}$

As $vec(\Lambda^{'}) = (D^{'}V_{33}^{-1}D)^{-1}(D^{'}V_{33}^{-1}1_n) = H(D^{'}V_{33}^{-1}1_n)$, we have:

$$dvec(\Lambda^{'}) = \left(1_n^{'}V_{33}^{-1}D \otimes I_{p(k+1)}\right) dvec(H) + \left(H \otimes 1_n^{'}V_{33}^{-1}\right) dvec(D) + \left(1_n^{'} \otimes HD^{'}\right) dvec(V_{33}^{-1}). \tag{B.12}$$

Let us consider the following result

Result 2.

$$\frac{\partial vec(V_{33}^{-1})}{\partial vec(V)^{'}} = -\left(0_{n\times p(k+1)} \vdots V_{33}^{-1}\right) \otimes \left(0_{n\times p(k+1)} \vdots V_{33}^{-1}\right). \tag{B.13}$$

$$\frac{\partial vec(D)}{\partial vec(V)^{'}} = \begin{pmatrix} I_p & 0_{p\times kp} & 0_{p\times n} \\ 0_{kp\times p} & I_{kp} & 0_{kp\times n} \end{pmatrix} \otimes \left(0_{n\times p} \vdots 0_{n\times kp} \vdots I_n\right). \tag{B.14}$$

$$\begin{aligned}
\frac{\partial vec(H)}{\partial vec(V)^{'}} = -K_{p(k+1),p(k+1)} &\left[H \begin{pmatrix} I_p & 0_{p\times kp} & 0_{p\times n} \\ 0_{kp\times p} & I_{kp} & 0_{kp\times n} \end{pmatrix} \otimes \left(0_{p(k+1)\times p(k+1)} \vdots HD^{'}V_{33}^{-1}\right) \right] \\
&- H \begin{pmatrix} I_p & 0_{p\times kp} & 0_{p\times n} \\ 0_{kp\times p} & I_{kp} & 0_{kp\times n} \end{pmatrix} \otimes \left(0_{p(k+1)\times p(k+1)} \vdots HD^{'}V_{33}^{-1}\right)
\end{aligned} \tag{B.15}$$

$$+\left(0_{p(k+1)\times p(k+1)} \vdots HD^{'}V_{33}^{-1}\right)\otimes\left(0_{p(k+1)\times p(k+1)} \vdots HD^{'}V_{33}^{-1}\right).$$

By the previous result, we can make the following developments (where K is a commutation matrix as defined by Magnus & Neudecker (2007, p.54) and $s = p(k+1)$);

$$\left(1_n^{'}\otimes HD^{'}\right)\frac{\partial vec(V_{33}^{-1})}{\partial vec(V)^{'}} = -\left(0_{1\times p(k+1)} \vdots 1_n^{'}V_{33}^{-1}\right)\otimes\left(0_{p(k+1)\times p(k+1)} \vdots HD^{'}V_{33}^{-1}\right); \tag{B.16}$$

$$\left(H\otimes 1_n^{'}V_{33}^{-1}\right)\frac{\partial vec(D)}{\partial vec(V)^{'}} = \left[H\begin{pmatrix} I_p & 0_{p\times kp} \\ 0_{kp\times p} & I_{kp}\end{pmatrix} \vdots 0_{p(k+1)\times n}\right]\otimes\left(0_{1\times p(k+1)} \vdots 1_n^{'}V_{33}^{-1}\right); \tag{B.17}$$

$$\begin{aligned}
\left(1_n^{'}V_{33}^{-1}D\otimes I_s\right)\frac{\partial vec(H)}{\partial vec(V)^{'}} &= -K_{1,s}\left(I_s\otimes 1_n^{'}V_{33}^{-1}D\right)\left[H\begin{pmatrix} I_p & 0_{p\times kp} & 0_{p\times n} \\ 0_{kp\times p} & I_{kp} & 0_{kp\times n}\end{pmatrix}\otimes\left(0_{s\times s} \vdots HD^{'}V_{33}^{-1}\right)\right] \\
&\quad - 1_n^{'}V_{33}^{-1}DH\begin{pmatrix} I_p & 0_{p\times kp} & 0_{p\times n} \\ 0_{kp\times p} & I_{kp} & 0_{kp\times n}\end{pmatrix}\otimes\left(0_{s\times s} \vdots HD^{'}V_{33}^{-1}\right) \\
&\quad + 1_n^{'}V_{33}^{-1}D\left(0_{s\times s} \vdots HD^{'}V_{33}^{-1}\right)\otimes\left(0_{s\times s} \vdots HD^{'}V_{33}^{-1}\right) \\
&= -H\begin{pmatrix} I_p & 0_{p\times kp} & 0_{p\times n} \\ 0_{kp\times p} & I_{kp} & 0_{kp\times n}\end{pmatrix}\otimes\left(0_{1\times s} \vdots (vec(\Lambda^{'}))^{'}D^{'}V_{33}^{-1}\right) \\
&\quad - (vec(\Lambda^{'}))^{'}\begin{pmatrix} I_p & 0_{p\times kp} & 0_{p\times n} \\ 0_{kp\times p} & I_{kp} & 0_{kp\times n}\end{pmatrix}\otimes\left(0_{s\times s} \vdots HD^{'}V_{33}^{-1}\right) \\
&\quad + \left(0_{1\times s} \vdots (vec(\Lambda^{'}))^{'}D^{'}V_{33}^{-1}\right)\otimes\left(0_{s\times s} \vdots HD^{'}V_{33}^{-1}\right).
\end{aligned} \tag{B.18}$$

The expression of $\frac{\partial vec(\Lambda^{'})}{\partial vec(V)^{'}}$ is obtained by summing (B.16), (B.17) and (B.18). This completes the proof of Lemma 1 ■

Proof of Result 1

We have $dvec(\Lambda^{'}) = \left(1_n^{'}V_{33}^{-1}\otimes I_{p(k+1)}\right) dvec\left((D^{'}V_{33}^{-1}D)^{-1}D^{'}\right)$. The result follows by applying the results (A16) and (A17) from Kan & Robotti (2009). □

Proof of Result 2 - (B.13)

First of all, note that by the chain rule and by the theorem 3 in Magnus & Neudecker (2007, p.171),

$$\frac{\partial vec(V_{33}^{-1})}{\partial vec(V)'} = \frac{\partial vec(V_{33}^{-1})}{\partial vec(V_{33})'} \cdot \frac{\partial vec(V_{33})}{\partial vec(V)'} = \left(-V_{33}^{-1} \otimes V_{33}^{-1}\right) \cdot \frac{\partial vec(V_{33})}{\partial vec(V)'}. \tag{B.19}$$

Moreover, we have $V_{33} = AVA'$, with $A = \left(0_{n\times p(k+1)} \vdots I_n\right)$. So $\frac{\partial vec(V_{33})}{\partial vec(V)'} = A \otimes A$, and $\frac{\partial vec(V_{33}^{-1})}{\partial vec(V)'} = -\left(V_{33}^{-1}A \otimes V_{33}^{-1}A\right) = -\left(0_{n\times p(k+1)} \vdots V_{33}^{-1}\right) \otimes \left(0_{n\times p(k+1)} \vdots V_{33}^{-1}\right)$. □

Proof of Result 2 - (B.14)

We will use here the expression of $dvec(D)$ as given by (B.9). We have $V_{31} = AVB_1$ and $V_{32} = AVB_2$, with $A = \left(0_{n\times p} \vdots 0_{n\times kp} \vdots I_n\right)$, $B_1 = \left(I_p \vdots 0_{p\times kp} \vdots 0_{p\times n}\right)'$ and $B_2 = \left(0_{kp\times p} \vdots I_{kp} \vdots 0_{kp\times n}\right)'$. So $\frac{\partial vec(V_{31})}{\partial vec(V)'} = B_1' \otimes A$, $\frac{\partial vec(V_{32})}{\partial vec(V)'} = B_2' \otimes A$ and $\frac{\partial vec(D)}{\partial vec(V)'} = \begin{pmatrix} B_1' \\ B_2' \end{pmatrix} \otimes A$. The result follows. □

Proof of Result 2 - (B.15)

We note as previously $s = p(k+1)$. We have $H = (D'V_{33}^{-1}D)^{-1}$; so by the chain rule,

$$\frac{\partial vec(H)}{\partial vec(V)'} = -(H \otimes H) \cdot \frac{\partial vec(D'V_{33}^{-1}D)}{\partial vec(V)'}. \tag{B.20}$$

Let us determine the expression of $\frac{\partial vec(D'V_{33}^{-1}D)}{\partial vec(V)'}$. By applying the total differential, we have $d(D'V_{33}^{-1}D) = (dD')V_{33}^{-1}D + D'(dV_{33}^{-1})D + D'V_{33}^{-1}(dD)$. Therefore,

$$\begin{aligned}
dvec(D'V_{33}^{-1}D) &= (D'V_{33}^{-1} \otimes I_{p(k+1)})dvec(D') + \underbrace{(D' \otimes D')dvec(V_{33}^{-1}) + (I_{p(k+1)} \otimes D'V_{33}^{-1})dvec(D)}_{\Xi} \\
&= (D'V_{33}^{-1} \otimes I_{p(k+1)})K_{n,p(k+1)}dvec(D) + \Xi \\
&= K_{p(k+1),p(k+1)}(I_{p(k+1)} \otimes D'V_{33}^{-1})dvec(D) + \Xi \\
&= (K_{p(k+1),p(k+1)} + I_{(p(k+1))^2})(I_{p(k+1)} \otimes D'V_{33}^{-1})dvec(D) + (D' \otimes D')dvec(V_{33}^{-1}) \\
&= (K_{s,s} + I_{s^2})(I_s \otimes D'V_{33}^{-1})dvec(D) + (D' \otimes D')dvec(V_{33}^{-1})
\end{aligned} \tag{B.21}$$

We already have $\frac{\partial vec(V_{33}^{-1})}{\partial vec(V)'}$ and $\frac{\partial vec(D)}{\partial vec(V)'}$ (resp. in equations B.13 and B.14). By using these expressions, we have:

$$\begin{aligned}
\frac{\partial vec(D'V_{33}^{-1}D)}{\partial vec(V)'} = (K_{s,s} + I_{s^2}) &\left[\begin{pmatrix} I_p & 0_{p\times kp} & 0_{p\times n} \\ 0_{kp\times p} & I_{kp} & 0_{kp\times n} \end{pmatrix} \otimes \left(0_{s,s} \vdots D'V_{33}^{-1}\right)\right] \\
&- \left(0_{s,s} \vdots D'V_{33}^{-1}\right) \otimes \left(0_{s,s} \vdots D'V_{33}^{-1}\right).
\end{aligned} \tag{B.22}$$

By the properties of the commutation matrix, $(H \otimes H)K_{s,s} = K_{s,s}(H \otimes H)$ (as H is a $s \times s$ -matrix). Therefore by using (B.20) and(B.22), we have the result through a straightforward calculation. □

B2.4 Proof of Proposition 2

We have $\lambda_t = vec(\lambda_t^{'}) = vec(Z_{t-1}^{'}\Lambda^{'} I_{k+1}) = (I_{k+1} \otimes Z_{t-1}^{'})vec(\Lambda^{'})$; the result follows by proposition 1. ■

B2.5 Proof of Proposition 3

In this proof, I consider the useful factors as the firsts factors, whereas the useless factor is the second factor. Subsequently, I consider the following notations: $D_1 \equiv D_{us}$, $D_2 \equiv D_{\bar{us}}$, $\Lambda_1 \equiv \Lambda_{us}$ and $\Lambda_2 \equiv \Lambda_{\bar{us}}$. I analyze here both the distributions of $vec(\hat{\Lambda}_1)$ and $vec(\hat{\Lambda}_2)$, even if proposition3 is only about $vec(\Lambda_1)$.

