7-2019

A quantile-based asset pricing model

Tomohiro ANDO

Jushan BAI

Mitohide NISHIMURA

Jun YU
Singapore Management University, yujun@smu.edu.sg

Citation

ANDO, Tomohiro; BAI, Jushan; NISHIMURA, Mitohide; and YU, Jun. A quantile-based asset pricing model. (2019). *SMU Economics and Statistics Working Paper Series, Paper No. 15-2019*. 1-51.
Available at: https://ink.library.smu.edu.sg/soe_research/2290

SMU ECONOMICS & STATISTICS

A Quantile-based Asset Pricing Model

Tomohiro Ando, Jushan Bai, Mitohide Nishimura, Jun Yu

July 2019

Paper No. 15-2019

A Quantile-based Asset Pricing Model

July 13, 2019

Tomohiro Ando,[1] Jushan Bai,[2] Mitohide Nishimura,[3] and Jun Yu[4]

Abstract

It is well-known that the standard estimators of the risk premium in asset pricing models are biased when some price factors are omitted. To address this problem, we propose a novel quantile-based asset pricing model and a new estimation method. Our new asset pricing model allows for the risk premium to be quantile-dependent and our estimation method is applicable to models with unobserved factors. It avoids biased estimation results and always ensures a positive risk premium. The method is applied to the U.S., Japan, and U.K. stock markets. The empirical analysis demonstrates the clear benefits of our approach.

JEL Classification: G12, G15

Keywords: Five-factor model; Quantile-based asset pricing model; Risk premium

[1]Melbourne Business School, Melbourne University, T.Ando@mbs.edu. 200 Leicester Street, Carlton, Victoria 3053, Australia.

[2]Department of Economics, Columbia University, jb3064@columbia.edu. 1019 International Affairs Building 420 West 118 Street New York, NY 10027 USA

[3]Nikko Asset Management Co. Ltd, mitohide.nishimura@nikkoam.com. The opinions expressed in this article are the author's own and do not reflect the view of Nikko Asset Management.

[4]Corresponding author. School of Economics and Lee Kong Chian School of Business, Singapore Management University, yujun@smu.edu.sg. 90 Stamford Rd. Singapore 178903

1 Introduction

Factor-based asset pricing models are highly popular for several compelling reasons. First, they can explain a cross-section of expected stock returns. Second, they offer frameworks to test the validity of asset pricing models. Third, they allow users to estimate the risk premium of factors, usually via a two-pass regression procedure. Most factor-based asset pricing models share several common features. First, they assume that factors are observed. Second, a linear relationship between expected returns and factors is adopted. Third, it is assumed that all relevant factors are included in the models. Finally, it is assumed that no irrelevant factor is used in the models.

These assumptions have important implications for model estimation, specification analysis, model comparison, estimation of the risk premium and other applications of a model. For example, when a model is misspecified, a more (or less) important factor may become less (or more) important. Additionally, in a misspecified model, the estimated risk premium can be negative, although in theory, the risk premium, which is what an investor should be compensated for bearing the source of risk, must be nonnegative.

Empirical researchers may resort to economic theory for guidance on a functional form and factors. For example, the arbitrage pricing theory of Ross (1976) predicts a linear relationship between expected returns and factors. However, no economic theory specifies a complete list of factors. Most asset pricing models simply cannot include all sources of relevant risk. Moreover, functional forms, such as linearity, usually come from the assumption about the utility function for the representative agent. When the assumption is wrong, the linear relationship between the expected returns and factors may no longer be valid.

Serious attempts to evaluate competing asset pricing models by accounting for model misspecification have been made in the recent literature. Kan et al. (2013)

obtained the asymptotic distribution of R^2. They showed that the asymptotic distribution of the difference between the sample R^2 of two candidate models depends on whether the models are correctly specified and whether they are nested. Kan and Robotti (2015) considered multiple model comparison tests. To search for useful factors, Feng et al. (2017) proposed a new model selection method to evaluate the marginal contribution of a new factor when a set of a large number of factors have been included in the model. The method is robust to model misspecification in the sense that a large number of factors may include redundant factors.

While these recent studies are attractive, these methods typically prepare a set of factors somewhat arbitrarily. However, there is no theoretical guarantee that the set of factors indeed contain all true risk factors required to explain the asset return. Giglio and Xiu (2017) proposed a three-pass method to estimate the risk premium. This method is shown to be valid when the observed factors are a strict subset of the true factors and when the observed factors are subject to measurement errors. The impact of model misspecification is also studied by Kan and Zhang (1999), Kan et al. (2013), Shanken and Zhou (2007), Gospodinov et al. (2013), Kleibergen and Zhan (2018), etc. However, these previous studies assume that the sensitivity to risk factors is constant over quantiles, while the possibility of quantile dependence of sensitivity to risk factors has been reported in the literature (see, for example, Ando and Bai (2018)). To measure the quantile-dependent risk premium, no systematic solution has been proposed thus far.

This paper directly addresses the abovementioned problems by introducing a new asset pricing model and a new estimation and inferential procedure. Our model and estimation method contain several salient features. First, the model assumes a linear relationship between quantiles of returns (instead of expected return) and factors. Second, both observed and unobserved factors are allowed in our models. As a benefit of our approach, we can avoid the omitted variable bias when some important common factors are missing in the model. Third, restrictions on monotonicity of the risk premium are imposed in our estimation method. Fourth, we develop asymptotic theory to the estimator under a double asymptotic argument (that is, the number of assets and the number of time series observations both go to infinity), facilitating statistical inference. Finally, unlike standard two-pass regressions to estimate the risk premium (see, e.g., Fama and Macbeth (1973), Ferson and Harvey (1991), Shanken (1992), Jagannathan and Wang (1998), Lewellen et al. (2010), Bai and Zhou (2015) and Gagliardini et al. (2016)), this paper estimates the risk premium based on our novel quantile-based asset pricing model. This new asset pricing model allows for the risk premium to be quantile-dependent. Moreover, our novel approach always ensures a positive risk premium.

Our quantile-based asset pricing model plays an important role when the risk premium is quantile-dependent. More specifically, the previous methods for estimating the risk premium ignore the quantile dependency of the risk premium. One strong assumption implicitly imposed on the previous methods is that the level of the risk premium does not depend on any quantile point. In other words, the risk premium is constant, regardless of the scenario of the market. Intuitively, however, it is natural to consider that investors demand a higher market risk premium when the market faces pessimistic trends. Thus, it is reasonable to assume that the risk premium depends on the quantile. Indeed, our empirical analysis reveals that the risk premium indeed depends on the quantile.

We also make theoretical contributions by developing an asymptotic theory for the proposed procedure. Due to the presence of estimation errors in unobservable common factor structures, the development of these results is nontrivial. If the estimation error for the factor structure is not negligible, then it is important to investigate the statistical properties of the proposed risk premium estimator by taking into account the effect of the estimated factor structure. In our asymptotic framework, the time-series dimension and the individual dimension are diverging. Therefore, we develop a novel strategy for establishing the asymptotic theory.

The model and the estimation method are first applied to a large set of equity portfolios, that is, the universe of component stocks from S&P500 in the U.S., TOPIX from Japan, and the FTSE all-share index from the U.K.. The number of unobserved common factors is identified, and the factors are estimated. The risk premia are obtained. We then investigate the impact of passive funds on the stock markets using the estimated common factor. More specifically, we explore whether the capital flows from/to passive funds have any impact on the risk premium. Our empirical results indicate that passive flow is related to the risk premium.

The remainder of this paper is organized as follows. Section 2 introduces the model and the estimation method. Section 3 establishes the asymptotic theory for our quantile-based 2-pass procedure. Section 4 reports and discusses empirical results based on the large set of equity portfolios. Section 5 examines the impact of passive funds on the risk premium. Our method reveals that the passive flow influences the risk premia in the U.S., Japan, and U.K. stock markets. Section 6 concludes. The proofs of the theorems are collected in the Appendix. The Appendix also contains a set of assumptions imposed on our procedure.

2 The Method

Our method proceeds in two steps. First, we estimate a quantile-based asset pricing model to extract unobserved common factors and their loadings from a large panel of asset returns. This first step allows us to avoid the omitted variable bias problem. Second, we introduce quantile-based 2-pass procedure motivated from Fama and MacBeth (1973). The risk premium, which depends on quantile points, is then estimated.

2.1 A quantile-based asset pricing model

Suppose that an excess return is measured over T time periods together with some common factors. For the i-th financial instrument ($i = 1, ..., N$), at time t, its return y_{it} is observed together with a p-dimensional vector of observable factors $\boldsymbol{x}_{it} = (x_{it,1}, ..., x_{it,p})'$. As shown below, our method is useful to study the risk premium associated with the common factors in an asset pricing model.

Consider the following structure for the τ-th conditional quantile function of y_{it}:

$$Q_{y_{it}}\left(\tau|\boldsymbol{x}_{it}, \boldsymbol{b}_{i,\tau}, \boldsymbol{f}_{t,\tau}, \boldsymbol{\lambda}_{i,\tau}\right) = \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} + \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}, \quad i = 1, \ldots, N,\ t = 1, \ldots, T, \qquad (1)$$

where $\boldsymbol{b}_{i,\tau} = (b_{i,0,\tau}, b_{i,1,\tau}, ..., b_{i,p,\tau})'$ is a p-dimensional vector of regression coefficients. Following Ando and Bai (2018), we have the unobservable factor structure $\boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}$, where $\boldsymbol{f}_{t,\tau}$ is an $r_\tau \times 1$ vector of unobservable factors and $\boldsymbol{\lambda}_{i,\tau}$ represents the unobservable factor loadings. Note that the dimension of unobservable structures may vary over quantiles. Studies on factors that explain the cross section of expected stock returns have reported several hundred factors; see, for example, Harvey et al. (2015). Thus, the p-dimensional observable factor may not be sufficient to capture the cross-sectional variation of asset returns well. To increase an explanatory power of asset pricing, the unobservable factor structure $\boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}$ is crucial. For the linear factor models that focus on the conditional mean of asset return y_{it}, refer to Chamberlain and Rothschild (1983), Connor and Korajzcyk (1986), Bai and Ng (2002), Bai (2009), Ando and Bai (2017) and the references therein.

Compared to typical asset pricing models in the literature, Model (1) has a few unique features. First, instead of assuming that the expected return of y_{it} is a linear function of factors, we assume the conditional quantile of y_{it} is a linear function of factors. As the distribution of y_{it} typically has heavy tails and quantiles are robust against outliers, we expect that the quantile regression and the estimates are more robust than those in the asset pricing models based on the conditional mean and the ordinary least squares regression. Second, we do not make serious attempts to find all observed factors to explain quantiles of y_{it}. Instead, we believe that some factors are unobserved, and we include them as the latent variables. As these factors are common across i, we hope to consistently estimate them when $N \to \infty$. Finally, we assume both N and T go to infinity. As we will show below, under a mild condition, the structure of the latent variables can be consistently detected.

To estimate the unknown parameters $B_\tau = (\boldsymbol{b}_{1,\tau}, ..., \boldsymbol{b}_{N,\tau})'$, $\Lambda_\tau = (\boldsymbol{\lambda}_{1,\tau}, ..., \boldsymbol{\lambda}_{N,\tau})'$, and $F_\tau = \left(\boldsymbol{f}_{1,\tau}, ..., \boldsymbol{f}_{T,\tau}\right)'$, a panel quantile approach is needed. In particular, given value of r, we estimate B_τ, Λ_τ and F_τ by minimizing

$$\ell_\tau(Y|X, B_\tau, F_\tau, \Lambda_\tau) = \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau \left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}\right),$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile loss function. Denote the estimators by $\hat{\boldsymbol{b}}_{i,\tau}$, $\hat{\boldsymbol{f}}_{t,\tau}$, $\hat{\boldsymbol{\lambda}}_{i,\tau}$. Under a set of mild conditions, as reported in the following proposition, Ando and Bai (2018) showed that the asymptotic distribution of the estimated common factor $\hat{\boldsymbol{f}}_{t,\tau}$ and factor loadings $\hat{\boldsymbol{\lambda}}_{i,\tau}$ is a multivariate normal distribution.

Proposition 1 *(Ando and Bai (2018) Theorem 2) Suppose that Assumption A $\sim$ Assumption E hold. Then, we have*

$$T^{1/2}\left(\hat{\boldsymbol{\lambda}}_{i,\tau} - \boldsymbol{\lambda}_{i,0,\tau}\right) \sim N(0, \Sigma_{i,\tau}), \quad \text{and} \quad N^{1/2}\left(\hat{\boldsymbol{f}}_{t,\tau} - \boldsymbol{f}_{t,0,\tau}\right) \sim N(0, \Theta_{t,\tau}),$$

where $\Sigma_{i,\tau} = \tau(1-\tau)\Gamma_{i,0,\tau}^{-1}V_{0,\tau}\Gamma_{i,0,\tau}^{-1}$ *and* $\Theta_{t,\tau} = \tau(1-\tau)\Psi_{t,0,\tau}^{-1}R_{0,\tau}\Psi_{t,0,\tau}^{-1}$ *with*

$$\Gamma_{i,0,\tau} := \text{plim}_{T\to\infty}T^{-1}\sum_{t=1}^{T} g_{it,0}\boldsymbol{f}_{t,0,\tau}\boldsymbol{f}_{t,0,\tau}', V_{0,\tau} := \text{plim}_{T\to\infty}T^{-1}\sum_{t=1}^{T}\boldsymbol{f}_{t,0,\tau}\boldsymbol{f}_{t,0,\tau}',$$

$$\Psi_{t,0,\tau} := \text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N} g_{it,0}\boldsymbol{\lambda}_{i,0,\tau}\boldsymbol{\lambda}_{i,0,\tau}', R_{0,\tau} := \text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_{i,0,\tau}\boldsymbol{\lambda}_{i,0,\tau}',$$

$g_{it,0} := g\left(0|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0,\tau}, \boldsymbol{\lambda}_{i,0,\tau}\right)$ *and* $g(\cdot)$ *being the true conditional density function of* $y_{it} - Q_{y_{it}}(\tau|\boldsymbol{x}_{it}, \boldsymbol{b}_{i,0,\tau}, \boldsymbol{f}_{t,0,\tau}, \boldsymbol{\lambda}_{i,0,\tau})$.

This proposition implies that the estimated common factor and the estimated factor loadings converge to their respective true values with $\sqrt{T}$ and $\sqrt{N}$ convergence rates. Note that the true dimension of unobservable structures is unknown. Following Ando and Bai (2018), the number of common factors is selected by minimizing the following information criterion:

$$IC_\tau(r) = \log\left[\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{b}}_{i,\tau}(r) - \hat{\boldsymbol{f}}_{t,\tau}(r)'\hat{\boldsymbol{\lambda}}_{i,\tau}(r)\right)\right] + r \times q(N,T), \quad (2)$$

where $\hat{\boldsymbol{b}}_{i,\tau}(r)$, $\hat{\boldsymbol{\lambda}}_{i,\tau}(r)$ and $\hat{\boldsymbol{f}}_{t,\tau}(r)$ is the estimated model parameters given the number of common factors r, and $q(N,T)$ is the penalty term to capture model complexity. In this paper, we use $q(N,T) = \log\left(\frac{NT}{N+T}\right)\left(\frac{N+T}{NT}\right)$. The following proposition, which also follows the approach of Ando and Bai (2019), ensures that this penalty term allows us to determine the true dimension of unobservable structure r_0.