Case 1: the SDF is correctly specified

As the factor 2 is useless, $D_2 = 0_{n \times p}$. So we have $1_n = D_1 vec(\Lambda_1^{'}) + D_2 vec(\Lambda_2^{'}) = D_1 vec(\Lambda_1^{'})$. D_1 and D_2 are expectations (as well as U, defined in assumption 2). So I use their empirical counterparts $\hat{D}_1$, $\hat{D}_2$ and D_3 (which are assumed to be consistent), and therefore consider the following linear model:

$$\begin{aligned} \hat{U}^{-\frac{1}{2}}1_n &= \hat{U}^{-\frac{1}{2}}\hat{D}_1 vec(\Lambda_1^{'}) + \hat{U}^{-\frac{1}{2}}\hat{D}_2 vec(\Lambda_2^{'}) + \varepsilon \\ &= \tilde{D}_1 vec(\Lambda_1^{'}) + \tilde{D}_2 vec(\Lambda_2^{'}) + \varepsilon. \end{aligned} \tag{B.23}$$

I consider here the matrices $\tilde{D}_1 \equiv \hat{U}^{-\frac{1}{2}}\hat{D}_1$ and $\tilde{D}_2 \equiv \hat{U}^{-\frac{1}{2}}\hat{D}_2$. I also consider the projection matrices $\hat{M}_1 = I_n - \tilde{D}_1(\tilde{D}_1^{'}\tilde{D}_1)^{-1}\tilde{D}_1^{'}$ and $\hat{M}_2 = I_n - \tilde{D}_2(\tilde{D}_2^{'}\tilde{D}_2)^{-1}\tilde{D}_2^{'}$, so I can have the two following models:

$$\hat{M}_1\hat{U}^{-\frac{1}{2}}1_n = \hat{M}_1\tilde{D}_2 vec(\Lambda_2^{'}) + \hat{M}_1\varepsilon \tag{B.24}$$

$$\hat{M}_2\hat{U}^{-\frac{1}{2}}1_n = \hat{M}_2\tilde{D}_1 vec(\Lambda_1^{'}) + \hat{M}_2\varepsilon. \tag{B.25}$$

Then I consider the following additional assumption;

Assumption 6.
(i) The error vector in the model (B.23) is such that $E(\varepsilon\varepsilon^{'}|D_1, D_2) = \Sigma$ *(conditional homoskedasticity)*
(ii) $vec\left(\sqrt{T}\hat{U}^{-\frac{1}{2}}\hat{D}_2\right) = vec\left(\sqrt{T}\tilde{D}_2\right) \xrightarrow{p} w_2 \sim \mathcal{N}(0_{np}, I_{np})$
(iii) $\sqrt{T}\hat{U}^{-\frac{1}{2}}\left(1_n - \hat{D}_1 vec(\Lambda_1^{'})\right) \xrightarrow{p} w_1 \sim \mathcal{N}(0_n, V_{w_1})$ *(when the SDF is correctly specified)*

Assumption 6(i) is a regularity condition on the errors of the model (B.23). Assumptions 6(ii and iii) are coherent with assumptions made by Gospodinov et al. (2014) (see the proof of their proposition

1 in the supplemental material). Like them I also assume that w_1 and w_2 are independent of each other, to satisfy the non-correlation between the useless factor and the useful ones. By assumption 6(ii), we have $\sqrt{T}\tilde{D}_2 = \sqrt{T}\hat{U}^{-\frac{1}{2}}\hat{D}_2 \xrightarrow{d} vec^{-1}(w_2) = W_2$; with vec^{-1} the inverse vectorization operator and W_2 a $n \times p$ - matrix.

Note that $\hat{M}_1$ is a consistent estimator of the matrix $M_1 = I_n - U^{-\frac{1}{2}}D_1(D_1^{'}U^{-1}D_1)^{-1}D_1^{'}U^{-\frac{1}{2}}$, and $\hat{M}_2$ converges toward $M_{w_2} = I_n - W_2(W_2^{'}W_2)^{-1}W_2^{'}$ (with $W_2 = vec^{-1}(w_2)$, defined as previously).

From the model (B.25), we have;

$$\begin{aligned}
\sqrt{T}(vec(\hat{\Lambda}_1^{'}) - vec(\Lambda_1^{'})) &= \sqrt{T}\left[\left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\tilde{D}_1^{'}\hat{M}_2\hat{U}^{-\frac{1}{2}}1_n - \left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1 vec(\Lambda_1^{'})\right] \\
&= \left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\tilde{D}_1^{'}\hat{M}_2\sqrt{T}\left(\hat{U}^{-\frac{1}{2}}1_n - \tilde{D}_1 vec(\Lambda_1^{'})\right) \\
&= \left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\tilde{D}_1^{'}\hat{M}_2\hat{U}^{-\frac{1}{2}}\sqrt{T}\left(1_n - \hat{D}_1 vec(\Lambda_1^{'})\right) \\
&\xrightarrow{p} \left(D_1^{'}U^{-\frac{1}{2}}M_{w_2}U^{-\frac{1}{2}}D_1\right)^{-1}D_1^{'}U^{-\frac{1}{2}}M_{w_2}U^{-\frac{1}{2}}w_1 \qquad \text{(B.26)}
\end{aligned}$$

Therefore, we have $vec(\hat{\Lambda}_1^{'}) - vec(\Lambda_1^{'}) = O_p(T^{-\frac{1}{2}})$; the consistency follows. M_{w_2} is a function of the random variable w_2, which follows a normal distribution. Since w_1 also follows a normal distribution, the asymptotic distribution of $vec(\hat{\Lambda}_1^{'})$ is given as a ratio of two distributions: the first one is the product of two normal distributions, and the second one is a normal distribution. This completes the proof for the "correctly specified case". ■

Lemma 2. *(asymptotic properties of $vec(\hat{\Lambda}_2^{'})$ when the SDF is correctly specified)*
If the SDF is correctly specified, the asymptotic behavior of the row vector associated with the useless factor is such that:

$$vec(\hat{\Lambda}_2^{'}) = O_p(1) \qquad \text{(B.27)}$$

Proof of Lemma 2. From the model (B.24), we have:

$$\begin{aligned}
vec(\hat{\Lambda}_2^{'}) &= \left(\tilde{D}_2^{'}\hat{M}_1\tilde{D}_2\right)^{-1}\tilde{D}_2^{'}\hat{M}_1\hat{U}^{-\frac{1}{2}}1_n \\
&= \left(\tilde{D}_2^{'}\hat{M}_1\tilde{D}_2\right)^{-1}\tilde{D}_2^{'}\hat{M}_1\hat{U}^{-\frac{1}{2}}\left(1_n - \hat{D}_1 vec(\Lambda_1^{'})\right) \qquad \textit{(as } \tilde{D}_1 = \hat{U}^{-\frac{1}{2}}\hat{D}_1, \textit{ and } \hat{M}_1\tilde{D}_1 = 0\textit{)} \\
&= \left((\sqrt{T}\tilde{D}_2)^{'}\hat{M}_1(\sqrt{T}\tilde{D}_2)\right)^{-1}(\sqrt{T}\tilde{D}_2^{'})\hat{M}_1\sqrt{T}\hat{U}^{-\frac{1}{2}}\left(1_n - \hat{D}_1 vec(\Lambda_1^{'})\right) \\
&\xrightarrow{d} (W_2^{'}M_1W_2)^{-1}W_2M_1w_1 \qquad \text{(B.28)}
\end{aligned}$$

Therefore, $vec(\hat{\Lambda}_2^{'}) = O_p(1)$. □

Case 2: the SDF is misspecified

Here, we do not have no more the equality $1_n = D_1 vec(\Lambda_1^{'})$, since the error vector is not null (we have $e(\Lambda_1) = D_1 vec(\Lambda_1^{'}) - 1_n = O_p(1)$). I still consider assumption 6(i and ii), as well as the matrices $\hat{M}_1$, $\hat{M}_2$, M_{w_2} and W_2 as defined previously. We have;

$$\begin{aligned} vec(\hat{\Lambda}_1^{'}) - vec(\Lambda_1^{'}) &= \left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\tilde{D}_1^{'}\hat{M}_2\hat{U}^{-\frac{1}{2}}1_n - \left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1 vec(\Lambda_1^{'}) \\ &= \left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\tilde{D}_1^{'}\hat{M}_2\left(\hat{U}^{-\frac{1}{2}}1_n - \tilde{D}_1 vec(\Lambda_1^{'})\right) \\ &\xrightarrow{p} \left(D_1^{'}U^{-\frac{1}{2}}M_{w_2}U^{-\frac{1}{2}}D_1\right)^{-1}D_1^{'}U^{-\frac{1}{2}}M_{w_2}U^{-\frac{1}{2}}\left(1_n - D_1 vec(\Lambda_1^{'})\right) \\ &\equiv \left(D_1^{'}U^{-\frac{1}{2}}M_{w_2}U^{-\frac{1}{2}}D_1\right)^{-1}D_1^{'}U^{-\frac{1}{2}}M_{w_2}U^{-\frac{1}{2}}(-e(\Lambda_1)) \end{aligned} \tag{B.29}$$

Therefore, $vec(\hat{\Lambda}_1^{'}) - vec(\Lambda_1^{'}) = O_p(1)$. This completes the proof for the "misspecified case". ■

Lemma 3. *(asymptotic properties of $vec(\hat{\Lambda}_2^{'})$ when the SDF is misspecified)*
If the SDF is misspecified, the asymptotic behavior of the row vector associated with the useless factor is such that:

$$vec(\hat{\Lambda}_2^{'}) = O_p(\sqrt{T}) \tag{B.30}$$

Proof of Lemma 3. From the model (B.24), we have:

$$\frac{1}{\sqrt{T}}vec(\hat{\Lambda}_2^{'}) = \left((\sqrt{T}\tilde{D}_2^{'})\hat{M}_1(\sqrt{T}\tilde{D}_2)\right)^{-1}(\sqrt{T}\tilde{D}_2^{'})\hat{M}_1\hat{U}^{-\frac{1}{2}}1_n \xrightarrow{p} (W_2M_1W_2)^{-1}W_2M_1U^{-\frac{1}{2}}1_n \tag{B.31}$$

Therefore, $vec(\hat{\Lambda}_2^{'}) = O_p(\sqrt{T})$ □

B2.6 Proof of Proposition 4

We still have a SDF with the constant factor, k useful factors and a useless factor. Then I still consider the notations and the additional assumption made in the previous proof.

Asymptotic properties of $\mathbf{W}(\mathbf{vec}(\hat{\mathbf{\Lambda}}_1^{'}))$

We know that:

$$\begin{aligned} vec(\hat{\Lambda}_1^{'}) &= \left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\tilde{D}_1^{'}\hat{M}_2\hat{U}^{-\frac{1}{2}}1_n \\ &= \left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\tilde{D}_1^{'}\hat{M}_2\left(\tilde{D}_1 vec(\Lambda_1^{'}) + \tilde{D}_2 vec(\Lambda_2^{'}) + \varepsilon\right) \\ &= vec(\Lambda_1^{'}) + \left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\tilde{D}_1^{'}\hat{M}_2\varepsilon \end{aligned} \tag{B.32}$$

So $vec(\hat{\Lambda}_1^{'}) - vec(\Lambda_1^{'}) = \left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\tilde{D}_1^{'}\hat{M}_2\varepsilon$. Therefore, by Assumption 5,

$$\begin{aligned}
W(vec(\hat{\Lambda}_1^{'})) &= \left[\left(\sqrt{T}\tilde{D}_1^{'}\hat{M}_2\varepsilon\right)^{'}\left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\right](\hat{\Sigma}_{11})^{-1}\left[\left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\sqrt{T}\tilde{D}_1^{'}\hat{M}_2\varepsilon\right] \\
&\xrightarrow{p} r_{\varepsilon}^{'}\left(D_1^{'}U^{-\frac{1}{2}}M_{w_2}U^{-\frac{1}{2}}D_1\right)^{-1}\left(Avar(\sqrt{T}vec(\hat{\Lambda}_1^{'}))\right)^{-1}\left(D_1^{'}U^{-\frac{1}{2}}M_{w_2}U^{-\frac{1}{2}}D_1\right)^{-1}r_{\varepsilon}\ ,
\end{aligned} \tag{B.33}$$

where $Avar(\sqrt{T}vec(\hat{\Lambda}_1^{'}))$ is the asymptotic variance of $\sqrt{T}vec(\hat{\Lambda}_1^{'})$ (remember that $\tilde{D}_1 \equiv U^{-\frac{1}{2}}\hat{D}_1$ and $\tilde{D}_2 \equiv U^{-\frac{1}{2}}\hat{D}_2$).