Proposition 2 *(Ando and Bai (2018) Theorem 3) Suppose that Assumption A $\sim$ Assumption E hold. Under the model selection criterion (2), we have a consistent model selector of the true dimension of the interactive effects (i.e., the true number of common factors)* $r_{0,\tau}$.

Based on this proposition, one can expect that the estimated factor structure can well capture the unobservable cross-sectional dependence. The estimated factor structure plays an important role to adjust the omitted factors that potentially explain the cross-sectional returns. By analyzing the common factors, it may be possible to speculate what factors other than the observed factors are driving the market.

Our procedure also obtains the r_τ-dimensional factor structure $\hat{\boldsymbol{f}}_{t,\tau}$ and the corresponding factor loading $\hat{\boldsymbol{\lambda}}_{i,\tau}$ for a set of quantile points $\tau_1, ..., \tau_K$. Here, K denotes the number of quantile points. Therefore, for any quantile point τ, we can calculate the asset return adjusted by the unobservable structure $\hat{\boldsymbol{f}}'_{t,\tau}\hat{\boldsymbol{\lambda}}_{i,\tau}$. This adjusted return contains the information on the risk premium for the set of observables $\boldsymbol{x}_{it}$.

2.2 Quantile-dependent risk premium

After the quantile function is estimated, we obtain $y_{it} - \hat{\boldsymbol{f}}'_{t,\tau_k}\hat{\boldsymbol{\lambda}}_{i,\tau_k}$, which is the return adjusted by the omitted unobservable structure. This adjustment is important, as endogeneity would be an issue otherwise. This section proposes a new method to estimate risk premium in the second pass. Our 2-pass approach is motivated by the well-known 2-pass approach of Fama and MacBeth (1973). Our approach has several important properties, which we explain first. The practical implementation of our approach is then described.

Property 1: Quantile-dependent

Intuitively, an investor may ask for more risk premia for negative returns and less for positive returns. This implies that the risk premia will vary over the quantile points of asset returns. From this perspective, the following relationship is expected:

$$Q_{z_{it}}\left(\tau|\boldsymbol{x}_{it}, \boldsymbol{r}(\tau), \boldsymbol{b}_{i,\tau}\right) \equiv \boldsymbol{r}(\tau)'\boldsymbol{b}_{i,\tau}, \tag{3}$$

where $\boldsymbol{r}(\tau) = (r_1(\tau), ..., r_p(\tau))$ is the p-dimensional risk premium parameter, and $z_{it} \equiv z_{it} - \boldsymbol{f}'_{t,0,\tau}\boldsymbol{\lambda}_{i,0,\tau}$ is the asset return adjusted by the omitted unobservable structure.

When the risk premium and the regression coefficient do not depend on the quantiles such that $\boldsymbol{r}(\tau) = \boldsymbol{r}$ and $\boldsymbol{b}_{i,\tau} = \boldsymbol{b}_i$ for all τ, it is obvious that the quantile function does not depend on τ either. In this case, the quantile function (3) reduces to the linear model $\boldsymbol{r}'\boldsymbol{b}_i$. In other words, our model reduces to the model employed in the 2-pass approach of Fama and MacBeth (1973). Needless to say, this constant linear model can be estimated using the 2-pass approach of Fama and MacBeth (1973). However, our empirical results reveal that the risk premium is quantile-dependent.

Property 2: Positiveness of risk premium

We note that the risk premium should be positive, that is, $\boldsymbol{r}(\tau) \geq 0, \quad \tau \in (0, 1)$. Although our approach allows for the risk premium to be negative, it would be difficult to understand the negative risk premium. Additionally, if the risk premium is negatively estimated, the estimation procedure would face some technical issues. One possible reason for the estimated risk premium being negative is model misspecification. Because our approach carefully avoids omitted variable bias and endogeneity, we can address the issue of misspecification.

Property 3: Monotonicity of risk premium

As discussed in Property 1, an investor may ask for more risk premia for negative returns and less for positive returns. From this point of view, the risk premium should satisfy the monotonicity restriction

$$\boldsymbol{r}(\tau_a) \geq \boldsymbol{r}(\tau_b), \quad \tau_a < \tau_b, \tag{4}$$

for any quantile points $0 < \tau_a < \tau_b < 1$. If an investor's attitude toward risk does not depend on any quantiles, the risk premium reduces to the constant risk premium. Our approach estimates the constant risk premium when the risk premium does not depend on any quantiles τ.

Property 4: Monotonicity of the quantile function of asset return

Finally, the quantile function itself should satisfy the monotonicity restriction. More specifically, $Q_{z_{it}}(\tau|\boldsymbol{x}_{it}, \boldsymbol{r}(\tau), \boldsymbol{b}_{i,\tau}, \boldsymbol{\lambda}_{i,\tau}) = \boldsymbol{r}(\tau)'\boldsymbol{b}_{i,\tau}$ should satisfy the monotonicity restriction,

$$\boldsymbol{r}(\tau_a)'\boldsymbol{b}_{i,\tau_a} \leq \boldsymbol{r}(\tau_b)'\boldsymbol{b}_{i,\tau_b}, \tag{5}$$

for any quantile points $0 < \tau_a < \tau_b < 1$. This implies that the risk premium $\boldsymbol{r}(\tau)$ should satisfy a monotone restriction of $Q_{z_{it}}(\tau|\boldsymbol{x}_{it}, \boldsymbol{r}(\tau), \boldsymbol{b}_{i,\tau})$ from the definition of the quantile function.

Property 5: Time-varying

In the last 15 years, investors have witnessed the subprime crisis in the U.S., the collapse of Lehman Brothers in 2007–2008, and the subsequent sovereign debt crisis in Europe. It is natural to expect that the risk premium surged during these chaotic periods compared with normal periods. Thus, we need to measure the risk premium by taking account its time-varying property.

In the next section, we propose a practical implementation procedure to obtain the risk premium that satisfies the restrictions (3), (4) and (5).

2.3 Estimation of risk premium

To accommodate the time-varying property of the risk premium, we use the rolling strategy, which can be handled by using the certain window of the historical data. In our empirical analysis, the past 250 days are used. For simplicity of notation, we drop the time dependency of the risk premium.

Taking account of the properties of the risk premium in (3), (4) and (5), we estimate the risk premium by solving

$$\hat{\boldsymbol{r}}(\tau) = \text{argmin} \frac{1}{KN} \sum_{k=1}^{K} \sum_{i=1}^{N} \rho_{\tau_k} \left(y_{it} - \hat{\boldsymbol{f}}'_{t,\tau_k} \hat{\boldsymbol{\lambda}}_{i,\tau_k} - \boldsymbol{r}(\tau_k)' \boldsymbol{b}_{i,\tau_k} \right),$$

under the following restrictions:

$$\boldsymbol{r}(\tau_k)' \boldsymbol{b}_{i,\tau_k} \leq \boldsymbol{r}(\tau_{k+1})' \boldsymbol{b}_{i,\tau_{k+1}}, \quad \text{for } k = 1, ..., K-1, \ i = 1, ..., N, \tag{6}$$

$$\boldsymbol{r}(\tau_k) \geq \boldsymbol{r}(\tau_{k+1}). \text{ for } \ k = 1, ..., K-1, \tag{7}$$

$$\boldsymbol{r}(\tau_k) \geq 0, \text{ for } \ k = 1, ..., K, \tag{8}$$

where $\{\tau_1, \tau_2, ..., \tau_K; \ \tau_k \leq \tau_{k+1}\}$ are a set of K quantile points. In a practical implementation, we use a set of $K = 5$ quantile points, that is, $\tau_1 = 0.05$, $\tau_2 = 0.25$, $\tau_3 = 0.5$, $\tau_4 = 0.75$ and $\tau_5 = 0.95$. It is possible to use finer grids.

There are two issues. One is the value of regression parameter $\boldsymbol{b}_{i,\tau_{k+1}}$, which should also satisfy the restriction (6). The other is that the risk premium should satisfy these restrictions (7) and (8). However, the direct minimization of a loss function under a very large number of restrictions, whose order is $O(KN)$, is extremely time-consuming. This problem can be solved as follows.

Step 1 Similar to Bondell et al. (2010), for each i, we first transform the observable factor structure $\boldsymbol{x}_{it}$ into the unit hypercube $[0,1]^p$. This transformation aims to satisfy the restriction (6). Note that once the transformation is performed, we then transformed back after the estimation while retaining the noncrossing property. Hereafter, we denote $\boldsymbol{x}_{it}$ as the transformed vector in the unit hypercube $[0,1]^p$.

Then, we estimate the regression coefficient vector $\boldsymbol{b}_{i,\tau}$ by minimizing the following objective function:

$$\ell(\boldsymbol{b}_{i,\tau}) = \frac{1}{KT}\sum_{k=1}^{K}\sum_{t=1}^{T}\rho_{\tau_k}\left(y_{it} - \hat{\boldsymbol{f}}'_{t,\tau_k}\hat{\boldsymbol{\lambda}}_{i,\tau_k} - \boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_k}\right), \tag{9}$$

subject to the monotone restriction of the quantile function. This problem can be solved by directly applying the method of Bondell et al. (2010). Then, we obtain a set of $\hat{\boldsymbol{b}}_{i,\tau_k}$ for $i = 1, ..., N$, $k = 1, ..., K$. Note that the estimated regression coefficient satisfies the monotone property of the quantile function.

Step 2 For each element of $\hat{\boldsymbol{b}}_{\tau}$, we first transform the estimated regression coefficients $\hat{\boldsymbol{b}}_{i,\tau}$ into the hypercube $[-1, 0]^p$. This mapping is intended to ensure the monotone property of quantile function (6) and the restrictions on the risk premium (7) and (8) simultaneously.

Then, the risk premium parameter $\boldsymbol{r}(\tau)$ at time t can be estimated by solving

$$\hat{\boldsymbol{r}}(\tau) = \text{argmin}\left[\frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\rho_{\tau_k}\left(y_{it} - \hat{\boldsymbol{f}}'_{t,\tau_k}\hat{\boldsymbol{\lambda}}_{i,\tau_k} - \boldsymbol{r}(\tau_k)'\hat{\boldsymbol{b}}_{i,\tau_k}\right)\right], \tag{10}$$

under the following restriction:

$$\boldsymbol{r}(\tau_1) \geq \boldsymbol{r}(\tau_2) \geq ... \geq \boldsymbol{r}(\tau_K) \geq 0.$$

This estimation can be implemented by the restricted optimization problem. Once we obtain the estimate $\hat{\boldsymbol{r}}(\tau)$, this parameter vector is transformed back to the original space. We then obtain the estimates of the risk premium.

In the next section, we establish a large sample theory for our proposed procedure.

3 Large Sample Theory

In this section, we provide an asymptotic theory of our estimators of the risk premium. There are several technical challenges. First, the estimated factor structure

$\hat{\boldsymbol{f}}'_{t,\tau_k}\hat{\boldsymbol{\lambda}}_{i,\tau_k}$ is plugged into the objective function in (9). Because the corresponding estimation problem involves an estimated factor structure, we must understand whether the estimation error due to the estimated factor is negligible. Second, the estimated risk premium is subject to the estimation uncertainty not only of $\hat{\boldsymbol{f}}'_{t,\tau_k}\hat{\boldsymbol{\lambda}}_{i,\tau_k}$ but also of $\hat{\boldsymbol{b}}_{i,\tau_k}$. If the estimation errors are not negligible, then it is important to investigate the statistical properties of the estimated $\boldsymbol{b}_i$ by taking into account the effect of estimated factor structures. However, these issues are not well understood. This section establishes the asymptotic property of our estimators by taking these errors into account.

Let $\tilde{\boldsymbol{b}}_{i,\tau_k}$ be the constrained infeasible estimator, which is obtained as the minimizer of

$$\tilde{\ell}(\boldsymbol{b}_{i,\tau_k}) \equiv \frac{1}{KT}\sum_{k=1}^{K}\sum_{t=1}^{T}\rho_{\tau_k}\left(y_{it} - \boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k} - \boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_k}\right),$$

subject to the following restrictions:

$$\boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_{k-1}} \le \boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_k}, \text{ for } \ t = 1, ..., T, \ k = 2, ..., K.$$

The following theorem establishes the relationship between $\tilde{\boldsymbol{b}}_{i,\tau_k}$ and $\hat{\boldsymbol{b}}_{i,\tau_k}$.

Theorem 1 *Suppose that the conditions $A \sim F$ hold. Then, for any $u \in R^{pK}$,*

$$\left|P\left(\sqrt{T}\left(\hat{\boldsymbol{b}}_{i,\tau_k} - \boldsymbol{b}_{i,0,\tau_k}\right) \le u\right) - P\left(\sqrt{T}\left(\tilde{\boldsymbol{b}}_{i,\tau_k} - \boldsymbol{b}_{i,0,\tau_k}\right) \le u\right)\right| \to 0.$$

so the constrained estimators share the same limiting distribution.

Theorem 1 implies that we can ignore the estimation error of the factor structure. In other words, our approach captures the omitted common factors accurately and thus avoids omitted variable bias.

Define $\tilde{\boldsymbol{r}}(\tau)$ as the risk premium estimator, which is obtained by solving

$$\tilde{\boldsymbol{r}}(\tau) = \text{argmin}\left[\frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\rho_{\tau_k}\left(y_{it} - \boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k} - \boldsymbol{r}(\tau_k)'\boldsymbol{b}_{i,0,\tau_k}\right)\right], \tag{11}$$

under the restriction $\boldsymbol{r}(\tau_1) \ge \boldsymbol{r}(\tau_2) \ge ... \ge \boldsymbol{r}(\tau_K) \ge 0$. In (11), the estimated parameters are $\boldsymbol{f}_{t,0,\tau}$, $\boldsymbol{\lambda}_{i,0,\tau}$ and $\boldsymbol{b}_{i,0,\tau}$ are the true parameter values. Although the risk premium parameter $\boldsymbol{r}(\tau)$ was obtained by using the objective function (10), Theorem 2 ensures that $\tilde{\boldsymbol{r}}(\tau)$ and $\hat{\boldsymbol{r}}(\tau)$ has the same limiting distribution. Additionally, we further define the classical quantile regression estimator

$$\bar{\boldsymbol{r}}(\tau) = \operatorname{argmin} \left[\frac{1}{NK} \sum_{k=1}^{K} \sum_{i=1}^{N} \rho_{\tau_k} \left(y_{it} - \hat{\boldsymbol{f}}'_{t,\tau_k} \hat{\boldsymbol{\lambda}}_{i,\tau_k} - \boldsymbol{r}(\tau_k)' \hat{\boldsymbol{b}}_{i,\tau_k} \right) \right].$$

Note that no constraints are imposed on the risk premium parameter. The following theorem implies that the theoretical results on the standard quantile regression is applicable to our estimator of the risk premium.

Theorem 2 *Suppose that the conditions $A \sim F$ hold. Let $\hat{\boldsymbol{r}}(\tau)$ and $\tilde{\boldsymbol{r}}(\tau)$ be the constrained and unconstrained risk premium estimators, respectively, for the set of quantiles $\boldsymbol{\tau}$. Then, for any $u \in R^{pK}$,*

$$\left| P\left(\sqrt{N} \left(\hat{\boldsymbol{r}}(\tau) - \boldsymbol{r}_0(\tau) \right) \leq u \right) - P\left(\sqrt{N} \left(\tilde{\boldsymbol{r}}(\tau) - \boldsymbol{r}_0(\tau) \right) \leq u \right) \right| \to 0.$$

so the constrained estimators share the same limiting distribution. Moreover, the estimator $\hat{\boldsymbol{r}}(\tau)$ has the same limiting distribution as the classical quantile regression estimator $\bar{\boldsymbol{r}}(\tau)$.

Theorem 2 implies that we can ignore the estimation error of the factor structure and the regression coefficients. Thus, the risk premium estimator will approach the true value with $\sqrt{N}$ convergence rate. Additionally, inference for the $\sqrt{N}$-consistent restricted estimator of the risk premium can be achieved by using the known asymptotic results for classical quantile regression.

4 Empirical Results 1: Risk Premium

We apply our modeling procedure to the dataset from several major stock markets around the world: those of the U.S., Japan, and the U.K..

4.1 Data

The universe of stocks is the S&P 500, TOPIX, and the FTSE All-Share Index. The composite stocks of the index at the end of the year are used as the universe of the following year. Daily stock prices are obtained from Bloomberg, and the data period spans 2006 to 2017.

Table 1 shows the 5%, 25%, 50%, 75% and 95% quantile points of daily stock return for every year. With respect to S&P 500 stock returns, a range between 5% quantile and 95% quantile is typically within −3% to 3%. In 2008 and 2009, the range is much larger due to the global financial crisis, which increased the volatility dramatically. Similar observations can be made for the TOPIX and the FTSE.

For the observable factors in the quantile-based asset pricing model, we employ Fama-French's 5 factors (Fama and French (2015)) and quantify the price of risk on these factors. We obtain Fama-French's 5 factors (North American factors, Japanese factors, and European factors) from French's website.[5] Japanese 5 factors and European 5 factors were converted to JPY and GBP currency, respectively. For the risk-free rate, we employ the three-month deposit rate.

To analyze the universe of stocks for each country, a quantile-based asset pricing model is specified as

$$\begin{aligned} Q_{y_{it}}(\tau|\boldsymbol{x}_{it}) \;=\;& \alpha_{i,\tau} + Mkt_t \times \beta_{Mkt,i,\tau} + HML_t \times \beta_{HML,i,\tau} + SMB_t \times \beta_{SMB,i,\tau} \\ & + RMW_t \times \beta_{RMW,i,\tau} + CMA_t \times \beta_{CMA,i,\tau} + \boldsymbol{f}'_{t,\tau}\boldsymbol{\lambda}_{i,\tau}, \end{aligned} \tag{12}$$

where Mkt_t, HML_t, SMB_t, RMW_t and CMA_t are Fama-French's five factors at time t. Here, Mkt is the return on a region's value-weighted market portfolio minus the risk-free rate, SMB (small minus big) is the average return on the nine small stock portfolios minus the average return on the nine big stock portfolios, HML (high minus low) is the average return on the two value portfolios minus the average return on the two growth portfolios, RMW (robust minus weak) is the average

[5] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html

return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios, and CMA (conservative minus aggressive) is the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios. When Fama-French's five factors capture the behaviors of stock returns very well, the unobserved factor structure $\boldsymbol{f}'_{t,\tau}\boldsymbol{\lambda}_{i,\tau}$ in (12) will become redundant. However, as discussed in the next section, the unobserved factor structure is important for capturing the behaviors of stock returns.

4.2 Estimated common factors

In this section, we report how the number of unobservable factors varies over quantiles, time and country. The period used for estimation is the past 250 days up to the end of every month. Then, we roll the estimation period every month.

Figure 1 (a) shows the selected number of factors for each of the percentiles in the U.S. stock market. It can be seen that there are large variations in the selected number of factors over time. First, it is easy to observe an increase in the selected number of factors in 2007-2008 during the global financial crisis. The selected number of factors is the U.S. is $\hat{r} = 10$ at the 95% quantile in July 2007. During this period, two hedge funds under the Bear Stearns umbrella, which had purchased a large amount of subprime mortgage securities, failed, and the stock market declined. Notably, the selected number of factors increased only in July, and a smaller number of factors are selected before and after July 2007. We also observe the increase in the number of factors in September-October 2008 when Lehman Brothers collapsed. The influence of the subsequent European debt crisis, caused by a concern of Greek departure from the EU, does not seem to have increased the number of factors in the U.S.. This is consistent with the fact that the effect of the European debt crisis is local within Europe. During this time period, the U.S. market remained generally stable. During October 2012-January 2013, we can observe a surge in the number of factors. This period experienced political turmoil due to the expiration of large tax cuts and the forced reduction of fiscal expenditures. In 2013 and 2014, the selected number of factors remained at a relatively low level. During this time period, the U.S. stock market had upward trends and maintained a strong market environment. After 2015, due to the concern of the Chinese economy, the devaluation of RMB and the EU withdrawal referendum in the U.K. on June 26, 2016, the number of factors increased. In summary, we see that the number of factors increases when there are strong shocks to the stock market and when the stock price fluctuates sharply. The influence of shocks is remarkable at lower/upper tails compared with the median, where stock price fluctuation is not large.

Figure 1 (b) shows the selected number of factors for each of the percentiles in the Japanese stock market. The trend of the selected number of factors is similar to that of the U.S. stock market. It generally increases when the stock price fluctuates sharply. For example, in April 2013, share prices surged because the Bank of Japan announced quantitative and qualitative monetary easing (quantitative-qualitative easing, QQE). The announcement clarified that the Bank of Japan was purchasing financial assets such as government bonds and exchange-traded funds (ETFs) more than ever. As expected, we observe an increase in the number of selected factors.

Finally, Figure 1 (c) shows the selected number of factors for each of the percentiles in the U.K. stock market. Again, there is a strong tendency for the number of factors at the upper and lower quantiles (5%, 95%) to be larger than that at the 50% quantile. The number of factors also exhibits a trend similar to that in the U.S. and Japanese stock markets. Unlike in the U.S. stock market, the European debt crisis increased the number of selected factors in the U.K. stock market, as expected. Moreover, it can be seen that the factor number has increased greatly as a result of the EU withdrawal referendum in the U.K..

4.3 Estimated risk premium

Figure 2 shows the estimated risk premia of Fama-French's 5 factors. The risk premia of these factors are obtained by estimating

$$Q_{z_{it}}(\tau|\boldsymbol{x}_{it}) = r_{Mkt}(\tau)\times\beta_{Mkt,i,\tau}+r_{HML}(\tau)\times\beta_{HML,i,\tau}+r_{SMB}(\tau)\times\beta_{SMB,i,\tau} +r_{RMW}(\tau)\times\beta_{RMW,i,\tau}+r_{CMA}(\tau)\times\beta_{CMA,i,\tau}, \quad (13)$$

where $\{\beta_{Mkt,i,\tau},\beta_{HML,i,\tau},\beta_{SMB,i,\tau},\beta_{RMW,i,\tau},\beta_{CMA,i,\tau}\}$ $(i=1,...,N)$ are obtained in Step 1 described in Section 2.2. Then, the risk premia of Fama-French's 5 factors that depend on the quantile $\{r_{Mkt}(\tau),r_{HML}(\tau),r_{SMB}(\tau),r_{RMW}(\tau),r_{CMA}(\tau)\}$ are obtained by using Step 2 given in Section 2.2. In this section, we discuss these estimation results.

4.3.1 Comparison over factors

Regarding the market factor (Mkt), the risk premium moves up and down between 0 and 6 for all countries. The risk premium at the 5% quantile occasionally rises to an exceptionally high level. In addition, except at the 5% quantile, the risk premium at the other quantiles tends to remain at a similar level. This implies that the investors request a greater risk premium for the 5% quantiles.

For the size factor (SMB), the level of risk premium in the U.S. market is lower than that in the other two markets. In fact, the risk premium at the 5% quantile ranges from nearly 0 to 1 in the U.S., while it ranges from nearly 0 to 6 in the Japanese and U.K. markets. The risk premium is not as sensitive to quantiles in the U.S. and Japan, for each quantile, suggesting that the risk premium required for large stocks is similar to that for small stocks. However, the risk premium is very sensitive to quantiles in the U.K.. In particular, a large risk premium is requested for small stocks.

Third, the estimated risk premium for the value factor (HML) ranges from nearly 0 to about 4 in the U.S. market. In contrast, it ranges from nearly 0 to 10 and from nearly 0 to 8 for the Japanese and U.K. markets, respectively. In the U.S., a large premium is required at the 5% quantile compared to the other quantiles. This indicates that investors require a large risk premium at lower quantiles. On the other hand, in Japan and in the U.K., the level of risk premium varies across quantiles. For example, the risk premium at 5% quantile is always larger than other quantiles in Japan and the U.K..

For the profit margin (RMW) factor, the ranges of the risk premium in the U.S., Japan and the U.K. are from nearly 0 to 3.5, nearly 0 to 7 and nearly 0 to 8, respectively. In the U.S., the risk premium is very small at all quantiles during the periods between 2006 and 2008 and between 2013 and 2015. This means that as long as the stock market is on a steady upward trend, the demand for a risk premium on the margin factor is nearly zero regardless of the performance of the stock. In Japan and the U.K., we observe a tendency for the risk premium to decrease as the quartile rises for the entire period.

Finally, for the investment attitude (CMA) factor, the risk premium ranges from nearly 0 to about 5 in the U.S. In contrast, the range of the Japanese stock market is nearly doubled, that is, from nearly 0 to 10. Similarly, in the U.K., it ranges from 0 to 9. Its difference in the risk premium at the 5% and 25% quantiles is much larger than in other countries.

We also explored whether the risk premium is sensitive to major economic events, such as the 2007-2008 financial crisis and the 2009 European debt crisis. Figure 2 indicates that the SMB and RMW factors in the U.S., the Mkt, SMB, HML and RMW factors in Japan, and the SMB, RMW and CMA factors in the U.K. exhibit larger changes compared to the other factors. These factors rise around 2007-2009, indicating that investors require larger risk premia on these factors in a risk-averse market environment. We note that the risk premia on the SMB and RMW factors increased in each country during this time period. Thus, these factors are strongly related to investors' risk-aversion attitude. This result is consistent with that in

Liew and Vassalou (2000), which noted that style portfolios would capture certain aspects of the business cycle risk.

4.3.2 Comparison over quantiles

Figure 3 shows how the total sum of estimated risk premia of Fama-French's 5 factors vary over quantiles. We can see that the sum of the risk premia of Fama-French's 5 factors at the 5% quantile in the U.S. remains at a lower level throughout the period compared to those of Japan and the U.K.. In terms of the share of a risk premium of an individual factor to the total risk premium, after the middle of 2011, the share of CMA became relatively high in the U.S. Although HML, Mkt and RMW occasionally increase, SMB continues to be at a low level throughout the period. The total risk premium of the Fama-French 5 factors in Japan and the U.K. is much higher than in the U.S. It ranges from 5 to 10 throughout the time period. Similar to the U.S., the risk premium of the CMA factor occupies a large proportion in the total risk premium. We can see that RMW and HML occasionally increase.

In terms of the total risk premium at the 25% quantile, it remains at a level of approximately 2 or less in the U.S. market. This level is much smaller than the total risk premium at the 5% quantile. The shares of HML and CMA in the total risk premium are higher than the others. This observation is similar to those at 5% quantile except that the proportion of Mkt is relatively large. In Japan, the total risk premium of Fama-French's 5 factors remains at a higher level. We can also see that the risk premium of the CMA, RMW and HML factors are relatively large compared to the other two factors. In the U.K., the overall level drops more than in Japan. Additionally, HML and SMB become increasingly important in addition to CMA and RMW in the total risk premium.

At the 50% quantile, the total risk premium in the U.S. continues to be lower than those in Japan and the U.K. In the U.S., the overall trend of the share of a risk premium of individual factors to the total risk premium is similar to those of the

25th percentile point. The share of risk premium of CMA becomes smaller compared to that of Mkt. In Japan, the total risk premium becomes much smaller compared with those of the 25% quantiles during some of the time period. In the U.K., the risk premium level becomes smaller in 2007–2008 and 2011–2012 compared with the other periods. The risk premium for the 75% and 95% quantiles continues to exhibit similar changes to those for the 25% to 50% quantiles. The main difference is that CMA is smaller during 2009-2010 in Japan and the U.K..

4.3.3 Estimation of risk premium by Fama-MacBeth regression

In this section, we compare our estimation result with that based on Fama-Macbeth's approach. We considered two versions of Fama-Macbeth's approach. The first version directly applies Fama-Macbeth's procedure to the Fama-French 5 factors. In the second version, the first stage applies the method in Section 2.1 and creates the factor returns for each quantile. Then, we apply Fama-Macbeth's procedure to these quantile factors. Similar to the previous section, returns of the past 250 days are used for the estimation.

Figure 4 shows the estimated risk premium based on the first approach. The crucial difference between Fama-Macbeth's method and our proposed method is that the former may obtain negative estimates of the risk premium. We can see that estimated risk premium of some factors has been negative for a long period of time. In the U.S., for example, the estimated risk premium for the Mkt factor is negative for the second half of 2008 through the first half of 2009 and for the year 2015-2016. For the RMW and CMA factors, the length of periods of positive risk premium is comparable to that of the negative period. Notably, the SMB and HML factors have a long negative period. Similarly, the estimated risk premia in Japan and the U.K. can take negative values. In contrast, our approach ensures a positive risk premium.

We also apply Fama-Macbeth's approach to the factor returns for each quantile by using the estimated beta per quantile obtained. As in the analysis in Section 4.2, the period used for estimation is the past 250 days up to the end of every month. Then, we roll the estimation period every month. Figure 5 shows the estimated risk premium based on Fama-Macbeth's approach for the U.S., Japan and the U.K. markets. We can see that the estimated risk premium may take negative values. Additionally, the monotonicity of the risk premium over quantile does not hold; the risk premium at a lower quantile (higher risk) is sometimes smaller than that at a higher quantile (low risk). This is simply because there is no constraint on the risk premium between the quantiles during the estimation process.

When we check the signs of the estimated risk premium, it can be seen that the signs are approximately in agreement irrespective of the quantile. In addition, the sign of the risk premium continues for a certain period. Then, the positive risk premium and the negative risk premium are alternately repeated. Needless to say, in all three countries, the magnitude of the estimated risk premium from our method and those from Fama-MacBeth's method are different.

4.3.4 Discussion

Our proposed method provides a useful tool for practitioners. For example, by monitoring the movement of the risk premium on the Fama-French five factors, we can see what kind of risk premium is required by investors and how much it varies over quantiles. Additionally, paying attention to the movement of the lower quantile point may be useful for detecting anomalies such as sudden changes in the market.

The proposed method is also useful for formulating an investment strategy. When we construct a portfolio, it is common to pay attention only to the exposure to the style factors. However, the risk premium of the factor is regarded as constant, regardless of the quantile. In contrast, our method suggests that the risk premia may vary over quantiles. By adopting the proposed method, there is a possibility of constructing a portfolio that can expect a more precise acquisition of the risk premium.