We have by Assumption 5, $\left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\sqrt{T}\tilde{D}_1^{'}\hat{M}_2\varepsilon \xrightarrow{p} \left(D_1^{'}U^{-\frac{1}{2}}M_{w_2}U^{-\frac{1}{2}}D_1\right)^{-1}r_{\varepsilon}$. Therefore,

$$\begin{aligned}
\Sigma_{11} = Avar(\sqrt{T}vec(\hat{\Lambda}_1^{'})) &= Avar\left[\left(\tilde{D}_1^{'}\hat{M}_2\tilde{D}_1\right)^{-1}\sqrt{T}\tilde{D}_1^{'}\hat{M}_2\varepsilon\right] \\
&= \left(D_1^{'}U^{-\frac{1}{2}}M_{w_2}U^{-\frac{1}{2}}D_1\right)^{-1}V_{r_\varepsilon}\left(D_1^{'}U^{-\frac{1}{2}}M_{w_2}U^{-\frac{1}{2}}D_1\right)^{-1}
\end{aligned} \tag{B.34}$$

Since r_{ε} follows a normal distribution, we have by replacing in (B.33):

$$W(vec(\hat{\Lambda}_1^{'})) \xrightarrow{p} r_{\varepsilon}^{'}V_{r_\varepsilon}^{-1}r_{\varepsilon} \sim \chi^2(s)\ . \tag{B.35}$$

Asymptotic properties of $\mathbf{W}(\mathbf{vec}(\hat{\mathbf{\Lambda}}_\mathbf{2}^{'}))$

We know that $vec(\hat{\Lambda}_2^{'}) = \left(\tilde{D}_2^{'}\hat{M}_1\tilde{D}_2\right)^{-1}\tilde{D}_2^{'}\hat{M}_1\hat{U}^{-\frac{1}{2}}1_n$. Therefore, by assumption 6,

$$\begin{aligned}
W(vec(\hat{\Lambda}_2^{'})) &= \left[1_n^{'}\hat{U}^{-\frac{1}{2}}\hat{M}_1(\sqrt{T}\tilde{D}_2)\left(\tilde{D}_2^{'}\hat{M}_1\tilde{D}_2\right)^{-1}\right](\hat{\Sigma}_{22})^{-1}\left[\left(\tilde{D}_2^{'}\hat{M}_1\tilde{D}_2\right)^{-1}(\sqrt{T}\tilde{D}_2^{'})\hat{M}_1\hat{U}^{-\frac{1}{2}}1_n\right] \\
&= \left[1_n^{'}\hat{U}^{-\frac{1}{2}}\hat{M}_1(\sqrt{T}\tilde{D}_2)\left((\sqrt{T}\tilde{D}_2^{'})\hat{M}_1(\sqrt{T}\tilde{D}_2)\right)^{-1}\right]T^2(\hat{\Sigma}_{22})^{-1}\left[\left((\sqrt{T}\tilde{D}_2^{'})\hat{M}_1(\sqrt{T}\tilde{D}_2)\right)^{-1}(\sqrt{T}\tilde{D}_2^{'})\hat{M}_1\hat{U}^{-\frac{1}{2}}1_n\right] \\
&\xrightarrow{p} \left[1_n^{'}U^{-\frac{1}{2}}M_1W_2\left(W_2^{'}M_1W_2^{'}\right)^{-1}\right]T^2\left(Avar(\sqrt{T}vec(\hat{\Lambda}_2^{'}))\right)^{-1}\left[\left(W_2^{'}M_1W_2^{'}\right)^{-1}W_2^{'}M_1U^{-\frac{1}{2}}1_n\right]\ .
\end{aligned} \tag{B.36}$$

Again, by assumption 6, we have:

$$\begin{aligned}
\left((\sqrt{T}\tilde{D}_2^{'})\hat{M}_1(\sqrt{T}\tilde{D}_2)\right)^{-1}(\sqrt{T}\tilde{D}_2^{'})\hat{M}_1\hat{U}^{-\frac{1}{2}}1_n &\xrightarrow{p} \left(W_2^{'}M_1W_2\right)^{-1}W_2^{'}M_1U^{-\frac{1}{2}}1_n \\
&\xrightarrow{p} \left(W_2^{'}M_1W_2\right)^{-1}vec(1_n^{'}U^{-\frac{1}{2}}M_1W_2I_p) \\
&\xrightarrow{p} \left(W_2^{'}M_1W_2\right)^{-1}\left(I_p \otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)vec(W_2) \\
&\xrightarrow{p} \left(W_2^{'}M_1W_2\right)^{-1}\left(I_p \otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)w_2
\end{aligned}$$

Therefore,

$$
\begin{aligned}
\Sigma_{22} &= Avar(\sqrt{T}vec(\hat{\Lambda}_2^{'})) \\
&= Avar\left[T\left((\sqrt{T}\tilde{D}_2^{'})\hat{M}_1(\sqrt{T}\tilde{D}_2)\right)^{-1}(\sqrt{T}\tilde{D}_2^{'})\hat{M}_1\hat{U}^{-\frac{1}{2}}1_n\right] \\
&= T^2\left[\left(W_2^{'}M_1W_2\right)^{-1}\left(I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)\right]I_{np}\left[\left(I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)^{'}\left(W_2^{'}M_1W_2\right)^{-1}\right] \\
&= T^2\left(W_2^{'}M_1W_2\right)^{-1}\left[\left(I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)\left(I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)^{'}\right]\left(W_2^{'}M_1W_2\right)^{-1}
\end{aligned}
\tag{B.37}
$$

By replacing in (B.36), we have:

$$
W(vec(\hat{\Lambda}_2^{'})) \xrightarrow{p} 1_n^{'}\hat{U}^{-\frac{1}{2}}M_1W_2\left(\left(I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)\left(I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)^{'}\right)^{-1}W_2^{'}M_1U^{-\frac{1}{2}}1_n \tag{B.38}
$$

Note that $W_2^{'}M_1U^{-\frac{1}{2}}1_n = \left(I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)vec(W_2) = \left(I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)w_2$. Therefore, $var(W_2^{'}M_1U^{-\frac{1}{2}}1_n) = \left(I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)I_{np}\left(I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)^{'} = \left(I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)\left(I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1\right)^{'}$. Since w_2 follows a normal distribution, we have from equation (B.38);

$$
W(vec(\hat{\Lambda}_2^{'})) \xrightarrow{p} \left((I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1)w_2\right)^{'}\left[var\left((I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1)w_2\right)\right]^{-1}\left((I_p\otimes 1_n^{'}U^{-\frac{1}{2}}M_1)w_2\right)\sim\chi^2(p)
$$

Chapter 3

Linear asset pricing models and nonlinear risk-adjustments*

3.1 Introduction

Stochastic discount factor models (hereafter SDF models) are a general representation of asset pricing models. They provide an appealing interpretation of the valuation mechanisms, as they link asset prices to the expected discounted value of their future payoffs. Once the SDF is known, agents can deduce asset prices from their future payoffs. Therefore, having a good estimation of SDFs is a key aspect of asset pricing.

One of the most common ways to approximate SDFs is to use linear factor models. The first attempt was the CAPM, with SDFs proxied as linear functions of the portfolio of aggregate wealth (Sharpe, 1964; Lintner, 1965). Since then, several linear factor models have been proposed to improve SDFs approximations.[1] However, the extreme multiplicity of factors proposed in the existing literature proves that it is very difficult to find an overall best linear factor model for that purpose.[2] Therefore, the question about the conditions under which a linear approximation is better than another is of interest.

Besides linear approximations, there are also nonlinear approximations encompassing

*I am grateful to René Garcia for his invaluable guidance.

[1] As few examples of later SDFs approximations through linear factor models, see: Fama & French (1993); Jagannathan & Wang (1996); Carhart (1997); Fama & French (2015)

[2] See Harvey et al. (2016) for precision on the "factors zoo" in asset pricing literature

specifications where SDFs are polynomial functions of factors (see Chapman, 1997; Harvey & Siddique, 2000), and specifications where any particular form is given to SDFs, then estimated through semi-parametric or non-parametric methods (see Dittmar, 2002; Wang, 2003; Almeida & Garcia, 2017). Nonlinear approximations have the advantage of being more informative (as they are less restrictive) and are especially required when returns are non-normal. Moreover, nonlinear SDFs are suitable for pricing higher-order systematic comoments (Almeida & Garcia, 2017).

Many linkages have been established between linear factor models in Arbitrage Pricing Theory (APT) and higher-order moments, as additional factors from CAPM can capture some of these moments. CAPM assumes that investors only care about means and variances, since systematic risks in that model are measured as contributions to the variance of the market returns. But we know from Scott & Horvath (1980) that preferences for higher-order moments also matter, particularly when the distribution of the expected utility is not fully determined by means and variances, when the returns have asymmetric distributions, or in the presence of tail risks. Harvey & Siddique (2000) explain in this vein how portfolios' skewness influence asset returns, while Chung, Johnson & Schill (2006) (hereafter CJS) show that the size and the value factors (*SMB* and *HML*) from the Fama-French's three-factor model become insignificant when systematic comoments of orders 3-10 are added in the model

This paper aims to generalize the idea of CJS by analyzing the influence of higher-order moments on the pricing abilities of linear factors models. The idea is to explain, according to investors' preferences for higher-order moments, how evolves the relevance of linear factor models in APT. We assume that there are some investors whose behaviors cannot be fully characterized with means and variances and therefore, analyze how their pricing strategies evolve according to their preferences for higher-order moments. Unlike CJS, we do not add systematic higher-order comoments as factors, but rather measure the explanatory power of linear factor models with respect to the optimal SDFs induced by various levels of preference for higher order moments. In general, analysis are performed on a limited number of higher-order moments, the most popular being the skewness and the kurtosis.[3] We do not adopt such a method in this paper since our analysis is more global: all higher-order moments are

[3] The effects of these two moments on asset pricing models have been widely analyzed in the literature. For an exhaustive presentation of skewness and kurtosis implications for asset pricing models, see Kraus & Litzenberger (1976); Harvey & Siddique (2000); Dittmar (2002).

considered at the same time.

More specifically, we analyze the pricing abilities of some popular linear factor models (benchmarks) according to investors' preferences for higher-order moments. As each investor has particular preferences, the pricing strategies are not the same among investors. Therefore, the pricing abilities of benchmarks vary according to investors' characteristics. In order to explain these differences, we measure the pricing errors between linear factor models and the optimal SDF estimated by the discrepancy minimization method from Almeida & Garcia (2017). Under a no-arbitrage condition, this optimal SDF can price all higher-order related exposures to risk. Therefore, we use it as the "*reference SDF*". Four linear factor models are used as benchmarks: (i) the CAPM; (ii) the three-factor model of Fama & French (1993), FF3; (iii) the four-factor model of Carhart (1997), CARH; and (iv) the five-factor model of Fama & French (2015), FF5. The analysis covers 642 months from July 1963 to December 2016.

The results show that: (i) for investors with moderate preferences for higher-order moments (in absolute value), the benchmarks can be ranked as follows: CAPM $\prec$ FF3 $\prec$ CARH $\sim$ FF5;[4] (ii) for investors with important preferences for higher-order moments, the ranking is CAPM $\prec$ FF3 $\prec$ FF5 $\prec$ CARH; and (iii) for investors with extreme preferences for higher-order moments, the ranking is CAPM $\prec$ FF5 $\prec$ FF3 $\prec$ CARH. A dynamic analysis confirms that since the year 2000, FF5 and CARH are the two best models for investors with moderate preferences for higher-order moments. For investors for whom these preferences are important or extreme, CARH is the best model.