5 Empirical Results 2: Impact of Passive Funds on the Risk Premium

According to Morningstar's 2017 report, in the U.S., the financial market experienced capital inflows of $691.6 billion to passive funds, while active funds saw capital outflows of $7.0 billion. The year 2017 is no exception. Migration from active funds to passive funds has been a long trend since 2006. The trend may be explained by a low management cost of passive funds, as well as the difficulty in finding skilled active fund managers. The difficulty in finding a skilled fund manager is related to the debate on whether active funds have generated excess returns. This topic has been widely discussed by academics for a long time (see, e.g., Jensen (1969), Brinson et al. (1986), Fama and French (2010), Cremers et al. (2016) and Crane and Crotty (2018)).

In this section, carrying over the empirical results from Section 4, we further explore the impact of passive funds on the risk premium. More specifically, we first study the relationship between the Mkt factor and liquidity and show that these two measures are related. As the next step, the impact of cash flow to passive funds on the risk premium on the Mkt factor will be investigated.

5.1 Data: flows into passive funds

To quantify the impact of flows into passive funds, we study the passive funds linked to the S&P 500, TOPIX, and the FTSE All-Share Index in each country. Appendix B provides the details of the data acquisition process from Bloomberg, including how to create a list of funds. For each of the mutual funds, we define the time series named "Flow_t" to measure the liquidity as follows:

$$\text{Flow}_t = \left(\frac{\text{TNA}_t}{\text{NAV}_t} - \frac{\text{TNA}_{t-1}}{\text{NAV}_{t-1}} \right) \times \text{NAV}_t,$$

where TNA_t is the total net asset at period t and NAV_t is net asset value at period t. When the number of outstanding shares is missing, we estimate it by linear interpolation. As a result, the number of mutual funds and the exchange-traded funds (ETFs) that track the S&P 500 were 61 and 14, respectively. The number of mutual funds and ETFs for TOPIX were 61 and 6, and those for the FTSE All-Share Index were 6 and 3, respectively.

Figure 6 summarizes the daily next flow of passive funds linked to the three indices of S&P500, TOPIX, and FTSE All-Share every year. The net flow is normalized by the trading value of the stock market. It is worth noting that the net flow is normalized by the total of the transaction (sell and buy) at the stock market. Therefore, the normalized share becomes larger when the net transaction at the stock market is employed. We can see that the inflow to the passive fund trading has been positive in all three countries. It is natural to expect that the impact of this net flow on the stock price is not negligible.

5.2 Empirical results

5.2.1 The relationship between the Mkt factor and the liquidity factor

We first study the relationship between the Mkt factor and the liquidity factor. Because we study market-weighted passive funds that are representative of each country's market, it is most likely that Mkt receives a large impact from in/out flows to passive funds. The correlation between the exposure to Mkt for each quantile and the liquidity is calculated over time. We carried out the same analysis for the other Fama-French factors. However, no clear relationships are observed for the other 4 Fama-French factors.

Figure 7 shows the historical correlation between the Mkt factor and the liquidity factor for each of the quantiles. In the U.S. market, the correlation is largely positive at the 95th and 75th percentile points. On the other hand, the correlation is largely negative at low quantiles such as the 5th percentile point and the 25th percentile point. This implies that liquidity contributes to the large fluctuations in the stock market. Similar observations can be made for the Japanese and the U.K. stock markets. We can also observe that the magnitude of correlation of the U.K. is relatively smaller than that of the U.S. and Japanese markets. In summary, the analysis reveals that there is a strong correlation between the Mkt factor and the liquidity factor. In the next section, we further examine whether the passive flow is affecting the risk premium on the Mkt factor.

5.2.2 The relationship between the market risk premium and passive flow

We study the correlation between the market risk premium on the Mkt factor and the flows to passive funds. Figure 8 shows the calculated correlation from 2008 April to 2017 December for each of the three markets.

For the U.S. market, we can see relatively high correlation from 2007 to 2010. This period coincides with the term when the inflow to the passive funds accounted for a large share of the market inflow. Notably, there is no significant difference in the correlation between quantiles in this period. From these results, it is likely that the increase in inflow to passive funds equally affected both high to low quantiles. From 2011 to the end of 2016, the correlation was around 0. We note that that the magnitude of the correlation is different between the 5th percentile point and the other quantile points. For example, from October 2012 to June 2013, the correlation with respect to the 5 percentile points behaves differently from those of the other quantiles. This indicates that some other factors in addition to the inflows to passive fund factors are important when the largely negative stock return is observed.

Compared to the U.S. market, the correlation for the Japanese market behaves differently. The magnitude of correlation varies over quantiles. For example, in 2010 and 2015-16, a largely positive correlation is observed at the 5th percentile point only. Thus, the correlation with the passive flow is high when stock market returns are low. In April 2013, the Bank of Japan announced quantitative-qualitative easing

(QQE), purchasing financial assets such as government bonds and listing investment funds (ETFs) more than ever. We can see the effect such that the stock price is supported by this policy.

The correlation of the U.K. market also shows different behavior. The magnitudes of correlation at each quantile point are different compared to the U.K. and Japan. We can observe the negative correlation, while a positive is expected if cash flows are affecting market factors. Together with the observations of the U.S. and Japanese markets, the magnitude of the influence of capital flows in the U.K. market factors seems smaller compared with that in the U.S. and Japanese markets.

6 Conclusion

A quantile-based asset pricing model was introduced. The proposed method has several attractive features. First, the method automatically detects the set of necessarily common factors for a working asset pricing model. This is a very important feature because a more (or less) important factor may become less (or more) important when a model is misspecified. Second, the method always ensures that the estimated risk premium is positive. Third, the method can allow for the risk premium to vary over quantiles while keeping economic intuition. Note that the method obtains the constant risk premium when it does not vary over quantiles. This is an attractive feature because, as shown in our empirical analysis, the risk premium varies over quantiles. Fourth, the developed asymptotic theory ensures the consistency and the asymptotic normality of the estimated parameters. Finally, the method transforms a large number of stock returns simultaneously and thus is capable of large-scale data analysis.

To justify our procedure, we further studied the theoretical property of our proposed procedure. Due to the estimation errors in the unobservable common factor structures, we needed to develop the asymptotic results carefully. More specifically, we developed the asymptotic theory of the proposed risk premium estimator by taking into account the effect of estimated factor structure. Under the condition when the time-series dimension and individual dimension are large, we developed a novel strategy for establishing asymptotic theory.

The model and the estimation method are applied to the universe of component stocks from S&P500 in the U.S., TOPIX from Japan, and FTSE All-Share index from the U.K. Based on our approach, our empirical results revealed many interesting findings.

References

Ando, T. and Bai, J. (2017). Clustering huge number of time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, **112**, 1182–1198.

Ando, T. and Bai, J. (2018). Quntile co-movements in financial markets. *Journal of the American Statistical Association*, forthcoming.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, **77**, 1229–1279.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, **70**, 191–221.

Bai, J. and G. Zhou (2015). Fama–MacBeth two–pass regressions: Improving risk premium estimates. *Finance Research Letters*, **15**, 31–40.

Bondell, H. D., Reich, B. J. and Wang, H. (2010) Noncrossing quantile regression curve estimation. *Biometrika*, **97**, 825-838.

Brinson, G. P., Hood, L. R., and Beebower, G. L. (1986). Determinants of portfolio performance. *Financial Analysts Journal*, **42**, 39–44.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica*, **51**, 1305–1324.

Chan, L.K, Karceski , J. and Lakonishok , J. (2003) The level and persistence of growth rates. *Journal of Finance*, **58**, 643–684.

Connor, G. and Korajzcyk, R. (1986). Performance measurement with the arbitrage pricing theory: a new framework for analysis. *Journal of Financial Economics*, **15**, 373–394.

Crane, A. D. and Crotty, K. (2018). Passive versus active fund performance: Do index funds have skill? *Journal of Financial and Quantitative Analysis*, **53**, 33–64.

Cremers, M., Ferreira, M. A., Matos, P., and Starks, L. (2016). Indexing and active fund management: International evidence. *Journal of Financial Economics*, **120**, 539–560.

Fama, E. F., and French, K. R. (2015). Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance*, **65**, 1915–1947.

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

Fama, E. F., and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical test. *Journal of Political Economy*, **71**, 607—636.

Feng, H., Giglio, S., and Xiu, D. (2017). Taming the factor zoo. *Working Paper*, University of Chicago, Booth School of Business.

Ferson, W. E. and C. R. Harvey (1991). The variation of economic risk premiums. *Journal of Political Economy*, **99**, 385–415.

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

Giglio, S., and Xiu, D. (2017). Inference on risk premia in the presence of omitted factors. *Working Paper*, University of Chicago, Booth School of Business.

Gospodinov, N., R. Kan, and C. Robotti (2014). Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors. *Review of Financial Studies*, **27**, 2139–2170.

Harvey, C. R., Liu, Y., and Zhu, H. (2015). ... and the cross-section of expected returns. *Review of Financial Studies*, **29**, 5–68.

Ippolito, R. A. (1989). Efficiency with costly information: A study of mutual fund performance, 1965–1984. *Quarterly Journal of Economics*, **104**, 1–23.

Jagannathan, R. and Z. Wang (1998). An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *Journal of Finance*, **53**, 1285–1309.

Jensen, M. C. (1969). Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business*, **42**, 167–247.

Kan, R. and C. Zhang (1999). Two-Pass Tests of Asset Pricing Models. *Journal of Finance*, **54**, 203–235.

Kan, R., and Robotti, C. (2012). Evaluation of asset pricing models using two-pass cross-sectional regression. *Handbook of Computational Finance*, edt by Duan, J. C. et al, pp223–251.

Kan, R., Robotti, C. and Shanken, J. (2013). Pricing model performance of the two-pass cross-sectional regression methodology. *Journal Finance*, **63**, 223–251.

Kleibergen, F. and Z. Zhan (2018). Identification-robust inference on risk premia of mimicking portfolios of non-traded factors. *Journal of Financial Econometrics*, **16**, 155–190.

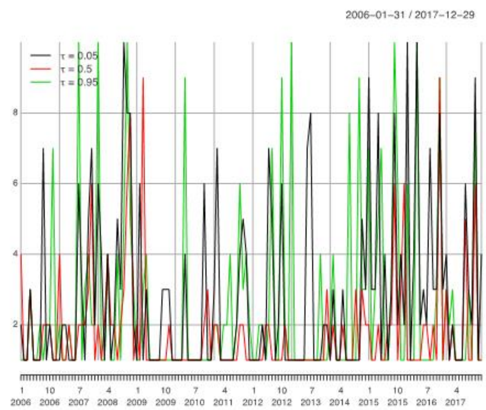
Liew, J., & Vassalou , M. (2000). Can book-to-market, size and momentum be risk factors that predict economic growth?. *Journal of Financial Economics*, **57**, 221–245.

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

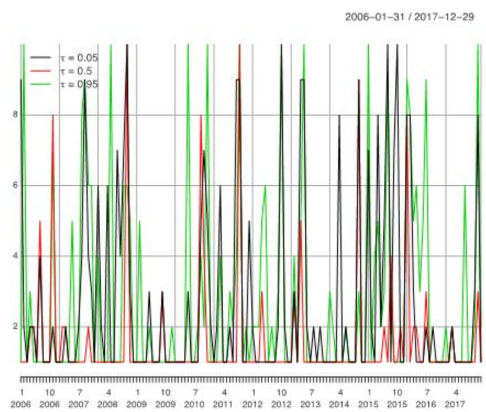
Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economics Theory*, **13**, 341–360.

Shanken, J. (1992). On the Estimation of Beta Pricing Models. *Review of Financial Studies*, **5**, 1-

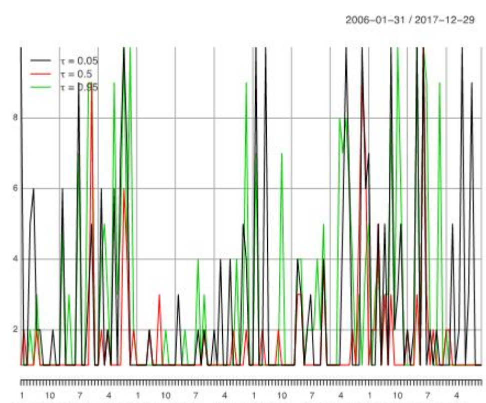
Shanken, J. and G. Zhou (2007). Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics*, **84**, 40–86.



(a) U.S.

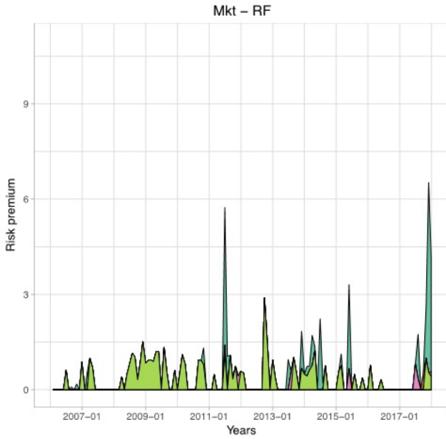


(b) Japan

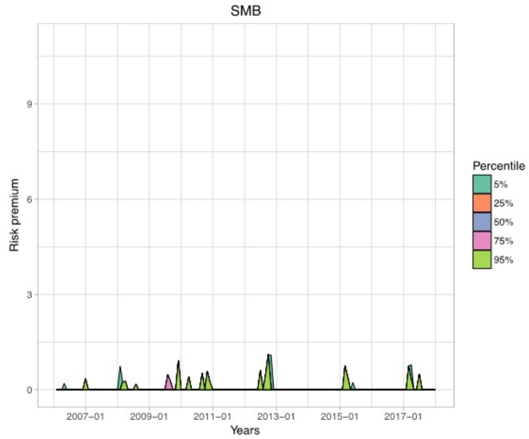


(c) UK

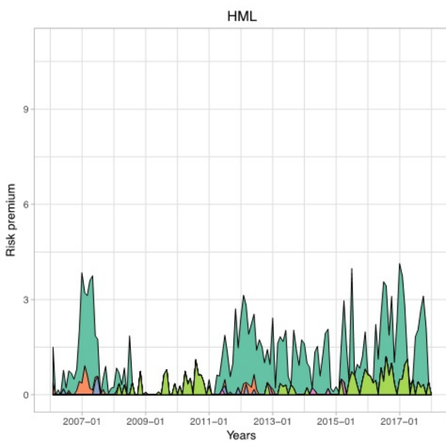
Figure 1: Selected number of factors at quantiles 5%, 50% and 95%, respectively.