We also analyze some linear factor models with non-traded factors. The results confirm the conclusions of Pondi (2018), as we observe that the human capital model analyzed in that paper and inspired by Lettau & Ludvigson (2001) and Gospodinov et al. (2014) is relevant for linear pricing. However, we show that this relevance is effective only for investors with null or very low preferences for higher-order moments.

This paper is linked to the literature of analyzing non-normal asset returns. We use the same non-linear optimization method as Almeida, Ardison & Garcia (2018), but unlike them, we do not analyze or propose a performance measure. Also, we generalize the idea of Chung et al. (2006), as we do not propose a particular specification with higher-order moments.

[4] Read the sign $\prec$ as "*have a lower pricing ability than*", and the sign $\sim$ as "*have a similar pricing ability than*"

However, the goal of this paper is different, since we do not just want to see if some higher-order moments can explain anomalies. The objective here is to measure how preferences for higher-order moments can influence the pricing abilities of linear factor models. This paper is also related to papers analyzing the usefulness of higher-order moments for explaining returns on markets (see Harvey & Siddique, 2000; Ando & Hodoshima, 2006; Smith, 2007), and to papers comparing the pricing abilities from different asset pricing models (see Barillas & Shanken, 2018). However, we limit our analysis to the Fama-French factor models.

Finally, this paper is related to those analyzing the links between optimal portfolios and higher-order moments (see Jondeau & Rockinger, 2006). However, our goal is not to propose an optimal portfolio. Optimal portfolio problems are only used here, as dual problems, in order to overcome dimensionality issues from the initial problems.

The rest of the paper is organized as follows: section 2 presents the links between optimal SDFs and preferences for higher-order moments. Section 3 presents the model used to compare linear factor models. Section 4 presents the analysis and the discussions. Section 5 concludes.

3.2 Nonlinear risk-adjustments: optimal SDF and higher-order moments

Conditions for an optimal SDF are presented in this section. We emphasize limits from conventional SDF estimation methods and present a recent method which aims to correct some of these limits.

3.2.1 Conditions for an optimal SDF

Let us consider an investor facing an intertemporal consumption and portfolio choice problem. Optimal decisions are summarized by the Euler equation

$$E(m_t R_t - 1_n) = 0_n \ , \tag{3.1}$$

where m_t is the SDF at time t, R_t the gross return on n assets at t, 1_n the n-vector of 1 and 0_n the n-vector of 0. We assume that there are T periods, so that $t \in \{1, \ldots, T\}$.

To define his dynamic strategy, the investor must evaluate the SDF at every period t.

Each period represents a state of nature, where decisions should be updated. The performance of a strategy can then be assessed with performance measures, summarized by the following equation (see Chen & Knez, 1996):

$$\alpha(R) = E(m_t R_t) - 1_n \ . \tag{3.2}$$

We assume that the market is incomplete, as the number of time periods (states of nature) is likely bigger than the number of priced assets in the market. The conditions that guarantee consistency of the performance measures presented above are derived under the following assumptions:[5]

Assumption 1. *If two portfolios have the same payoffs in every state of nature, then they should have the same price.*

Assumption 2. *If the payoff of a portfolio is (almost surely) non-negative and certainly strictly positive in some states of nature, then the price of that portfolio should be strictly positive.*

Assumption 1 is the law of one price as formulated by Garcia (2018). This assumption is a sufficient condition for the existence of an admissible SDF. Assumption 2 is the no-arbitrage condition as presented by Cochrane (2001). This assumption is a necessary and a sufficient condition for the existence of a strictly positive admissible SDF.

Because the market is incomplete, there may be a large set of admissible SDFs. Therefore, it could be hard to find the optimal one, which must be strictly positive in every state of nature. In general, some restrictions are made to overcome that difficulty. (i) The most common restriction is to write SDFs as functions of variables. Such a strategy is clearly limited, as all specifications cannot be considered. With that method, it is difficult to clearly define which factors should be considered and which specification is the best. This is why factor models have led to what is called today "*the factor zoo*" (see Harvey et al., 2016; Feng et al., 2017). (ii) Another restriction strategy is to estimate SDFs by using the mean-variance frontier of asset returns, like Hansen & Jagannathan (1991). That strategy does

[5] For a deeper presentation on consistency of performance measures in asset pricing, see Almeida & Garcia (2017)

not consider higher-order related exposures to risk.[6] SDFs estimated through that method fail to satisfy the strictly positivity constraint, as it may produce negative SDFs in some state of nature. Even the constrained Hansen-Jagannathan does not always achieve strictly positivity constraint, particularly when the priced returns are non-normal or when there are tail risks (see Almeida et al., 2018). (iii) Finally, a utility-based estimation can be used. As presented by Cochrane (2001), the value of a SDF at t can be interpreted as the marginal rate of substitution between periods t and $t+1$.[7] Therefore, once the investor's utility is known, we can easily estimate the SDF. However, utility-based estimations are not always accurate since they require equilibrium models with representative agents.

Following the issues emphasized above, we can say that optimal SDFs should be estimated without any restriction either on investor's utility, or on functional specifications applied on SDFs. As a solution, Almeida & Garcia (2017) recently proposed a method which optimizes on all the set of strictly positive admissible SDFs and relaxes the restrictions listed above. This method is utility-free and consider all higher-order related exposures to risk.

3.2.2 Finding the optimal SDF

Almeida & Garcia (2017) consider that the optimal SDF minimizes a convex discrepancy function ϕ, defined by:

$$\phi(m,\gamma) \equiv \phi_\gamma(m) = \frac{m^{\gamma+1}-1}{\gamma(\gamma+1)} \qquad (\gamma \in \mathbb{R}) \ . \tag{3.3}$$

The nature of the discrepancy function depends on the value of γ. As an example, for the popular Hansen & Jagannathan (1991)'s method (hereafter HJ method), the optimal SDF is obtained though the minimization of a quadratic function of pricing errors (which matches with the case $\gamma = 1$).[8] For a given value of γ, the optimal SDF m^*_γ is given as a solution of

[6]What we call "*higher-order related exposures to risk*" is all information from higher-order comoments with respect to asset returns

[7]More precisely, m_t is the rate at which the investor is willing to substitute its consumption at t for consumption at $t+1$

[8]For an exhaustive presentation of popular particular cases associated with the discrepancy functions presented here, see Almeida & Garcia (2017).

the following problem:

$$m_\gamma^* \equiv (m_{\gamma,1}^*, \dots m_{\gamma,T}^*) = \underset{m}{argmin} \;\; E(\phi_\gamma(m))$$
$$s.t. \left| \begin{array}{ll} E\left(m(R_t - 1_n)\right) = 0_n & (i) \\ E(m) = 1 & (ii) \\ m \gg 0 & (iii) \end{array} \right. \tag{3.4}$$

As precision, (iii) is a consequence of the arbitrage-free assumption, (ii) is taken without loss of generality, and (i) is a consequence of (3.1).

As Problem (3.4) is solved on a T-dimensional space, we can use its dual counterpart to reduce dimensionality issues and solve it on a n-dimensional space,[9] following Borwein & Lewis (1991). The dual problem is given as follow:

$$\lambda_\gamma^* = \underset{\alpha\in\mathbb{R}}{arg} \; \underset{\lambda}{sup} \; \alpha - E\left(\phi_\gamma^{*,+}(\alpha + \lambda'(R_t - 1_n))\right) \; . \tag{3.5}$$

$\phi_\gamma^{*,+}$ is a convex conjugate of ϕ_γ and is defined by $\phi_\gamma^{*,+}(z) = \underset{w>0}{sup} \; zw - \phi_\gamma(w)$. By denoting λ_γ^* the solution of this dual problem for a given value γ, the optimal SDF at t is given by:

$$m_{\gamma,t}^* = \begin{cases} \left(1 + \gamma\lambda_\gamma^{*\,\prime}(R_t - 1_n)\right)^{\frac{1}{\gamma}} \mathbb{1}_{\{1+\gamma\lambda_\gamma'(R_t - 1_n)>0\}}(\lambda_\gamma) & \text{if } \gamma > 0 \\ \left(1 + \gamma\lambda_\gamma^{*\,\prime}(R_t - 1_n)\right)^{\frac{1}{\gamma}} & \text{if } \gamma < 0 \\ exp\left(\lambda_0^{*\,\prime}(R_t - 1_n)\right) & \text{if } \gamma = 0 \end{cases} \tag{3.6}$$

Almeida & Garcia (2017) show that the dual problem from (3.4) can be interpreted as an optimal portfolio problem under the following HARA utility function:

$$u(w) = \frac{-1}{\gamma + 1}(1 - \gamma w)^{\frac{\gamma+1}{\gamma}} \qquad (1 - \gamma w > 0) \; , \tag{3.7}$$

where $w_t = -\lambda_\gamma'(R_t - 1_n)$ is interpreted as "the agent's wealth" at t and γ is the same parameter as in (3.3). Since w_t also depends on γ, we write it without γ for simplicity. The condition $(1 - \gamma w) > 0$ comes from the SDF estimation, as $1 - \gamma w_t \equiv 1 + \gamma\lambda_\gamma'(R_t - 1_n) > 0$ (see Almeida et al., 2018, for more details about estimation procedures).

[9]The number of assets n is smaller than the number of periods T, as the market is incomplete.

By a Taylor expansion of this utility function around the expected optimal wealth, the weights given to any higher-order moments have the following form:

$$W(\gamma) = g(\gamma)\,(1 - \gamma E(w_t))^{h(\gamma)} \ , \tag{3.8}$$

with g and h two functions whose expressions depend on the orders of moments.[10] As the weights mainly depend on γ, we can say that γ is an indicator of the level of preference for higher-order moments. For each γ, the optimal SDF m^*_γ is consistent with preferences for all higher-order moments, under the proportions (or the weights) indicated by the value of γ.

It is important to emphasize that each value of γ corresponds to a utility function in the dual problem (3.7). Therefore, as the initial discrepancy function depends on γ, there is a different discrepancy function for each type of investors (according to their utility functions in the dual problem). Each γ characterizes a different attitude toward risk and therefore, a different pricing strategy. In this paper, we will analyze these differences according to the levels of preference for higher-order moments (which are also characterized, as seen above, by the values taken by γ).

Moreover, about the dual optimal portfolio problem, note that $\gamma = 1$ matches with a quadratic utility (so that the initial problem is equivalent to HJ), $\gamma = 0$ matches with an exponential utility and $\gamma = -1$ matches with a logarithmic utility.

3.3 The model

Linear factor models are widely used in asset pricing literature, as the idea of linking asset returns with observable or estimated variables is appealing. In general, the pricing abilities of these models are measured according to their linear exposures to risk, without any consideration for heterogeneity among investors.[11] In this paper, we want to assess the pricing ability of linear factor models according to investors' preferences. We characterize heterogeneity with the levels of preference for higher-order moments among investors.

As emphasized above, the Almeida & Garcia (2017)'s method provides an optimal SDF which is utility-free and which consider all higher-order related exposures to risk. That method

[10]We will talk more deeply about the expressions of g and h later.

[11]Linear factor models in asset pricing are mainly equilibrium models with homogeneity of beliefs among investors.

also assures strictly positivity in every states of nature. Therefore, we are going to consider the optimal SDF from Almeida & Garcia (2017) as the "*reference SDF*" and measure the pricing errors between that reference and linear factor models. As we already know the limits of linear factor models (particularly for pricing nonlinear exposures to risk), the purpose of our analysis is to compare the most popular of these models, and to evaluate how their pricing abilities evolve according to investors' preferences for higher-order moments (remember that for each level of preference for higher-order moments γ, there is an optimal SDF m_γ^*).

3.3.1 Heterogeneity among investors: values of γ and preferences for higher-order moments

As seen above in equation (3.7), the dual problem used to estimate the optimal SDF can be interpreted as an optimal portfolio problem under a HARA utility function.[12] By a Taylor expansion of that utility function around the expected optimal wealth, we have the following proposition.