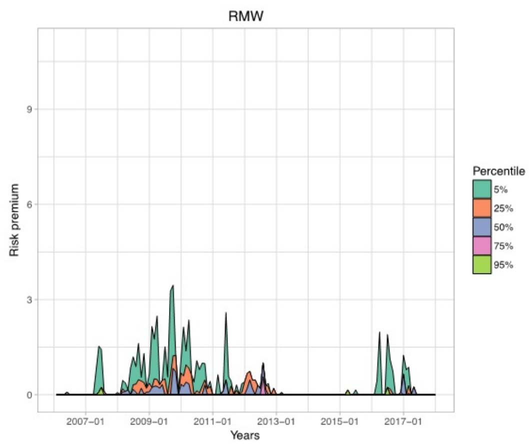
(a) SP500 Mkt



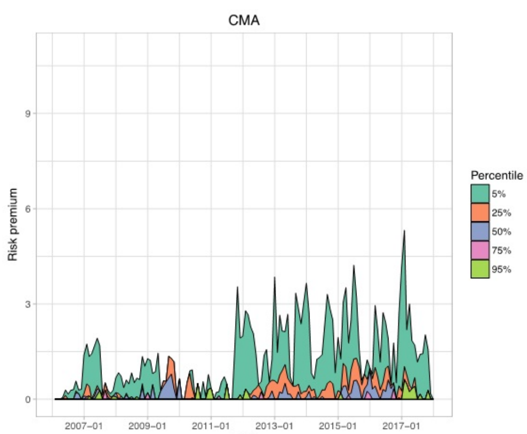
(b) SP500 SMB



(c) SP500 HML

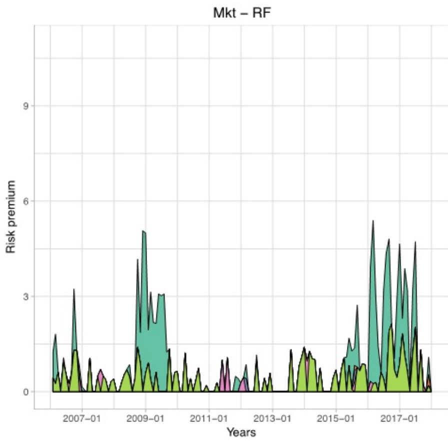


(d) SP500 RMW

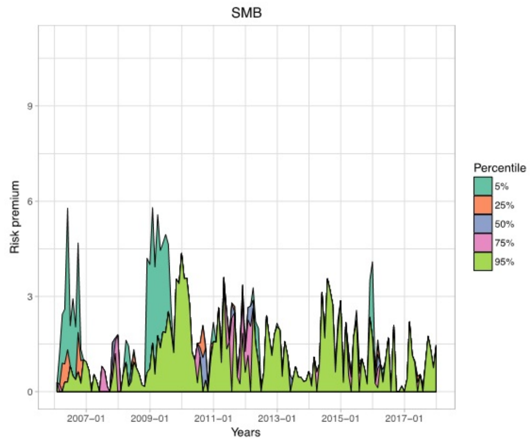


(e) SP500 CMA

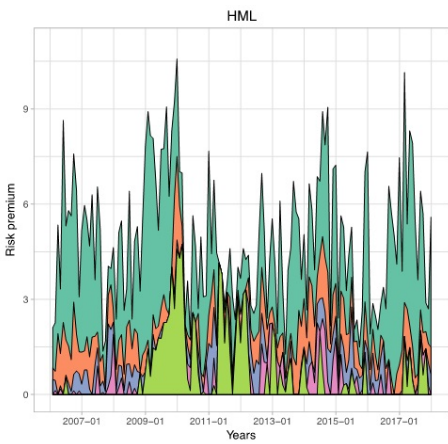
Figure 2: S&P500: Quantified price of risk at quantiles 5%, 25% 50%, 75% and 95%, respectively.



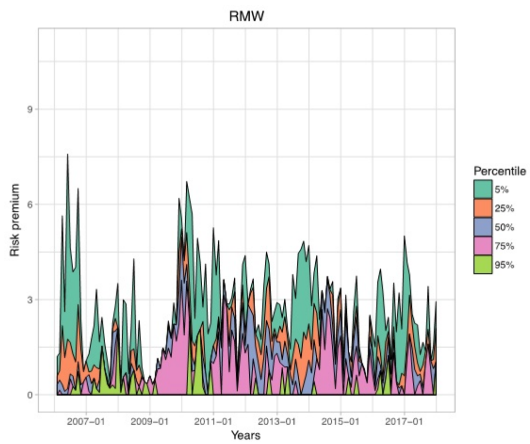
(a) TOPIX Mkt



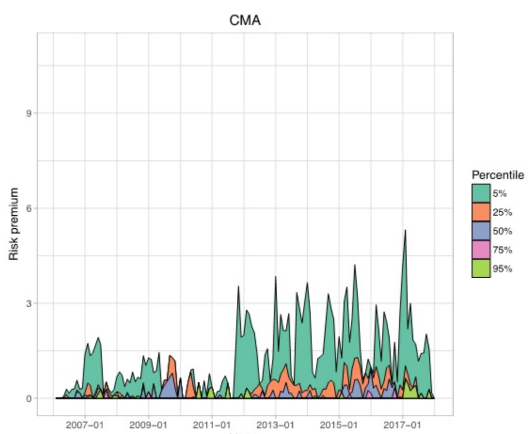
(b) TOPIX SMB



(c) TOPIX HML

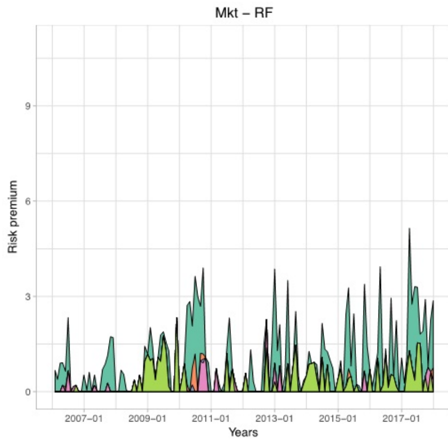


(d) TOPIX RMW

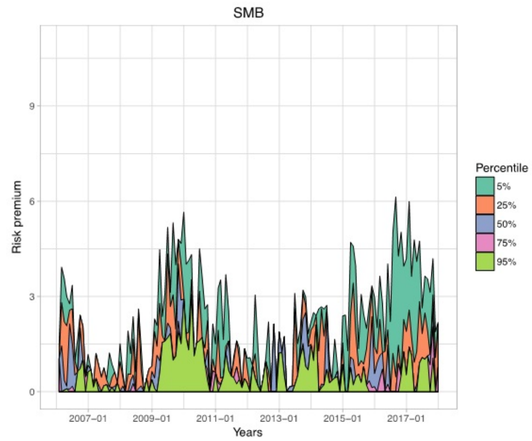


(e) TOPIX CMA

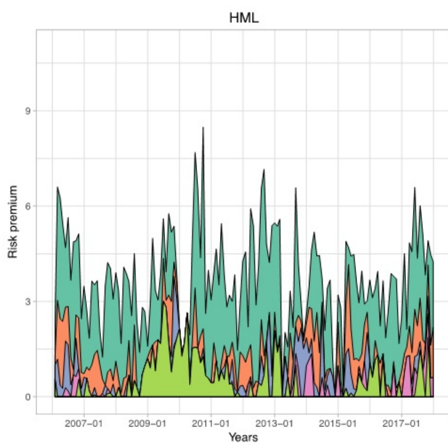
Figure 2: (Continued) TOPIX: Quantified price of risk at quantiles 5%, 25% 50%, 75% and 95%, respectively.



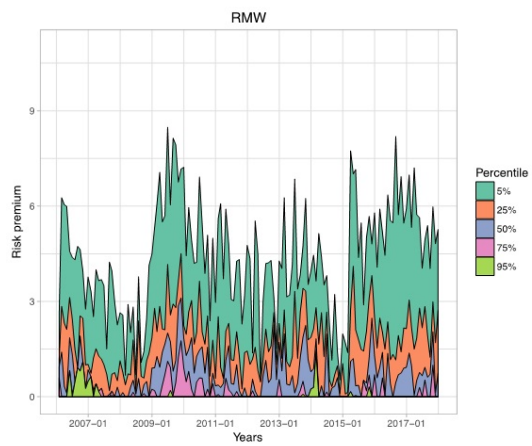
(a) FTSE Mkt



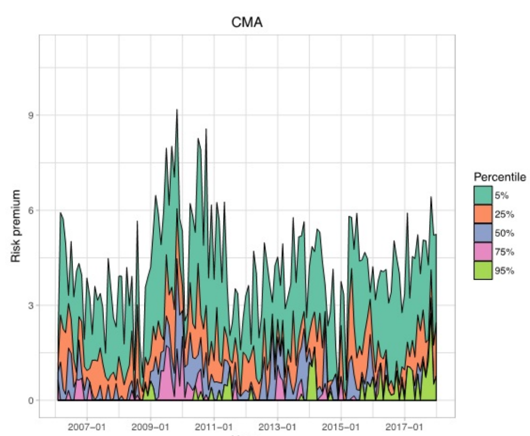
(b) FTSE SMB



(c) FTSE HML



(d) FTSE RMW



(e) FTSE CMA

Figure 2: (Continued) FTSE: Quantified price of risk at quantiles 5%, 25% 50%, 75% and 95%, respectively.

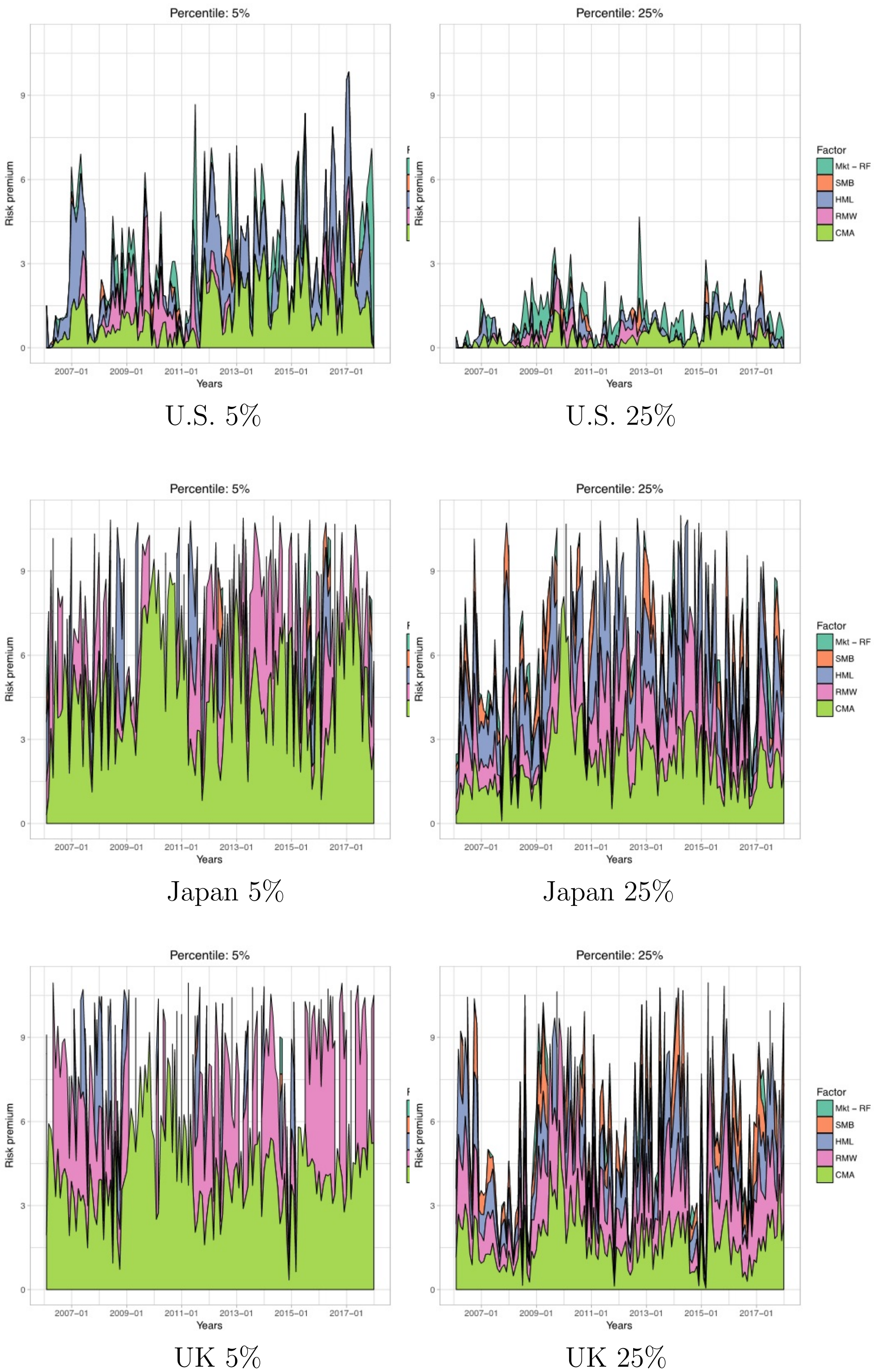
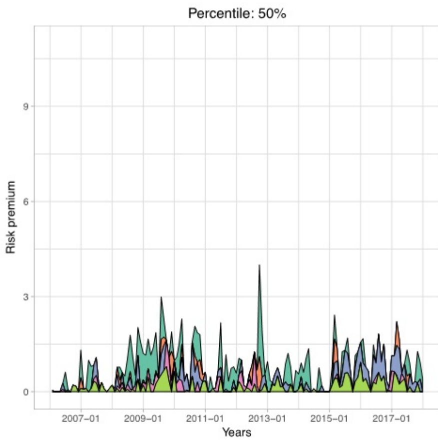
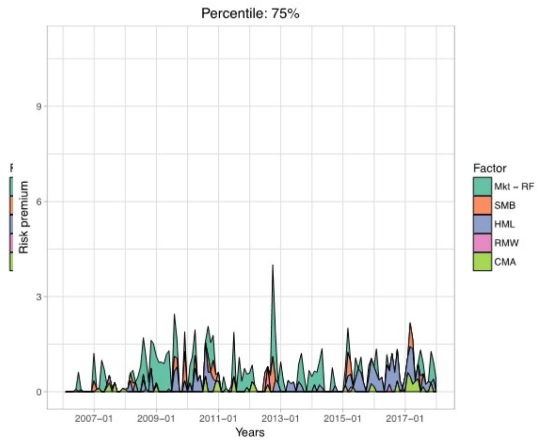


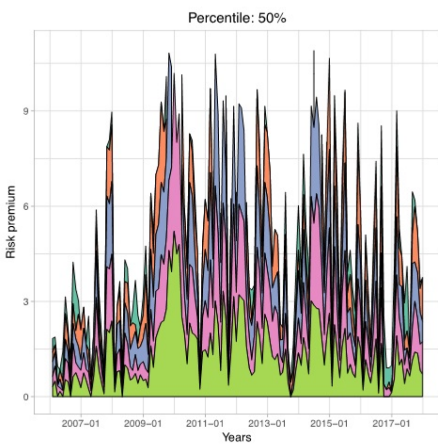
Figure 3: Total sum of risk premium of Fama-French 5 factors.



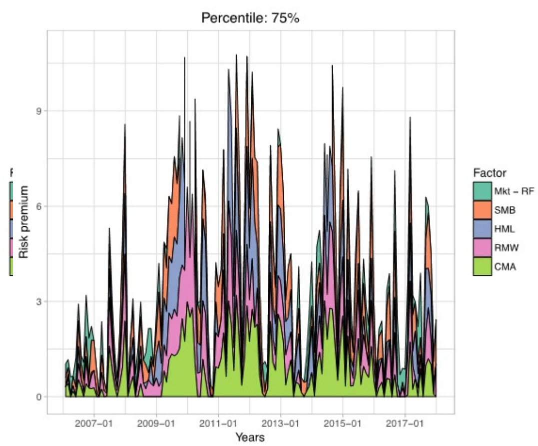
U.S. 50%



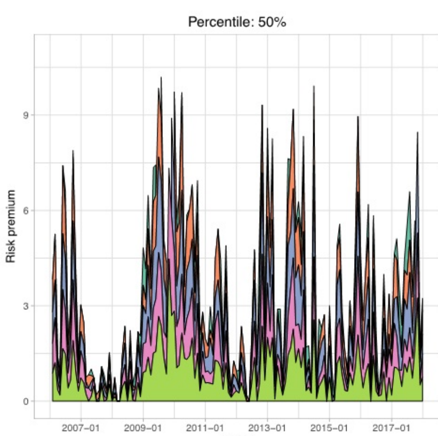
U.S. 75%



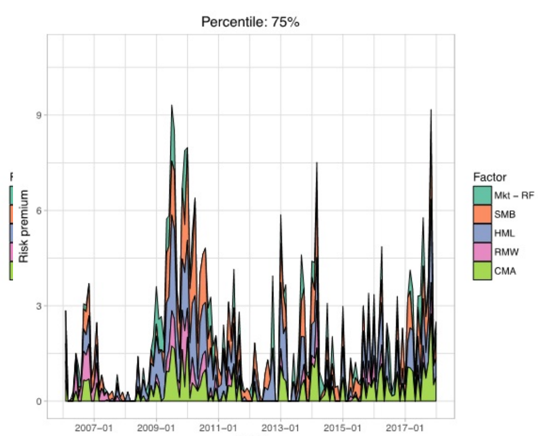
Japan 50%



Japan 75%

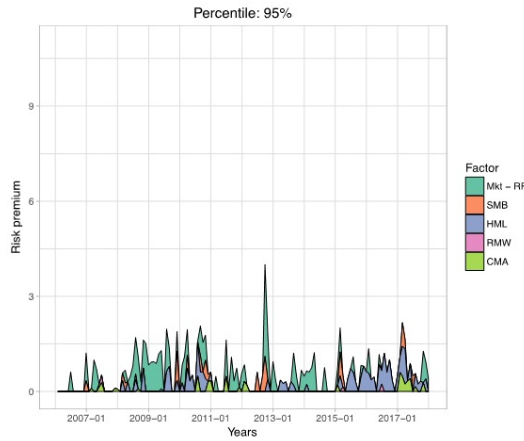


UK 50%

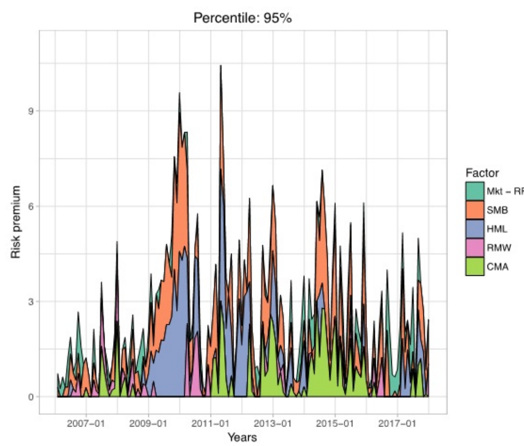


UK 75%

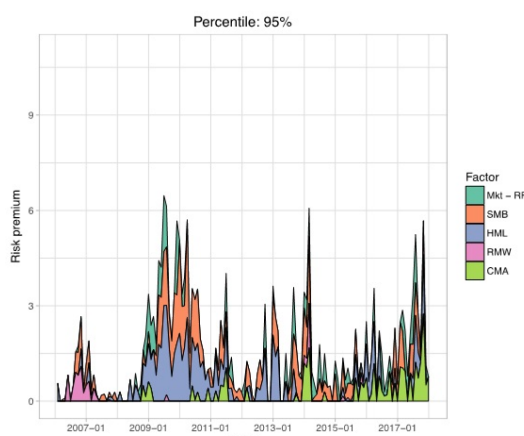
Figure 3: (Continued). Total sum of risk premium of Fama-French 5 factors.



U.S. 95%

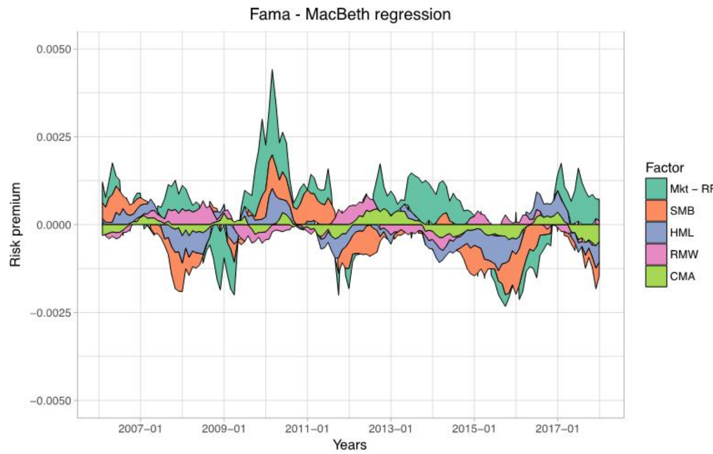


Japan 95%

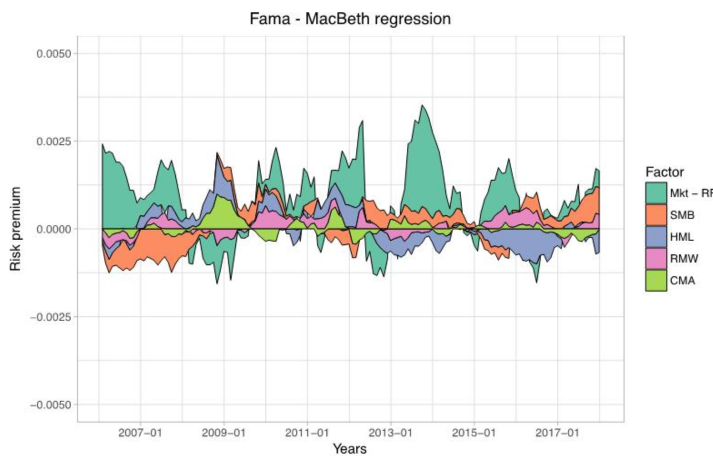


UK 95%

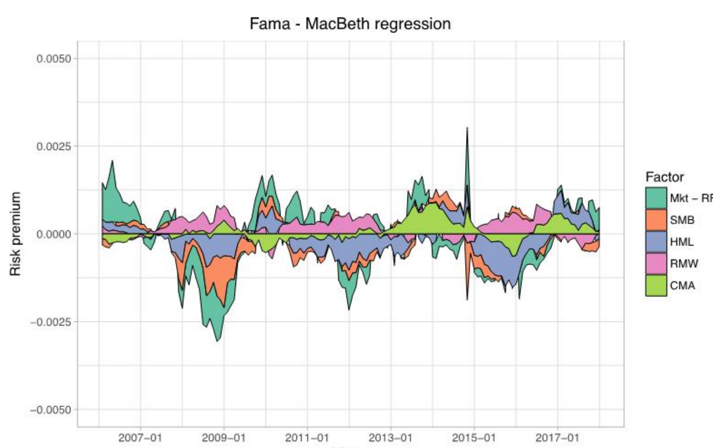
Figure 3: (Continued). Total sum of risk premium of Fama-French 5 factors.



(a) SP500

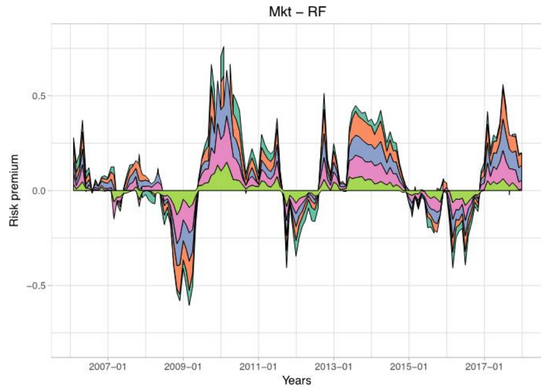


(b) TOPIX

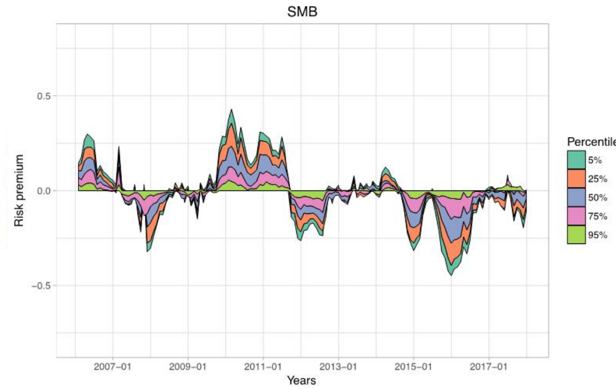


(c) FTSE

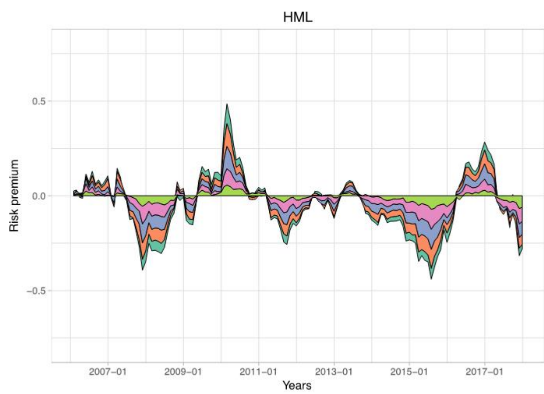
Figure 4: Quantified price of risk of Fama and French 5 factors. The results are obtained based on Fama-Macbeth (1973)'s approach.



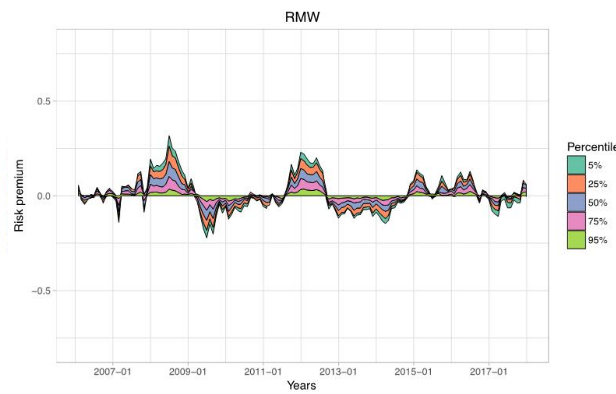
(a) SP500 Mkt



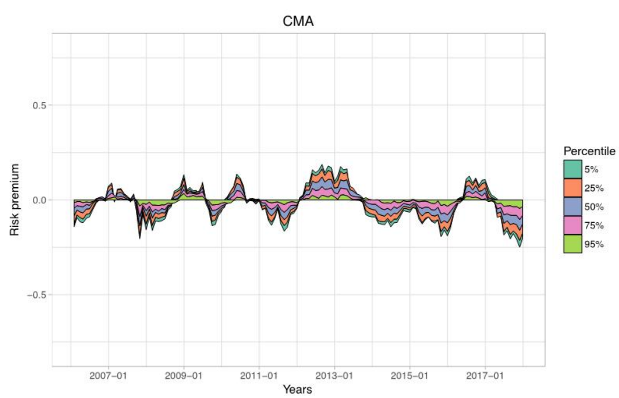
(b) SP500 SMB



(c) SP500 HML

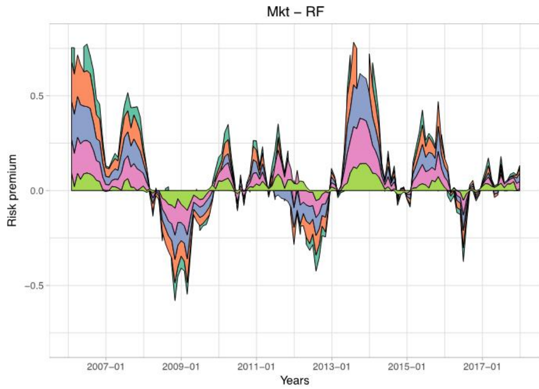


(d) SP500 RMW

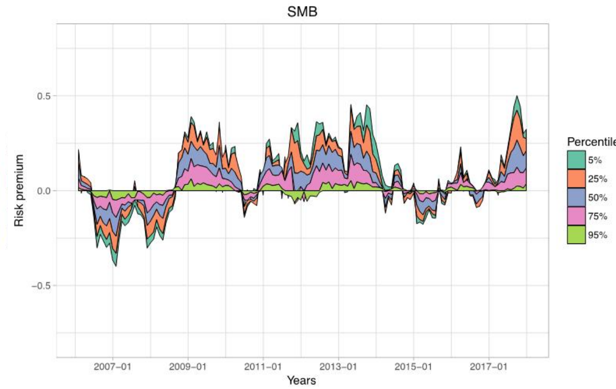


(e) SP500 CMA

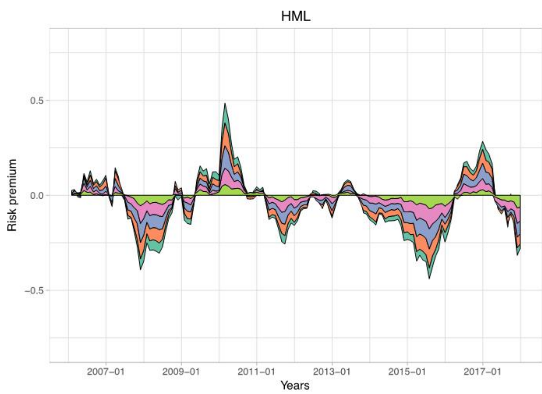
Figure 5: S&P500: Quantified price of risk at quantiles 5%, 25% 50%, 75% and 95%, respectively. The results are obtained based on Fama-Macbeth (1973)'s approach.