Proposition 1. *The weight associated with a moment of order $p \geq 2$, is given as follow:*

$$W_p(\gamma) = \frac{1}{p!}(-1)^{p+1} \prod_{j=0}^{p-2}(1 - j\gamma) \cdot (1 - \gamma E(w_t))^{\frac{1}{\gamma} - (p-1)} \ . \tag{3.9}$$

Since $1 - \gamma E(w_t) > 0$, the sign of $W_p(\gamma)$ only depends on the value of $(-1)^{p+1} \prod_{j=1}^{p-2}(1 - j\gamma)$. Therefore, the sign of $W_p(\gamma)$ only depends on γ. Note that the weight associated with the moment of order 1 is $(1 - \gamma E(w_t))^{\frac{1}{\gamma}}$. Now let us consider the following assumption.

Assumption 3.
(i) Each investor is risk-averse and has a positive marginal utility. His utility function is such that $u^{'} > 0$ and $u^{''} < 0$.
(ii) There is a strict consistency in preference direction for all investors on the market, and for all moments of order p. So, for all $p \in \mathbb{N}^*$, we have $u^{(p)} > 0$, or $u^{(p)} < 0$, or $u^{(p)} = 0$

Assumptions 3(i)-(ii) are assumptions A1, A2 and A3 from Scott & Horvath (1980). Under these assumptions, they show that odd moments are weighted positively while even

[12] For a value of γ, that utility function is given by $u(w) = \frac{-1}{\gamma+1}(1 - \gamma w)^{\frac{\gamma+1}{\gamma}}$, with $(1 - \gamma w) > 0$.

moments are weighted negatively (we will call this result "*conditions on the signs of weights*"). Therefore, we should have:

$$
\begin{array}{llll}
(order\ 3) & (1-\gamma) > 0 & (order\ 4) & -(1-\gamma)(1-2\gamma) < 0 \\
(order\ 5) & (1-\gamma)(1-2\gamma)(1-3\gamma) > 0 & (order\ 6) & -(1-\gamma)(1-2\gamma)(1-3\gamma)(1-4\gamma) < 0 \\
\ldots & & \ldots &
\end{array}
$$

Proposition 2. *The weight associated with a moment of order $p \geq 3$ is consistent with the "conditions on the signs of weights" if and only if $\gamma < \frac{1}{p-2}$. So, if we want to satisfy these conditions for all higher-order moments, we must consider γ such that $\gamma \leq 0$.*

Proposition 2 does not concern the weights associated with the moments or orders 1 and 2. These weights satisfy the "*conditions on the signs of weights*" without any restriction on γ since they are respectively given by $(1-\gamma E(w_t))^{\frac{1}{\gamma}}$ which is positive, and $-(1-\gamma E(w_t))^{\frac{1}{\gamma}-1}$ which is negative.[13]

When $\gamma \leq 0$, all the weights associated with higher-order moments are non-null. Moreover, the further the value of γ is below 0, the more preferences for odd moments increase and the more preferences for even moments decrease. For our analysis, we will assume negative values for γ, and consider two additional particular cases: $\gamma = 1$ and $\gamma = 0.5$. For precision, $\gamma = 1$ corresponds to the case where all preferences for higher-order moments are null (HJ model), while $\gamma = 0.5$ corresponds to the case where the preference for skewness is non-null, and all the preferences for a moment of an order above 3 are null.[14] Therefore, we are going to perform analysis for the following values of γ: -3, -2, -1, 0, 0.5 and 1 (Almeida et al., 2018, use values from -3.5 to 1). Moreover, we will assume that the preference for higher-order moments is (i) *moderate* when $\gamma = 1$, $\gamma = 0.5$ and $\gamma = 0$, (ii) *important* when $\gamma = -1$ and $\gamma = -2$, and (iii) *extreme* when $\gamma = -3$.

[13] Remember that the utility in the dual optimal portfolio problem is given under the condition $(1-\gamma w_t) > 0$ (see (3.7)).

[14] Following Scott & Horvath (1980), the case $\gamma = 0.5$ corresponds to a situation where an investor is willing to choose an asset with the greater skewness, even if that asset has a lower expected value in comparison to another.

3.3.2 Estimation and analysis procedures

In this subsection, we present the data and the linear factor models used as benchmarks in this paper. We also present how the relevance of each benchmark will be assessed, and how the pricing abilities of these benchmarks will be compared.

Data and benchmark models

Four linear factor models are used as benchmarks: (i) the CAPM; (ii) the three-factor model of Fama & French (1993), FF3; (iii) the four-factor model of Carhart (1997), CARH; and (iv) the five-factor model of Fama & French (2015), FF5. These linear factor models are among the most popular in asset pricing and have the particularity of being nested in each other. This will make the comparison among them much easier and will help us to see whether the pricing ability of a factor model, after an addition of new factors, depends on the preferences of investors (or precisely, on their preferences for higher-order moments).

As returns, we use 25 Fama-French monthly portfolios sorted on size and book-to-market, from July 1963 to December 2016. These data are from Kenneth French's website. Since the goal of this paper is not to explain returns, we only use these portfolios instead of adding additional portfolios, like Gospodinov et al. (2014). The relevance of SDFs is not analyzed here with respect to returns, but rather with respect to an estimated optimal nonlinear SDF. Therefore, even if the selected portfolios exhibit a factor structure, they are unlikely to be a source of data snooping.[15]

Benchmarks relevance: inference on the benchmarks parameters

Let us denote $y = f'\theta$, a linear factor model with θ a k-dimensional parameter and f a vector of factors. We consider a given value γ and the subsequent optimal SDF m^*_γ, estimated by the Almaida-Garcia method. To estimate the parameter θ_γ in our selected benchmarks, we follow Hansen & Jagannathan (1997) and solve the following optimization problem:

$$\theta^*_\gamma = \underset{\theta}{argmin} \|f^{'}\theta - m^*_\gamma\| \ . \tag{3.10}$$

[15] For robustness checks, we will apply the correction method proposed by Lewellen et al. (2010), and will add industry portfolios to the initial 25 portfolios

$\delta_\gamma = \underset{\theta}{min}\|f^{'}\theta - m^*_\gamma\|$ is a distance that measures the linear discrepancy between the linear factor model and the optimal SDF m^*_γ. Therefore, the optimal parameter θ^*_γ can be obtained through an Ordinary Least Squares (OLS) regression of m^*_γ on the factors f. We are going to use that strategy.[16]

Note that the relevance of benchmarks as presented in this paper, implies a statistical analysis of the significance of the factors from each linear factor model (i.e. an analysis of the significance of θ_γ, for each value of γ). Moreover, it is important to emphasize that we could have used a more general method like Weighted Least Squares (WLS) to estimate θ^*_γ. However, this strategy is not suitable in the present case, because we have to use a unique model-free weighting matrix for comparison purposes (as for the HJ distance).

Benchmarks comparisons

Our comparisons rely on a measurement of the goodness-of-fits between the benchmarks and the optimal SDF. For every value of γ, the idea is to measure how each linear factor model (benchmark) explains the optimal SDF m^*_γ, and then to compare the results among benchmarks. Benchmarks comparisons come after estimations of the benchmarks parameters. For that purpose, we will use Akaike Information Criteria (AIC)[17] and F-tests. We present below, how the F-statistics will be estimated and interpreted.

Let us assume that we have two linear models: a simple model (model 1), nested in a much more complex model (model 2). Let us also assume that the two specifications are given as follows:

$$(model\ 1) \quad y_t = \sum_{l=1}^{K_1} \alpha_l f_t \qquad t = 1, \dots, T \tag{3.11}$$

$$(model\ 2) \quad y_t = \sum_{l=1}^{K_1} \alpha_l f_t + \sum_{l=1}^{K_2} \beta_l g_t \qquad t = 1, \dots, T \tag{3.12}$$

The idea of the F-test is to measure whether the reduction of the residual sum of squares induced by moving from model 1 to model 2 (a gain), is canceled by the relative reduction of

[16]For inference purposes, heteroskedasticity and autocorrelation consistent standard errors will be estimated following Newey & West (1987)

[17]Akaike Information Criteria will be estimated according to the formula $AIC = 2K + TLog(RSS/T)$, where K is the number of parameters in the model, T the size of the sample on which the model is applied and RSS the residual sum of squares from the model.

the degree of freedom (a loss). For that purpose, we test the null $H_0 : \beta_1 = \cdots = \beta_{K_2} = 0$. By denoting RSS the residual sum of squares, the F-stat follows the Fisher-Snedecor distribution $F(K_2, T - K_1 - K_2)$, and is given by:

$$F = \frac{(RSS_1 - RSS_2)/K_2}{RSS_2/(T - K_1 - K_2)} \tag{3.13}$$

If the null is not rejected, then there is no evidence that the complex model (model 2) is better than the simple model (model 1). On the other hand, if the null is rejected, the probability of having a better fit with model 2 is higher than the size of the test; therefore, model 2 is selected.

For our pairwise comparisons, we will use the AIC as a complementary criterion. If the null is rejected while model 2 has a lower AIC than model 1, then we will conclude that model 2 is better. In all cases, we will only interpret the results for which there is coherence between the results from AIC and those from F-tests.

For the ranking of more than 2 models, we will still just consider consistent results between AIC and F-tests, with the exception that for this special case, we will assume that two models have similar explanatory power if the two criteria (AIC and F-tests) permute them in their respective rankings, while keeping the same structure among the rankings. As an illustration, suppose that we try to compare four models M_1, M_2, M_3 and M_4. Suppose, moreover, that AIC and F-tests respectively give the following rankings: $M_1 \prec M_2 \prec M_3 \prec M_4$, and $M_1 \prec M_3 \prec M_2 \prec M_4$. Then, we will conclude that models M_2 and M_3 have similar explanatory power and consider the following as the final ranking: $M_1 \prec M_2 \sim M_3 \prec M_4$.

3.4 Analysis and discussions

The main point of this section is to analyze the relevance of benchmarks and compare their pricing abilities with respect to the optimal SDF m^*_γ, induced by a level of preference for higher-order moments γ. As presented above, consistency in preference direction is satisfy for any higher-order moment when $\gamma \leq 0$. However, we will also analyze the particular cases $\gamma = 0.5$ and $\gamma = 1$, which will be considered as for the case $\gamma = 0$, like cases where the preference for higher-order moments is moderate (remember that the further γ is below 0, the more the weights on higher-order moments increase in absolute value). Moreover, for simplicity, we will call "*p-value of a factor*", the p-value of the t-test on the parameter associated with that factor. Also, during the analysis, we will sometimes associate values of γ with their corresponding utility functions in the dual optimal portfolio problem (see (3.5) and (3.7)).

3.4.1 Significance of factors and preference for higher-order moments

Table 3.1 presents the p-values from the t-tests on the regression parameters, following the OLS regressions of the optimal SDF on benchmarks. It appears that overall, all the factors in benchmarks are relevant when preferences match with the classical HJ case ($\gamma = 1$). Moreover, all the p-values increase as γ decreases. Therefore, the more the preference for higher-order moments increases, the more some factors become insignificant. This is particularly the case for the size factor *SMB* in FF3, CARH and FF5, and for the value factor *HML*, the profitability factor *RMW* and the investment factor *CMA*, all in FF5.

Note that the size factor *SMB* becomes insignificant very quickly (from $\gamma = 0$). On the other hand, *HML*, *RMW* and *CMA* become insignificant in FF5 when the preference for higher-order moments becomes relatively high (from $\gamma = -2$ for *RMW*, and from $\gamma = -3$ for the two others). Furthermore, note that the market factor *Mkt-Rf* is the only factor which is always relevant, irrespective of the benchmark and the level of preference for higher-order moments.

As there are some particularities associated with the benchmarks, we now analyze specifically some of these benchmarks.

Significance of factors in FF3. In FF3, the market factor *Mkt-Rf* and the value factor *HML* are significant for all of the selected values of γ. Their p-values indicate that these two factors are relevant even when the weights on higher-order moments are very high. This is not the case for the size factor *SMB*, whose significance is relatively weak, even for quadratic preferences (the p-value of this factor is 0.096 when $\gamma = 1$).

Significance of factors in CARH. For CARH, observations are quite the same as for FF3, with the exception that here, the relevance of the size factor *SMB* is relatively more effective when preferences are quadratic or exponential (p-value at 0.074 and 0.076 respectively when $\gamma = 1$ and $\gamma = 0$). We still note that the market factor and the value factor are significant even when the preference for higher-order moments is very high. Furthermore, although the momentum factor *Mom* is significant for all the selected values of γ, its significance becomes weak when the preference for higher-order moments becomes extreme; its p-value is 0.081 when $\gamma = -3$.

Significance of factors in FF5. FF5 is very different from the two benchmarks analyzed above, since, in addition to the size factor, all the other factors - with the exception of the market - become insignificant when the preference for higher-order moments becomes high. Moreover, unlike the previous two benchmarks, the significance of the size factor is stronger for quadratic and exponential preferences (p-values at 0.005 and 0.010 respectively when $\gamma = 1$ and $\gamma = 0$). Note that even if the profitability factor *RMW* and the investment factor *CMA* are both insignificant for $\gamma = -2$ and $\gamma = -3$, these two factors are relevant for quadratic, exponential and logarithmic preferences.

Table 3.1: P-values from t-tests on the regression parameters

		values of γ					
		-3	-2	-1	0	0.5	1
CAPM	Int.	0	0	0	0	0	0
	Mkt-Rf	0.033	0.034	0.012	0	0	0
FF3	Int.	0	0	0	0	0	0
	Mkt-Rf	0.014	0.010	0.002	0	0	0
	SMB	0.297	0.292	0.251	0.114	0.093	0.096
	HML	0.001	0	0	0	0	0
CARH	Int.	0	0	0	0	0	0
	Mkt-Rf	0.011	0.009	0.002	0	0	0
	SMB	0.301	0.295	0.252	0.101	0.076	0.074
	HML	0	0	0	0	0	0
	Mom	0.081	0.078	0.058	0.004	0	0
FF5	Int.	0	0	0	0	0	0
	Mkt-Rf	0.009	0.008	0.001	0	0	0
	SMB	0.213	0.201	0.121	0.010	0.005	0.005
	HML	0.135	0.051	0.009	0	0	0
	RMW	0.218	0.147	0.039	0	0	0
	CMA	0.108	0.093	0.060	0.015	0.008	0.006

The table gives the p-values from the t-tests on the regression parameters, as the optimal SDF is regressed on each benchmark though the OLS method. The tests are performed with robust standard errors, estimated following Newey & West (1987)

3.4.2 Benchmarks comparisons

Table 3.2 presents AIC from regressions of the optimal SDF on benchmarks. We can see that when the preference for higher-order moments is moderate, adding an additional Fama-French factor to a prior benchmark leads to a better asset pricing model. For $\gamma = 1$, $\gamma = 0.5$ and $\gamma = 0$, the following ranking remains true among the benchmarks: CAPM $\prec$ FF3 $\prec$ CARH $\prec$ FF5 (read the sign $\prec$ as "*have a lower pricing ability than*"). On the other hand, when the preference for higher-order moments becomes important, there are some permutations in the previous ranking. To understand these permutations, let us analyze the results from Table 3.3.

Table 3.3 presents the results from the F-tests used to pairwise compare benchmarks. It appears that regardless of the level of preference for higher-order moments, FF3, CARH, and FF5 have better pricing abilities than CAPM. Moreover, we see that CARH is always better than FF3, and that FF3 is better than FF5 only when the preference for higher-order moments becomes extreme ($\gamma = -3$). Finally, we observe that FF5 is better than CARH only when the preference for higher-order moments is moderate. The pricing ability of CARH becomes better than that of FF5 when the preference for higher-order moments becomes important (CARH is better than FF5 for $\gamma = -1$, $\gamma = -2$ and $\gamma = -3$).

From the results presented above, it appears that adding new factors to a prior asset pricing model does not always improve the pricing ability of the model. In some circumstances, it is better to perform asset pricing with much simpler models. On some level, these results confirm the conclusions of Kozak, Nagel & Santosh (2017).[18] In our analysis, distinctions among asset pricing models are made according to the levels of agents' preferences for higher-order moments. Under that perspective, we see that FF3, CARH and FF5 are all improvements from CAPM. When the preference for higher-order moments is moderate we have the following ranking among benchmarks: CAPM $\prec$ FF3 $\prec$ CARH $\prec$ FF5. On the other hand, when the preference for higher-order moments becomes important, the ranking is CAPM $\prec$ FF3 $\prec$ FF5 $\prec$ CARH. Lastly, when the preference for higher-order moments becomes extreme, the ranking becomes CAPM $\prec$ FF5 $\prec$ FF3 $\prec$ CARH.

[18] Kozak et al. (2017) show that a relatively small number of estimated factors explain the SDF better than empirical 4-factor or 5-factor models

Table 3.2: Benchmarks comparisons with AIC

	values of γ					
	-3	-2	-1	0	0.5	1
CAPM	136.91	9.98	-176.44	-338.43	-375.07	-385.90
FF3	133.67	2.70	-194.21	-374.33	-417.10	-429.77
CARH	132.17	0.04	-199.88	-383.02	-426.19	-438.54
FF5	136.15	4.40	-195.65	-383.73	-429.38	-442.59

The table gives Akaike information criteria from regressions of the optimal SDF on benchmarks

Table 3.3: Benchmarks comparisons with F-tests

		values of γ					
		-3	-2	-1	0	0.5	1
CAPM	*CAPM vs FF3*	0	0	0	0	0	0
	CAPM vs CARH	0	0	0	0	0	0
	CAPM vs FF5	0	0	0	0	0	0
FF3	*FF3 vs CARH*	0.005	0.001	0	0	0	0
	FF3 vs FF5	0.176	0.072	0.002	0	0	0
CARH	*CARH vs FF5*	1	1	1	0.013	0	0

The table gives the p-values from the F-tests on nested models among benchmarks. A non-rejection of the null means that the simple specification is better than the complex.

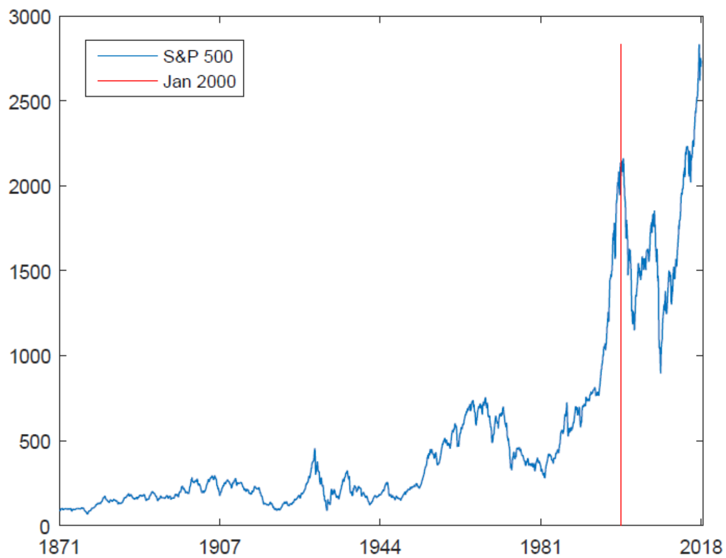
3.4.3 Time-consistency of results: dynamic analysis

In this subsection, we aim to determine how time-consistent the previous results are. For that purpose, we perform the previous analysis under two different periods of time: before 2000 and since 2000 (results in tables 3.4 and 3.5). We chose 2000 as threshold following the results from Pondi (2018) according to which, major changes have been observed in the pricing abilities of benchmarks since 2000. These changes can partly be explained by the intensification of activities on the markets since 2000. Figure 3.1 shows in this vein, with constant May 2018 dollars, the evolution of the inflation adjusted $S\&P$ 500 index from January 1871 to July 2018.[19] We see that since 2000, the index has reached levels never reached before, and the amplitudes of its cycles are much higher.

For the period before 2000, we observe that when the preference for higher-order moments is moderate ($\gamma = 1$, $\gamma = 0.5$, $\gamma = 0$), the following ranking holds among benchmarks: CAPM $\prec$ CARH $\prec$ FF3 $\prec$ FF5. On the other hand, when the preference for higher-order moments is important or extreme, we have the following ranking: CAPM $\prec$ FF5 $\prec$ CARH $\prec$ FF3. So,

[19]See Shiller (2000) for a presentation on the construction of historic $S\&P$ 500 prices

Figure 3.1: Evolution of the monthly $S\&P$ 500 index



Inflation adjusted $S\&P$ 500 index, with constant May 2008 dollars. The red line indicates January 2000. Data source: Shiller (2000), and multpl economic data provider (*www.multpl.com*)

the rankings we have for the period 1963-1999 are different from the overall rankings obtained above. If FF5 is still the best benchmark when the preference for higher-order moments is moderate, FF3 appears to be the best benchmark when the preference for higher-order moments is important or extreme (instead of CARH in the overall analysis).

For the period since 2000, we observe that when the preference for higher-order moments is moderate, we have CAPM $\prec$ FF3 $\prec$ CARH $\prec$ FF5. This result is the same as the overall result. So since 2000, FF5 is the best benchmark for pricing assets on markets where investors have moderate preferences for higher-order moments. When these preferences are important or extreme (particularly for $\gamma = -2$ and $\gamma = -3$), we have contradictory results between AIC ranking and F-tests. Therefore, we do not interpret results for these levels of preference.[20]

[20] One reason for contradictory results between AIC and F-tests may be the relatively small size of the sample "*since 2000*" (204 months, while the sample "*before 2000*" has 438 months)

Table 3.4: Benchmarks comparisons with AIC (dynamic analysis)

	Before 2000					
	values of γ					
	-3	-2	-1	0	0.5	1
CAPM	43.90	-14.34	-107.85	-208.28	-235.06	-243.02
FF3	37.91	-23.80	-126.06	-243.85	-277.59	-287.83
CARH	39.56	-22.51	-125.12	-242.72	-276.39	-286.70
FF5	40.75	-21.29	-124.49	-246.07	-281.96	-292.59
	Since 2000					
	values of γ					
	-3	-2	-1	0	0.5	1
CAPM	12.94	-8.93	-36.77	-66.25	-75.58	-77.98
FF3	14.37	-8.13	-37.01	-68.14	-78.16	-80.74
CARH	14.51	-8.08	-37.91	-70.91	-81.64	-83.95
FF5	15.23	-7.73	-38.02	-72.31	-83.85	-87.10

The table gives Akaike information criteria from regressions of the optimal SDF on benchmarks

Table 3.5: Benchmarks comparisons with F-tests (dynamic analysis)

		Before 2000					
		values of γ					
		-3	-2	-1	0	0.5	1
	CAPM vs FF3	0	0	0	0	0	0
CAPM	*CAPM vs CARH*	0	0	0	0	0	0
	CAPM vs FF5	0	0	0	0	0	0
FF3	*FF3 vs CARH*	0.374	0.204	0.122	0.160	0.178	0.161
	FF3 vs FF5	0.268	0.185	0.063	0.001	0	0
CARH	*CARH vs FF5*	0.175	0.184	0.077	0	0	0
		Since 2000					
		values of γ					
		-3	-2	-1	0	0.5	1
	CAPM vs FF3	0.055	0.027	0.008	0.001	0.001	0
CAPM	*CAPM vs CARH*	0.019	0.009	0.001	0	0	0
	CAPM vs FF5	0.012	0.004	0	0	0	0
FF3	*FF3 vs CARH*	0.041	0.036	0.011	0.001	0	0.001
	FF3 vs FF5	0.030	0.018	0.004	0	0	0
CARH	*CARH vs FF5*	0.092	0.056	0.030	0.006	0.002	0.001

The table gives the p-values from the F-tests on nested models among benchmarks. A non-rejection of the null means that the simple specification is better than the complex.