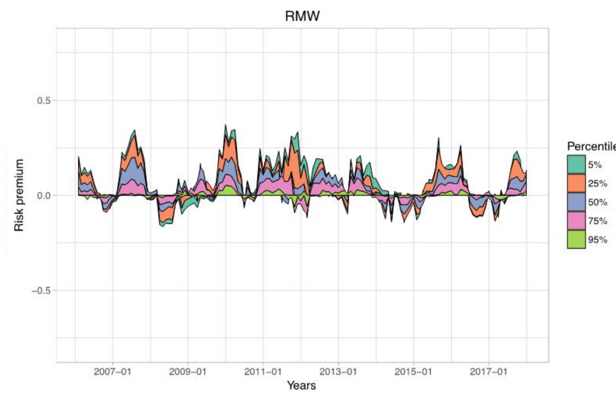
(a) TOPIX Mkt



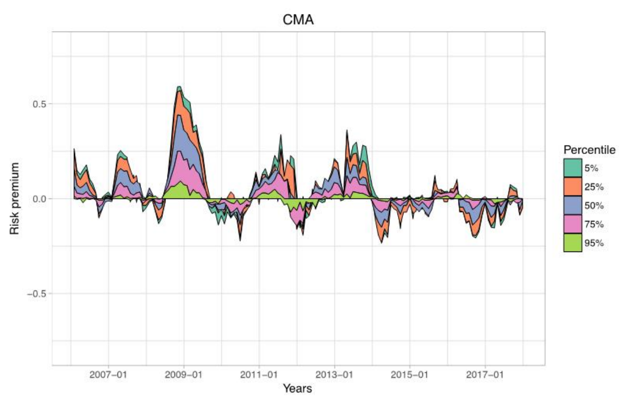
(b) TOPIX SMB



(c) TOPIX HML

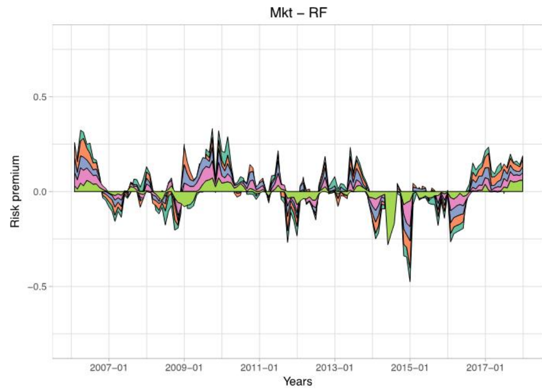


(d) TOPIX RMW

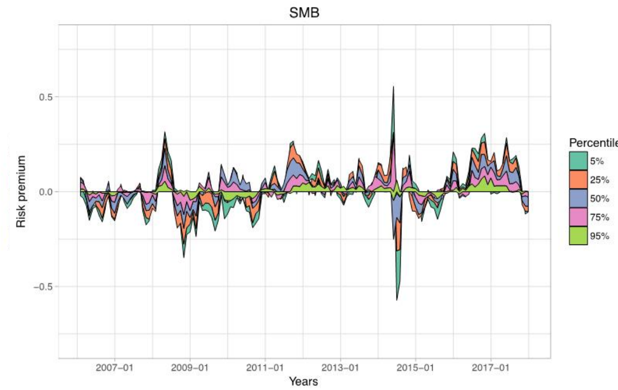


(e) TOPIX CMA

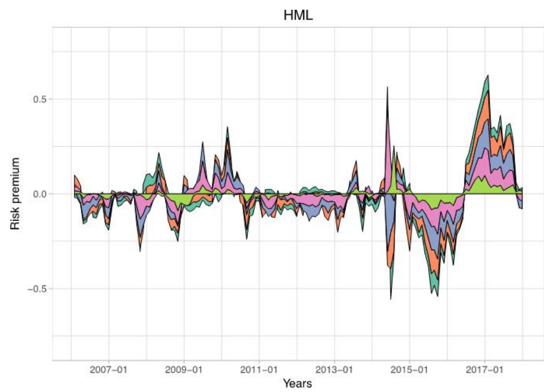
Figure 5: (Continued) TOPIX: Quantified price of risk at quantiles 5%, 25% 50%, 75% and 95%, respectively. The results are obtained based on Fama-Macbeth (1973)'s approach.



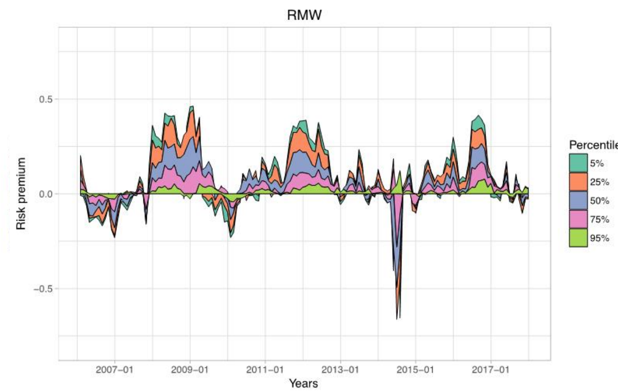
(a) FTSE Mkt



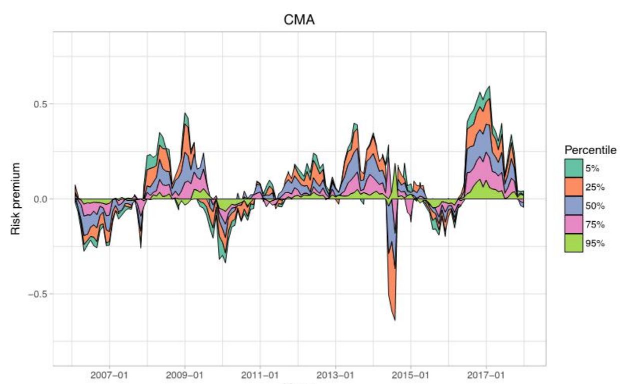
(b) FTSE SMB



(c) FTSE HML



(d) FTSE RMW



(e) FTSE CMA

Figure 5: (Continued) FTSE: Quantified price of risk at quantiles 5%, 25% 50%, 75% and 95%, respectively. The results are obtained based on Fama-Macbeth (1973)'s approach.

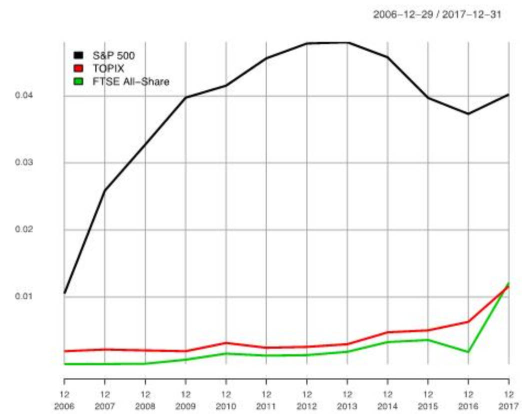
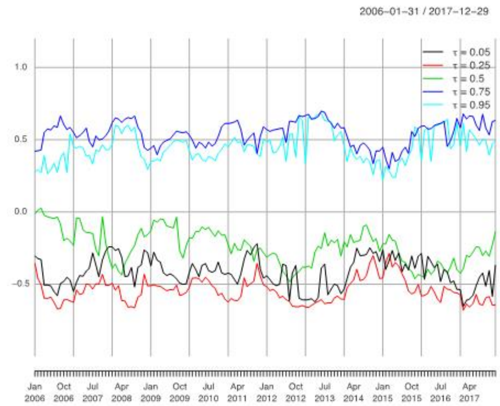
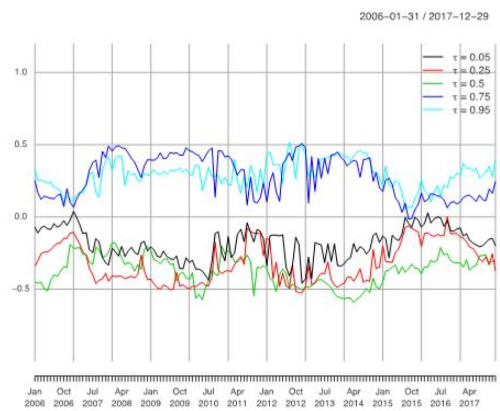


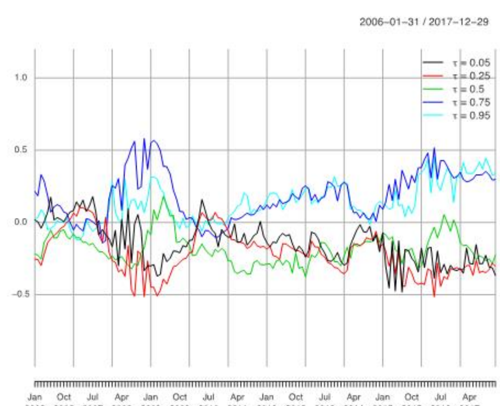
Figure 6: Capital inflow and outflow of passive funds.



(a) U.S.



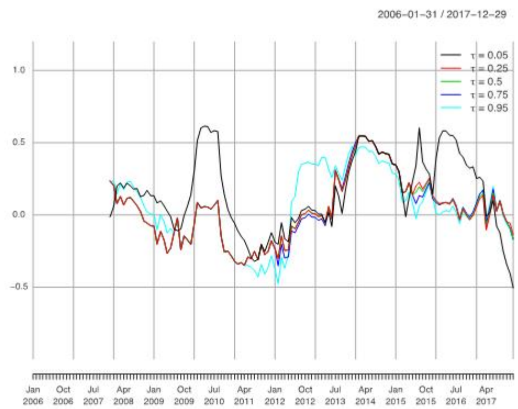
(b) Japan



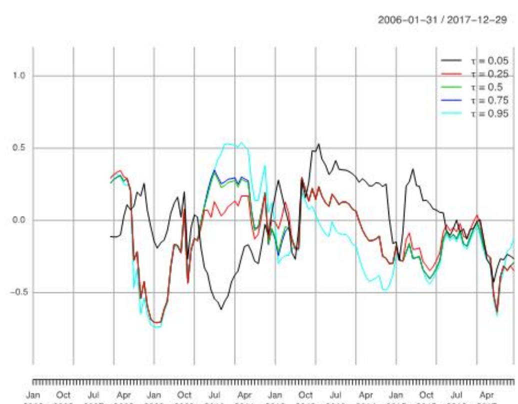
(c) UK

Figure 7: Correlation between the Mkt factor and the liquidity factor.

(a) U.S.



(b) Japan



(c) U.K.

Figure 8: Time series plot of the correlation between the Mkr risk premium and the capital inflow to passive funds.

Year	S&P 500 5%	25%	50%	75%	95%
2005	-0.022	-0.008	0.000	0.008	0.024
2006	-0.023	-0.007	0.000	0.008	0.025
2007	-0.029	-0.009	0.000	0.009	0.028
2008	-0.065	-0.020	-0.001	0.016	0.062
2009	-0.050	-0.014	0.001	0.016	0.055
2010	-0.030	-0.009	0.001	0.010	0.032
2011	-0.035	-0.011	0.000	0.011	0.035
2012	-0.024	-0.008	0.000	0.009	0.026
2013	-0.020	-0.006	0.001	0.009	0.023
2014	-0.021	-0.006	0.001	0.008	0.021
2015	-0.027	-0.009	0.000	0.009	0.026
2016	-0.027	-0.007	0.001	0.009	0.028
2017	-0.019	-0.005	0.001	0.007	0.020
	TOPIX				
Year	5%	25%	50%	75%	95%
2005	-0.024	-0.008	0.000	0.010	0.032
2006	-0.034	-0.012	0.000	0.011	0.035
2007	-0.033	-0.012	0.000	0.010	0.033
2008	-0.059	-0.020	-0.001	0.016	0.056
2009	-0.041	-0.014	0.000	0.013	0.045
2010	-0.031	-0.010	0.000	0.010	0.032
2011	-0.034	-0.011	0.000	0.011	0.036
2012	-0.029	-0.010	0.000	0.010	0.033
2013	-0.034	-0.010	0.000	0.013	0.040
2014	-0.029	-0.009	0.000	0.010	0.031
2015	-0.030	-0.009	0.000	0.010	0.032
2016	-0.037	-0.011	0.000	0.011	0.038
2017	-0.022	-0.007	0.000	0.008	0.027
	ASX				
Year	5%	25%	50%	75%	95%
2005	-0.021	-0.005	0.000	0.006	0.024
2006	-0.026	-0.007	0.000	0.008	0.028
2007	-0.032	-0.009	0.000	0.009	0.031
2008	-0.058	-0.018	-0.001	0.014	0.053
2009	-0.043	-0.012	0.000	0.015	0.050
2010	-0.031	-0.009	0.000	0.010	0.033
2011	-0.034	-0.011	0.000	0.010	0.033
2012	-0.027	-0.008	0.000	0.009	0.030
2013	-0.025	-0.007	0.000	0.009	0.027
2014	-0.025	-0.007	0.000	0.007	0.025
2015	-0.025	-0.007	0.000	0.008	0.027
2016	-0.029	-0.008	0.000	0.009	0.032
2017	-0.021	-0.006	0.000	0.007	0.023

Table 1: Summary statistics

A Assumptions and Technical Details

For clarity, we state the assumptions on the quantile function in (1) introduced in the main text.

A.1 Assumptions

Assumption A: Common factors

Let $\mathcal{F}$ be a compact subset of R^{r_τ}. The unobservable common factors $\boldsymbol{f}_{t,0,\tau} \in \mathcal{F}$ satisfy $T^{-1} \sum_{t=1}^{T} \boldsymbol{f}_{t,0,\tau} \boldsymbol{f}_{t,0,\tau}{}' \to \Sigma_{F_\tau}$ as $T \to \infty$, where Σ_{F_τ} is an $r_\tau \times r_\tau$ positive definite matrix.

Assumption B: Factor loadings and regression coefficients

Let $\mathcal{B}$ and $\mathcal{L}$ be compact subsets R^{p+1} and R^{r_τ}. The regression coefficient $\boldsymbol{b}_{i,0,\tau}$ and the factor-loading for the common factors satisfy $\boldsymbol{b}_{i,0,\tau} \in \mathcal{B}$ and $\boldsymbol{\lambda}_{i,0,\tau} \in \mathcal{L}$. In addition, the factor-loading matrix $\Lambda_{0,\tau} = (\boldsymbol{\lambda}_{1,0,\tau}, \ldots, \boldsymbol{\lambda}_{N,0,\tau})'$ satisfies $N^{-1}\Lambda_{0,\tau}'\Lambda_{0,\tau}$ being a $r_\tau \times r_\tau$ positive definite matrix for all N. Also, the matrix $N^{-1} \sum_{i=1}^{N} \boldsymbol{b}_{i,0,\tau}\boldsymbol{b}_{i,0,\tau}'$ is positive definite for all $\tau \in (0, 1)$.

Assumption C: Idiosyncratic error terms

(C1): The random variable $\varepsilon_{it,\tau} = y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0,\tau} - \boldsymbol{f}_{t,0,\tau}'\boldsymbol{\lambda}_{i,0,\tau}$ is independently distributed over i and t, conditional on X, $F_{0,\tau}$ and $\Lambda_{0,\tau}$. In addition, it satisfies $E\left[|\varepsilon_{it,\tau} - E[\varepsilon_{it,\tau}]|^K\right] < K!C_\varepsilon^K$ for $K \geq 1$ and a positive constant $C_\varepsilon < \infty$.

(C2): The conditional density function of $\varepsilon_{it,\tau}$ given $\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0,\tau}, \boldsymbol{\lambda}_{i,0,\tau}$, denoted as $g_{it}\left(\varepsilon_{it,\tau}|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0,\tau}, \boldsymbol{\lambda}_{i,0,\tau}\right)$, is continuous. In addition, for any compact set $\mathcal{C}$, there exists a positive constant $\bar{g} > 0$ (depending on $\mathcal{C}$) such that $\inf_{c \in \mathcal{C}} g_{it}(c|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0,\tau}, \boldsymbol{\lambda}_{i,0,\tau}) \geq \bar{g}$ for all i and t.

Assumption D: Predictors and design matrix

(D1): For a positive constant C_x, predictors satisfy $\sup_{it} \|\boldsymbol{x}_{it}\| < C_x < \infty$.

(D2): There exist positive constants C_1 and C_2 such that for each X_i,

$$0 < C_1 < \lambda_{\min}(T^{-1}(X_i, F_{0,\tau})'(X_i, F_{0,\tau})) < \lambda_{\max}(T^{-1}(X_i, F_{0,\tau})'(X_i, F_{0,\tau})) < C_2 < \infty,$$

where $X_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iT})'$, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and the largest eigenvalue of a matrix A, respectively. These inequalities hold with probability approaching 1 as $T \to \infty$.