3.4.4 Robustness checks

In this subsection, we check to see if our qualitative results are coherent regardless of the returns on which the model is applied. For that purpose, we add 10 industry portfolios to the initial 25 Fama-French portfolios and perform the same analysis as previously.[21] We use robustness checks to define the final rankings. Among the ranking presented in the previous sub-sections, only the robust ones will be kept.

Overall analysis

The results in tables 3.6 and 3.7 give the following rankings: when $\gamma = 1$, $\gamma = 0.5$ and $\gamma = 0$, we have CAPM $\prec$ FF3 $\prec$ FF5 $\prec$ CARH. When $\gamma = -1$, we have CAPM $\prec$ FF3 $\sim$ FF5 $\prec$ CARH. When $\gamma = -2$ and $\gamma = -3$, we have CAPM $\prec$ FF5 $\prec$ FF3 $\prec$ CARH. Therefore, the results are consistent with those obtained in section 3.4.2, particularly for the cases where the preference for higher-order moments is either important or extreme.

Note that the ranking obtained in section 3.4.2, for the case of moderate preferences for higher-order moments, is not fully confirmed here: FF5 is not better than CARH after the addition of the 10 industry portfolios. However, as FF5 remains the second-best model for that level of preference, we can still conclude that FF5 and CARH are the two best models when the preference for higher-order moments is moderate. Therefore, for a moderate level of preference for higher-order moments, we adopt the following final ranking: CAPM $\prec$ FF3 $\prec$ CARH $\sim$ FF5.

In summary, the final rankings over the entire sample are given as follows: (i) CAPM $\prec$ FF3 $\prec$ CARH $\sim$ FF5 for investors with moderate preferences for higher-order moments; (ii) CAPM $\prec$ FF3 $\prec$ FF5 $\prec$ CARH for investors with important preferences for higher-order moments; (iii) CAPM $\prec$ FF5 $\prec$ FF3 $\prec$ CARH for investors with extreme preferences for higher-order moments.

[21] We just add 10 industry portfolios here instead of 17 as in Pondi (2017) and Pondi (2018), because the solving program of problem (3.4) does not converge when we have both a large γ (in absolute value) and a large number of assets. With 35 portfolios, the program converges for all the selected values of γ

Table 3.6: Benchmarks comparisons with AIC (with additional returns)

	values of γ					
	-3	-2	-1	0	0.5	1
CAPM	198.86	125.69	-46.49	-234.88	-277.71	-289.43
FF3	197.24	122.42	-56.02	-258.20	-305.79	-318.80
CARH	195.23	117.97	-70.36	-286.60	-332.22	-341.06
FF5	200.62	125.50	-54.76	-263.03	-313.97	-329.20

The table gives Akaike information criteria from regressions of the optimal SDF on benchmarks.
10 industry portfolios are added to the initial 25 Fama-French portfolios used for table 3.2.

Table 3.7: Benchmarks comparisons with F-tests (with additional returns)

		values of γ					
		-3	-2	-1	0	0.5	1
CAPM	*CAPM vs FF3*	0.002	0	0	0	0	0
	CAPM vs CARH	0	0	0	0	0	0
	CAPM vs FF5	0.006	0.001	0	0	0	0
FF3	*FF3 vs CARH*	0.002	0	0	0	0	0
	FF3 vs FF5	0.491	0.353	0.044	0	0	0
CARH	*CARH vs FF5*	1	1	1	1	1	1

The table gives the p-values from the F-tests on nested models among benchmarks. A non-rejection of the null means that the simple specification is better than the complex.
10 industry portfolios are added to the initial 25 Fama-French portfolios used for table 3.3.

Dynamic analysis

We perform again the dynamic analysis (before 2000 and since 2000), after the addition of 10 industry portfolios. The results are presented in tables 3.8 and 3.9.

For the period 1963-1999, we observe that when the preference for higher-order moments is important or extreme, the ranking is CAPM $\prec$ FF5 $\prec$ CARH $\prec$ FF3. This result is the same as the one obtained in subsection 3.4.3. When the preference for higher-order moments is moderate, the ranking is CAPM $\prec$ FF5 $\prec$ FF3 $\prec$ CARH; this result, on the other hand, does not confirm the one obtained above.

Table 3.8: Benchmarks comparisons with AIC (dynamic analysis with additional returns)

	Before 2000					
	values of γ					
	-3	-2	-1	0	0.5	1
CAPM	105.89	55.66	-35.23	-130.60	-158.69	-166.62
FF3	102.18	49.67	-47.36	-153.54	-186.21	-195.55
CARH	103.31	50.56	-47.08	-154.30	-186.91	-196.09
FF5	105.58	53.11	-43.88	-150.42	-183.65	-193.65
	Since 2000					
	values of γ					
	-3	-2	-1	0	0.5	1
CAPM	45.01	33.21	7.84	-27.11	-38.43	-41.33
FF3	47.28	35.19	9.10	-27.20	-39.08	-42.11
CARH	44.62	32.05	4.63	-35.07	-48.25	-51.32
FF5	45.58	33.19	5.68	-33.43	-46.37	-49.22

The table gives Akaike information criteria from the regressions of the optimal SDF on benchmarks
10 industry portfolios are added to the initial 25 Fama-French portfolios used for table 3.4.

Table 3.9: Benchmarks comparisons with F-tests (dynamic analysis with additional returns)

		Before 2000					
		values of γ					
		-3	-2	-1	0	0.5	1
CAPM	*CAPM vs FF3*	0	0	0	0	0	0
	CAPM vs CARH	0	0	0	0	0	0
	CAPM vs FF5	0.001	0	0	0	0	0
FF3	*FF3 vs CARH*	0.158	0.112	0.047	0.012	0.013	0.016
	FF3 vs FF5	0.507	0.532	0.551	0.368	0.195	0.093
CARH	*CARH vs FF5*	1	1	1	1	1	1
		Since 2000					
		values of γ					
		-3	-2	-1	0	0.5	1
CAPM	*CAPM vs FF3*	0.142	0.102	0.046	0.010	0.005	0.005
	CAPM vs CARH	0.002	0.001	0	0	0	0
	CAPM vs FF5	0.002	0.001	0	0	0	0
FF3	*FF3 vs CARH*	0.001	0	0	0	0	0
	FF3 vs FF5	0.002	0.001	0	0	0	0
CARH	*CARH vs FF5*	0.128	0.166	0.146	0.375	0.611	1

The table gives the p-values from the F-tests on nested models among benchmarks. A non-rejection of the null means that the simple specification is better than the complex.
10 industry portfolios are added to the initial 25 Fama-French portfolios used for table 3.5.

For the period 2000-2016, we observe that when the preference for higher-order moments is moderate, the ranking is CAPM $\prec$ FF3 $\prec$ FF5 $\prec$ CARH; this result partially confirms the one obtained in subsection 3.4.3. The exception here originates from the permutation between CARH and FF5 as first and second best, respectively. Therefore, we will adopt the following final ranking when the preference for higher-order moments is moderate: CAPM $\prec$ FF3 $\prec$ CARH $\sim$ FF5. As we did not have coherent results in subsection 3.4.3 for $\gamma < 0$, we do not analyze the cases $\gamma = -1$, $\gamma = -2$ and $\gamma = -3$ here.

In summary, the final ranking over the period 1963-1999 is given as follows: CAPM $\prec$ FF5 $\prec$ CARH $\prec$ FF3 for investors with important or extreme preferences for higher-order moments. Also, the final ranking over the period 2000-2016 is given as follows: CAPM $\prec$ FF3 $\prec$ CARH $\sim$ FF5 for investors with moderate preferences for higher-order moments.

3.4.5 Additional analysis: models with non-traded factors

In general, linear asset pricing models with non-traded factors have poor pricing abilities. However, as shown by Pondi (2018) the relevance of these models can be emphasized with dynamical analysis. In this sub-section, we analyze some model with non-traded factors under various levels of preference for higher-order moments. The idea is to see whether the relevance of these models evolves according to preferences for higher-order moments.

We consider three new models: (i) the human capital model analyzed in Pondi (2018) and inspired by Lettau & Ludvigson (2001) and Gospodinov et al. (2014), **C-LAB**, (ii) the consumption CAPM, **CCAPM**, and (iii) the durable CCAPM of Yogo (2006), **D-CCAPM**.[22] New data are quarterly from Q3 1963 to Q4 2016.

Table 3.10 presents the results from the t-tests on the regression parameters of the models. We see that models with non-traded factors are globally insignificant, irrespective of the level of preference for higher-order moments (the intercept is the only "factor" constantly significant). However, it appears that for $\gamma = 1$ and $\gamma = 0.5$, the consumption-wealth ratio *Cay* in C-LAB is significant. Therefore, the relevance of C-LAB for linear pricing is effective only when the preference for higher-order moments is null of very low.

[22] For a deeper presentation of the three models and data sources, see Pondi (2018).

Table 3.10: P-values from t-tests on the regression parameters (quarterly models)

		values of γ					
		-3	-2	-1	0	0.5	1
CCAPM	Int.	0	0	0	0	0	0
	Cnd	0.357	0.370	0.281	0.126	0.097	0.113
D-CCAPM	Int.	0	0	0	0	0	0
	Mkt-Rf	0.526	0.505	0.418	0.275	0.223	0.181
	Cnd	0.471	0.464	0.389	0.221	0.200	0.250
	Cd	0.928	0.820	0.843	0.878	0.982	0.864
C-LAB	Int.	0	0	0	0	0	0
	Mkt-Rf	0.501	0.480	0.376	0.195	0.133	0.097
	Cay	0.890	0.991	0.693	0.121	0.040	0.017
	Lab	0.186	0.200	0.266	0.561	0.755	0.930

The table gives the p-values from the t-tests on the regression parameters, as the optimal SDF is regressed on each quarterly model though the OLS method. The tests are performed with robust standard errors, estimated following Newey & West (1987)

3.4.6 Coherence of the results with Chapter 2

In Pondi (2018) (Chapter 2), a ranking analysis is performed among Fama-French models by using the dynamical (misspecification-robust) significance of the factors.[23] Although the ranking criterion used in this paper is different, we propose in this subsection to assess in which extend our results are coherent with the rankings from that paper. Since Chapter 2 measures model misspecifications with a Hansen–Jagannathan distance, only our results from investors with moderate preferences for higher-order moments are considered for comparison purposes.

Our dynamical analysis has established, among investors with moderate preferences for higher-order moment, that FF5 and CARH are the two best Fama-French models since 2000 (followed by FF3 and CAPM). This result is exactly the same as in Chapter 2. Moreover, our static analysis (over the last 50 years) has also established the same ranking. This is a little bit different from the results in Chapter 2. There, the two models with the highest relevance frequencies over the last 50 years are FF5 and FF3 (followed by CARH and CAPM). The difference may be explained by the fact that Chapter 2 does not consider non-linear SDFs and therefore, cannot fully capture momentum-driven risk-adjustments. Overall the results from this paper are consistent with those from Chapter 2.

[23]For the ranking purpose, the criterion used in Pondi (2018) is the following: a model is significant at time t if at least one of its factors, other than the constant factor, is significant at t

3.5 Conclusion

In this paper, we analyze the pricing abilities of the Fama-French linear factor models according to investors' characteristics. We assume that investors' preferences are not fully characterized by the mean and the variance. Also, we assume that there is a heterogeneity among investors, whose pricing strategies depend on their preferences for higher-order moments. We show that: (i) for investors with moderate preferences for higher-order moments (in absolute value), the benchmarks can be ranked as follows: CAPM $\prec$ FF3 $\prec$ CARH $\sim$ FF5; (ii) for investors with important preferences for higher-order moments, the ranking is CAPM $\prec$ FF3 $\prec$ FF5 $\prec$ CARH; and (iii) for investors with extreme preferences for higher-order moments, the ranking is CAPM $\prec$ FF5 $\prec$ FF3 $\prec$ CARH. Furthermore, we show that although the relevance for linear pricing of the human capital model analyzed in Pondi (2018) is confirmed, this relevance is effective only for investors with null or very low preferences for higher-order moments.

Appendices for Chapter 3 (C)

C1 Proofs

C1.1 Proof of Proposition 1

The HARA utility function of the dual optimal portfolio problem is given by $u(w) = \frac{-1}{\gamma+1}(1-\gamma w)^{\frac{\gamma+1}{\gamma}}$, (with $(1-\gamma w) > 0$). A Taylor expansion of that utility function around the expected optimal wealth gives:

$$\begin{aligned} u(w_t) = u(E(w_t)) &+ (1-\gamma E(w_t))^{\frac{1}{\gamma}}(w_t - E(w_t)) \\ &+ \frac{1}{2}(-1)(1-\gamma E(w_t))^{\frac{1}{\gamma}-1}(w_t - E(w_t))^2 \\ &+ \frac{1}{3!}(1-\gamma)(1-\gamma E(w_t))^{\frac{1}{\gamma}-2}(w_t - E(w_t))^3 \\ &+ \frac{1}{4!}(-1)(1-\gamma)(1-2\gamma)(1-\gamma E(w_t))^{\frac{1}{\gamma}-3}(w_t - E(w_t))^4 \\ &+ \dots \end{aligned} \tag{C.1}$$

It follows that,

$$\begin{aligned} E(u(w_t)) = u(E(w_t)) & \\ &+ \frac{1}{2}(-1)(1-\gamma E(w_t))^{\frac{1}{\gamma}-1}E\left((w_t - E(w_t))^2\right) \\ &+ \frac{1}{3!}(1-\gamma)(1-\gamma E(w_t))^{\frac{1}{\gamma}-2}E\left((w_t - E(w_t))^3\right) \\ &+ \frac{1}{4!}(-1)(1-\gamma)(1-2\gamma)(1-\gamma E(w_t))^{\frac{1}{\gamma}-3}E\left((w_t - E(w_t))^4\right) \\ &+ \dots \end{aligned} \tag{C.2}$$

Therefore, Proposition 1 is true for orders 2, 3 and 4. Let us suppose by induction that the proposition is true for an order $p \geq 4$. We are going to show that the relation also holds for $p+1$.

By the assumption in induction, we have $u^{(p)}(w_t) = (-1)^{p+1}\prod_{j=0}^{p-2}(1-j\gamma)\cdot(1-\gamma w_t)^{\frac{1}{\gamma}-(p-1)}$.

Therefore,

$$u^{(p+1)}(w_t) = (-1)^{p+2}\prod_{j=0}^{p-1}(1-j\gamma)\cdot(1-\gamma w_t)^{\frac{1}{\gamma}-p}\ . \tag{C.3}$$

The result follows. ■

C1.2 Proof of Proposition 2

By Proposition 1, the weight $W_p(\gamma)$ associated with a moment of order $p \geq 3$ is given as follow:

$$W_p(\gamma) = -W_{p-1}(\gamma)\cdot\frac{(p-1)!}{p!}\left(1-(p-2)\gamma\right)\left(1-\gamma E(w_t)\right)^{-1} \tag{C.4}$$

If p is odd, then by the "*conditions on the signs of weights*", $W_p > 0$ and $W_{p-1} < 0$. Therefore, $(1-(p-2)\gamma) > 0$; so $\gamma < \frac{1}{p-2}$.
If p is even, we have $W_p < 0$ and $W_{p-1} > 0$. The result follows by the same argument as previously.
When p is very large, then $\gamma \leq \underset{p\to\infty}{lim}\left(\frac{1}{p-2}\right) = 0$. ■

Bibliographie

Bibliography

Almeida, C., Ardison, K., & Garcia, R. (2018). Nonparametric assessment of hedge fund performance. Working paper.

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

Almeida, C. & Garcia, R. (2017). Economic implications of nonlinear pricing kernels. *Management Science*, *63*(10), 3361–3380.

Ando, M. & Hodoshima, J. (2006). The robustness of asset pricing models: Coskewness and cokurtosis. *Finance Research Letters*, *3*(2), 133–146.

Ang, A. & Chen, J. (2007). Capm over the long run: 1926–2001. *Journal of Empirical Finance*, *14*(1), 1–40.

Ang, A. & Kristensen, D. (2012). Testing conditional factor models. *Journal of Financial Economics*, *106*(1), 132–156.

Antoine, B., Proulx, K., & Renault, E. (2018). Pseudo-true sdfs in conditional asset pricing models. Working Paper.

Barillas, F. & Shanken, J. (2018). Comparing asset pricing models. *Journal of Finance*, *73*(2), 715–754.

Bekaert, G. & Wu, G. (2000). Asymmetric volatility and risk in equity markets. *Review of Financial Studies*, *13*(1), 1–42.

Black, F. (1976). Studies of stock price volatility changes. *In: Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association*, 177–181.

Bollerslev, T., Engle, R. F., & Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, *96*(1), 116–131.

Borwein, J. M. & Lewis, A. S. (1991). Duality relationships for entropy-like minimization problems. *SIAM Journal on Control and Optimization*, *29*(2), 325–338.

Bryzgalova, S. (2016). Spurious factors in linear asset pricing models. Working Paper.

Burnside, C. (2016). Identification and inference in linear stochastic discount factor models with excess returns. *Journal of Financial Econometrics*, *14*(2), 295–330.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, *52*(1), 57–82.

Chapman, D. A. (1997). Approximating the asset pricing kernel. *Journal of Finance*, *52*(4), 1383–1410.

Chen, Z. & Knez, P. J. (1996). Portfolio performance measurement: Theory and applications. *Review of Financial Studies*, *9*(2), 511–555.

Chung, Y. P., Johnson, H., & Schill, M. J. (2006). Asset pricing when returns are nonnormal: Fama-french factors versus higher-order systematic comoments. *Journal of Business*, *79*(2), 923–940.

Cochrane, J. H. (2001). *Asset Pricing*. Princeton University Press.

Cosemans, M., Frehen, R., Schotman, P. C., & Bauer, R. (2016). Estimating security betas using prior information based on firm fundamentals. *Review of Financial Studies*, *29*(4), 1072–1112.

Dittmar, R. F. (2002). Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *Journal of Finance*, *57*(1), 369–403.

Engle, R. F., Lilien, D. M., & Robins, R. P. (1987). Estimating time varying risk premia in the term structure: The arch-m model. *Econometrica*, *55*(2), 391–407.

Engle, R. F. & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, *48*(5), 1749–1778.

Fama, E. F. & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, *25*(1), 23–49.

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

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

Fama, E. F. & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607–636.

Feng, G., Giglio, S., & Xiu, D. (2017). Taming the factor zoo. Working Paper.

Ferson, W. E. & Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of Political Economy*, *99*(2), 385–415.

Gagliardini, P., Ossola, E., & Scaillet, O. (2016). Time-varying risk premium in large cross-sectional equity dataset. *Econometrica*, *84*(3), 985–1046.

Gagliardini, P. & Ronchetti, D. (2014). Comparing asset pricing models by the conditional hansen-jagannathan distance. Working Paper.

Garcia, R. (2018). Extracting stochastic discount factors from data. Presentation, XVIII Econtro Brasileiro De Finanças.

Ghysels, E. (1998). On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *Journal of Finance*, *53*(2), 549–573.

Giglio, S. & Xiu, D. (2017). Inference on risk premia in the presence of omitted factors. Working Paper.

Gospodinov, N., Kan, R., & Robotti, C. (2014). Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors. *Review of Financial Studies*, *27*(7), 2139–2170.

Gospodinov, N., Kan, R., & Robotti, C. (2017). Spurious inference in reduced-rank asset-pricing models. *Econometrica*, *85*(5), 1613–1628.

Hansen, L. P. & Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy*, *99*(2), 225–262.

Hansen, L. P. & Jagannathan, R. (1997). Assessing specification errors in stochastic discount factor models. *Journal of Finance*, *52*(2), 557–590.

Hansen, L. P. & Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, *50*(5), 1269–1286.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the cross-section of expected returns. *Review of Financial Studies*, *29*(1), 5–68.

Harvey, C. R. & Siddique, A. (2000). Conditional skewness in asset pricing tests. *Journal od Finance*, *55*(3), 1263–1295.

Hou, K., Xue, C., & Zhang, L. (2014). A comparison of new factor models. NBER Working Paper.

Jagannathan, R. & Wang, Z. (1996). The conditional capm and the cross-section of expected returns. *Journal of Finance*, *51*(1), 3–53.

Jensen, G. R., Mercer, J. M., & Johnson, R. R. (1996). Business conditions, monetary policy, and expected security returns. *Journal of Financial Economics*, *40*(2), 213–237.

Jondeau, E. & Rockinger, M. (2006). Optimal portfolio allocation underhigher moments. *European Financial Management*, *12*(1), 29–55.

Kan, R. & Robotti, C. (2009). Model comparison using the hansen-jagannathan distance. *Review of Financial Studies*, *22*(9), 3449–3490.

Kan, R., Robotti, C., & Shanken, J. (2013). Pricing model performance and the two-pass cross-sectional regression methodology. *Journal of Finance*, *68*(6), 2617–2649.

Kan, R. & Zhang, C. (1999a). GMM tests of stochastic discount factor models with useless factors. *Journal of Financial Economics*, *54*(1), 103–127.

Kan, R. & Zhang, C. (1999b). Two-pass tests of asset pricing models with useless factors. *Journal of Finance*, *54*(1), 203–235.

Kelly, B. T., Pruitt, S., & Su, Y. (2017). Instrumented principal component analysis. Working paper, SSRN.

Kleibergen, F. (2009). Tests of risk premia in linear factor models. *Journal of Econometrics*, *149*(2), 149–173.

Kleibergen, F. & Zhan, Z. (2015). Unexplained factors and their effects on second pass r-squared's. *Journal of Econometrics*, *189*(1), 101–116.

Kleibergen, F. & Zhan, Z. (2018). Identification-robust inference on risk premia of mimicking portfolios of non-traded factors. *Journal of Financial Econometrics*, *16*(2), 155–190.

Knight, K. & Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, *28*(5), 1356–1378.

Kozak, S., Nagel, S., & Santosh, S. (2017). Shrinking the cross section. Working Paper 24070, NBER.

Kraus, A. & Litzenberger, R. H. (1976). Skewness preference and the valuation of risk assets. *Journal of Finance*, *31*(4), 1085–1100.

Lettau, M. & Ludvigson, S. (2001). Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy*, *109*(6), 1238–1287.

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

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, *47*(1), 13–37.

Lo, A. W. & Mackinlay, C. (1990). Data-snooping biases in tests of financial asset pricing models. *Review of Financal Studies*, *3*(3), 431–467.

Magnus, J. R. & Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.

Nagel, S. & Singleton, K. J. (2011). Estimation and evaluation of conditional asset pricing models. *Journal of Finance*, *66*(3), 873–909.

Newey, W. K. & West, K. D. (1987). A simple, positive-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Petkova, R. (2006). Do the Fama-French factors proxy for innovations in predictive variables? *The Journal of Finance*, *61*(2), 581–612.

Petkova, R. & Zhang, L. (2005). Is value riskier than growth? *Journal of Financial Economics*, *78*(1), 187–202.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, *7*(2), 186–199.

Pondi, E. M. (2017). Time-varying risk premia with intermittently useless factors. Working Paper.

Pondi, E. M. (2018). Dynamical relevance of factors in misspecified asset pricing models. Working Paper.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, *13*(3), 341–360.

Scott, R. C. & Horvath, P. A. (1980). On the direction of preference for moments of higher order than the variance. *Journal of Finance*, *35*(4), 915–919.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*(3), 425–442.

Shiller, R. J. (2000). *Irrational Exuberance*. Princeton University Press.

Smith, D. R. (2007). Conditional coskewness and asset pricing. *Journal of Empirical Finance*, *14*(1), 91–119.

Wang, K. Q. (2003). Asset pricing with conditioning information: A new test. *Journal of Finance*, *58*(1), 161–196.

White, H. (2000). A reality check for data snooping. *Econometrica*, *68*(5), 1097–1126.

Yogo, M. (2006). A consumption-based explanation of expected stock returns. *Journal of Finance*, *61*(2), 539–580.