(D3): Define $A_{i,\tau} = \frac{1}{T} X_i' M_{F_\tau} X_i$, $B_{i,\tau} = (\boldsymbol{\lambda}_{i,0,\tau} \boldsymbol{\lambda}_{i,0,\tau}') \otimes I_T$, $C_{i,\tau}' = \frac{1}{\sqrt{T}} \boldsymbol{\lambda}_{i,0,\tau}' \otimes (X_i' M_{F_\tau})$, $M_{F_\tau} = I - F_\tau (F_\tau' F_\tau)^{-1} F_\tau'$. Let $\mathcal{F}_\tau$ be the collection of F_τ such that $\mathcal{F}_\tau = \{F_\tau : F_\tau' F_\tau / T = I\}$. We assume

$$\inf_{F_\tau \in \mathcal{F}_\tau} \lambda_{\min} \left[\frac{1}{N} \sum_{i=1}^{N} E_{i,\tau}(F_\tau) \right] > 0,$$

where $E_{i,\tau}(F_\tau) = B_{i,\tau} - C_{i,\tau}' A_{i,\tau}^{-1} C_{i,\tau}$ and inf is taken under the fixed τ which is the focus.

Assumption E: Restrictions on N, T

N and T in Step 1 satisfy $T^{1/2}/N^{1-\gamma} \to 0$ and $N^{1/2}/T^{1-\gamma} \to 0$ for a small γ satisfying $1/16 < \gamma$.

Assumption F: Restrictions on τ_k

$N^{-1/2} \min_k (\tau_{k+1} - \tau_k) \to \infty$.

Remark 1 Assumptions A $\sim$ E are taken from Ando and Bai (2018). The full rank assumption in Assumptions A and B is imposed to ensure the number of common factors being r_τ. In Assumption C, we impose some mild conditions on the idiosyncratic errors. As given in Assumption D, we need to impose the regularity condition on design matrix $\mathrm{X_i}$ and common factor structure $F_{0,\tau}$. the usual rank condition is used for identification in (D2). (D3) is also imposed to ensure the consistency of the estimated parameters. Assumption E bands the diverging magnitudes of N and T. However, it is not strong assumption. Assumption F is used for Theorem 1 so that the inference for the $N^{1/2}$-consistent constrained estimator.

A.2 Proof of Theorem 1

We use Knight's identity,

$$\rho_\tau(u-\nu)-\rho_\tau(u) = -\nu\psi_\tau(u)+\int_0^\nu (I(u\le s)-I(u\le 0))ds,$$

with $\psi_\tau(u)=\tau-I(u\le 0)$. From Proposition 1, we have

$$T^{1/2}\left(\hat{\boldsymbol{\lambda}}_{i,\tau}-\boldsymbol{\lambda}_{i,0,\tau}\right)\sim N(0,\Sigma_{i,\tau}),\quad \text{and}\quad N^{1/2}\left(\hat{\boldsymbol{f}}_{t,\tau}-\boldsymbol{f}_{t,0,\tau}\right)\sim N(0,\Theta_{t,\tau}),$$

which implies

$$\begin{aligned}\max_i\|\hat{\boldsymbol{\lambda}}_{i,\tau}-\boldsymbol{\lambda}_{i,0,\tau}\| &= O_p(\log(N)/\sqrt{T}),\\ \max_t\|\hat{\boldsymbol{f}}_{t,\tau}-\boldsymbol{f}_{t,0,\tau}\| &= O_p(\log(T)/\sqrt{N}).\end{aligned}$$

For some positive constant C, we thus have

$$\begin{aligned}&\max_i\max_t\left\|\hat{\boldsymbol{f}}'_{t,\tau_k}\hat{\boldsymbol{\lambda}}_{i,\tau_k}-\boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k}\right\|\\ \le\ & C\times\max_t\left\|\hat{\boldsymbol{f}}_{t,\tau_k}-\boldsymbol{f}_{t,0,\tau_k}\right\|+C\times\max_i\left\|\hat{\boldsymbol{\lambda}}_{i,\tau_k}-\boldsymbol{\lambda}_{i,0,\tau_k}\right\|\\ =\ & O_p(\log(N)/\sqrt{T})+O_p(\log(T)/\sqrt{N})\\ =\ & o_p(1).\end{aligned}$$

Using these results, we have

$$\begin{aligned}&\frac{1}{KT}\sum_{k=1}^K\sum_{t=1}^T\rho_{\tau_k}\left(y_{it}-\hat{\boldsymbol{f}}'_{t,\tau_k}\hat{\boldsymbol{\lambda}}_{i,\tau_k}-\boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_k}\right)\\ =\ & \frac{1}{KT}\sum_{k=1}^K\sum_{t=1}^T\rho_{\tau_k}\left(y_{it}-\boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k}-\boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_k}-\left\{\hat{\boldsymbol{f}}'_{t,\tau_k}\hat{\boldsymbol{\lambda}}_{i,\tau_k}-\boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k}\right\}\right)\\ =\ & \frac{1}{KT}\sum_{k=1}^K\sum_{t=1}^T\rho_{\tau_k}\left(y_{it}-\boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k}-\boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_k}\right)\\ & -\frac{1}{KT}\sum_{k=1}^K\sum_{t=1}^T\left\{\hat{\boldsymbol{f}}'_{t,\tau_k}\hat{\boldsymbol{\lambda}}_{i,\tau_k}-\boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k}\right\}\psi_\tau\left(y_{it}-\boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k}-\boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_k}\right)\\ & +\frac{1}{KT}\sum_{k=1}^K\sum_{t=1}^T\int_0^{\hat{\boldsymbol{f}}'_{t,\tau_k}\hat{\boldsymbol{\lambda}}_{i,\tau_k}-\boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k}}\Big(I(y_{it}-\boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k}-\boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_k}\le s)\\ & -I(y_{it}-\boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k}-\boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_k}\le 0)\Big)ds\\ =\ & \frac{1}{KT}\sum_{k=1}^K\sum_{t=1}^T\rho_{\tau_k}\left(y_{it}-\boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k}-\boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_k}\right)+o_p(1).\end{aligned}$$

From now, we investigate the asymptotic property of the infeasible estimator $\tilde{\boldsymbol{b}}_{i,\tau_k}$, which is obtained as the minimizer of

$$\tilde{\ell}(\boldsymbol{b}_{i,\tau_k}) \equiv \frac{1}{KT}\sum_{k=1}^{K}\sum_{t=1}^{T}\rho_{\tau_k}\left(y_{it} - \boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k} - \boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_k}\right).$$

subject to the restrictions:

$$\boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_{k-1}} \leq \boldsymbol{x}'_{it}\boldsymbol{b}_{i,\tau_k}, \quad t = 1, ..., T,\ k = 2, ..., K.$$

Regarding $y_{it} - \boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k}$ is the response variable, this optimization problem is identical to that of Bondell et al. (2010). Thus, the asymptotic property of $\tilde{\boldsymbol{b}}_{i,\tau_k}$ directly follows from Theorem 1 of Bondell et al. (2010). Thus, we obtain the claim. This completes the proof of Theorem 1.

A.3 Proof of Theorem 2

Using the same argument used in the proof of Theorem 1, the objective function of the risk premium parameter is re-expressed as

$$\begin{aligned} & \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\rho_{\tau_k}\left(y_{it} - \hat{\boldsymbol{f}}'_{t,\tau_k}\hat{\boldsymbol{\lambda}}_{i,\tau_k} - \boldsymbol{r}(\tau_k)'\hat{\boldsymbol{b}}_{i,\tau_k}\right) \\ = \quad & \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\rho_{\tau_k}\left(y_{it} - \boldsymbol{f}'_{t,0,\tau_k}\boldsymbol{\lambda}_{i,0,\tau_k} - \boldsymbol{r}(\tau_k)'\boldsymbol{b}_{i,0,\tau_k}\right) + o_p(1), \end{aligned}$$

where used

$$\begin{aligned} \max_i \|\hat{\boldsymbol{\lambda}}_{i,\tau} - \boldsymbol{\lambda}_{i,0,\tau}\| &= O_p(\log(N)/\sqrt{T}), \\ \max_i \|\hat{\boldsymbol{b}}_{i,\tau} - \boldsymbol{b}_{i,0,\tau}\| &= O_p(\log(N)/\sqrt{T}), \\ \max_t \|\hat{\boldsymbol{f}}_{t,\tau} - \boldsymbol{f}_{t,0,\tau}\| &= O_p(\log(T)/\sqrt{N}) \end{aligned}$$

and Knight's identity.

Thus, it is enough to show the asymptotic equivalence of $\tilde{\boldsymbol{r}}(\tau)$ and $\bar{\boldsymbol{r}}(\tau)$. Let $\hat{\boldsymbol{z}} = N^{1/2}(\bar{\boldsymbol{r}} - \boldsymbol{r}_0)$ and $\tilde{\boldsymbol{z}} = N^{1/2}(\tilde{\boldsymbol{r}} - \boldsymbol{r}_0)$. Similar to Bondell et al. (2010), we can decompose

$$|P(\hat{\boldsymbol{z}} \leq v) - P(\tilde{\boldsymbol{z}} \leq v)| = |P(\hat{\boldsymbol{z}} \leq v|\hat{\boldsymbol{z}} \neq \tilde{\boldsymbol{z}}) - P(\tilde{\boldsymbol{z}} \leq v|\hat{\boldsymbol{z}} \neq \tilde{\boldsymbol{z}})| \times P(\hat{\boldsymbol{z}} \neq \tilde{\boldsymbol{z}}).$$

Because the first term in the product is bounded by 1, it suffices to show that $P(\hat{\boldsymbol{z}} = \bar{\boldsymbol{z}}) \to 1$. As discussed in the proof of Theorem 1 in Bondell et al. (2010), due to the formulation of the estimator, the event $\hat{\boldsymbol{z}} = \bar{\boldsymbol{z}}$ is equivalent to the event that the quantile estimator $\bar{\boldsymbol{r}}(\tau)'\boldsymbol{b}_{i,0,\tau}$ based on $\bar{\boldsymbol{r}}(\tau)$ maintains its appropriate quantile ordering. To show that the probability of this event goes to one, we consider the difference in the following quantity $N^{1/2}\left(\bar{\boldsymbol{r}}(\tau_{k+1})'\boldsymbol{b}_{i,0,\tau_{k+1}} - \bar{\boldsymbol{r}}(\tau_k)'\boldsymbol{b}_{i,0,\tau_k}\right)$ with $\tau_{k+1} > \tau_k$ and τ_{k+1} and τ_k are from a set of K pre-specified quantile levels $\tau_1 < ... < \tau_K$ in the estimation.

The difference $N^{1/2}\left(\bar{\boldsymbol{r}}(\tau_{k+1})'\boldsymbol{b}_{i,0,\tau_{k+1}} - \bar{\boldsymbol{r}}(\tau_k)'\boldsymbol{b}_{i,0,\tau_k}\right)$ can be decomposed as

$$\begin{aligned}
& N^{1/2}\left(\bar{\boldsymbol{r}}(\tau_{k+1})'\boldsymbol{b}_{i,0,\tau_{k+1}} - \bar{\boldsymbol{r}}(\tau_k)'\boldsymbol{b}_{i,0,\tau_k}\right) \\
= \;& N^{1/2}\left(\bar{\boldsymbol{r}}(\tau_{k+1})'\boldsymbol{b}_{i,0,\tau_{k+1}} - \boldsymbol{r}_0(\tau_{k+1})'\boldsymbol{b}_{i,0,\tau_{k+1}}\right) - N^{1/2}\left(\bar{\boldsymbol{r}}(\tau_k)'\boldsymbol{b}_{i,0,\tau_k} - \boldsymbol{r}_0(\tau_k)'\boldsymbol{b}_{i,0,\tau_k}\right) \\
& + N^{1/2}\left(\boldsymbol{r}_0(\tau_{k+1})'\boldsymbol{b}_{i,0,\tau_{k+1}} - \boldsymbol{r}_0(\tau_k)'\boldsymbol{b}_{i,0,\tau_k}\right). \qquad (14)
\end{aligned}$$

It is known that the unrestricted estimator $\bar{\boldsymbol{r}}$ is $N^{1/2}$-consistent. Therefore, the first two terms in (14) are $O_p(1)$ for any τ_a.

We next investigate the last term in (14). By the Mean Value Theorem, we have

$$N^{1/2}\left(\boldsymbol{r}_0(\tau_{k+1})'\boldsymbol{b}_{i,0,\tau_{k+1}} - \boldsymbol{r}_0(\tau_k)'\boldsymbol{b}_{i,0,\tau_k}\right) = (\tau_{k+1} - \tau_k)\frac{\partial}{\partial\tau}\boldsymbol{r}_0(\tau_{k^*})'\boldsymbol{b}_{i,0,\tau_{k^*}},$$

where $\tau_k \leq \tau_{k^*} \leq \tau_{k+1}$. Because $\boldsymbol{b}_{i,0,\tau} < 0$ and the negativity of $\boldsymbol{r}_0(\tau_k)$, we have

$$\frac{\partial}{\partial\tau}\boldsymbol{r}_0(\tau_{k^*})'\boldsymbol{b}_{i,0,\tau_{k^*}} > C > 0.$$

where C is some positive constant. Therefore, we have

$$N^{1/2}\left(\boldsymbol{r}_0(\tau_{k+1})'\boldsymbol{b}_{i,0,\tau_{k+1}} - \boldsymbol{r}_0(\tau_k)'\boldsymbol{b}_{i,0,\tau_k}\right) \geq C \times N^{1/2} \times (\tau_{k+1} - \tau_k).$$

By assumption, the right hand side diverges. This indicates that the third term in (14) dominates in the difference $N^{1/2}\left(\bar{\boldsymbol{r}}(\tau_{k+1})'\boldsymbol{b}_{i,0,\tau_{k+1}} - \bar{\boldsymbol{r}}(\tau_k)'\boldsymbol{b}_{i,0,\tau_k}\right)$ with probability tending to one. Noting that the difference $\tau_{k+1} - \tau_k > 0$, the difference will be positive. This implies that $\bar{\boldsymbol{r}}(\tau)$ and $\hat{\boldsymbol{r}}(\tau)$ share the same asymptotic distribution. This completes the proof of Theorem 2.

B Data acquisition procedure of mutual fund data

First, we obtained a list of passive funds that are classified as mutual funds. The screening criteria are as follows: General Attribute is 'Index Fund', Fund Type is 'Open-End Funds', and Fund Primary Share Class is 'Yes'. Then, we obtained a list of passive funds that are classified as ETFs. The screening criteria are as follows: Fund Type is 'Exchange Traded Products', and Fund Primary Share Class is 'Yes'. After we obtained the list of passive funds, we omitted the leveraged funds, bear funds, and misclassified funds. Specifically, we calculated the beta against the benchmark of each fund and then excluded funds that have beta less than 0.95 or greater than 1.05. The flow data for ETFs were directly obtained from Bloomberg. However, the flow data of ETFs in the U.S. were adjusted a lag of one day because the shares outstanding is reported by the ETF issuers with a one-day lag. There are some administrators who reported no lag data, but it has not been distinguished on Bloomberg data. [6] For this reason, we adjusted one day for every ETF in the U.S.

[6] We confirmed this point to Bloomberg.