A Tool Kit for Factor-Mimicking Portfolios

Kuntara Pukthuanthong
University of Missouri
Email: pukthuanthongk@missouri.edu

Richard Roll
California Institute of Technology
Email: rroll@caltech.edu

Junbo Wang
Louisiana State University
Email: junbowang@lsu.edu

Tengfei Zhang
Louisiana State University
Email: tzhan23@lsu.edu

Abstract

We relate factor-mimicking portfolios (FMPs) to the beta-pricing model and propose that each FMP should minimize the mispricing component of its underlying factor. We also examine FMP construction when the underlying factor contains noise and offer a new method to resolve this issue. For both classical and our newly proposed methods, we recommend enhanced necessary criteria. FMPs of several macroeconomic factors constructed by our method satisfy these criteria. We find that equity returns are priced by consumption growth, inflation, and the unemployment rate; and corporate bond returns are priced by consumption growth, industrial production, and the default spread.

Key words: factor-mimicking portfolios, non-traded factors, risk premium

JEL classification: G10, G12, G11

1. Introduction

Perhaps the most important question in asset pricing is whether different average returns across assets are rewarded for risk. Firm characteristics, such as size, book-to-market ratio, momentum, investment, and profitability, are related reliably to cross-sectional differences in asset returns (Fama and French 1993, 2016, 2018a; Hou, Xue, and Zhang 2015). However, whether they proxy for risk exposures still needs to be determined (Pukthuanthong, Roll, and Subrahmanyam 2019).

Theoretically, macroeconomics factors (e.g., Chen, Roll, and Ross 1986) and other non-traded factors (e.g., Adrian, Etula, and Muir 2014) capture the fundamental risks in the economy and thus should also explain the cross-sectional expected returns. However, observed changes in these factors contain measurement errors and provide only weak predictions of asset returns. To reduce factor noise, the previous literature (Huberman, Kandel, and Stambaugh 1987; Breeden, Gibbons, and Litzenberger 1989; Ferson, Siegel, and Xu 2006; Balduzzi and Robotti 2008; Giglio and Xiu 2019) recommends factor-mimicking portfolios (FMPs), which contain traded assets that are representations of underlying non-traded factors. The extant literature often uses FMPs in testing asset-pricing models (e.g., Cooper and Priestley 2011; Barillas et al. 2019; Pukthuanthong et al. 2019).[1]

Following Huberman, Kandel, and Stambaugh (1987) and Breeden, Gibbons, and Litzenberger (1989), an FMP is chosen to maximize the correlation with the underlying factor (i.e., a maximum correlation portfolio). In this paper, we provide a new theory of the FMP. Specifically, our FMP is constructed to jointly minimize the mispricing component of stock returns with respect to the underlying factors and to meet the requirement that the covariance of any test asset return with the underlying factor is the same as that with the FMP return. In other words, among the FMPs that represent the underlying factor, the least-mispriced portfolio is selected. Through the least-mispriced portfolio, we link the FMP construction with the beta pricing model, while the maximum correlation portfolio is built on mean-variance efficiency. Thus, the connection between the two

[1] The other applications are pointed out by Roll and Srivastava (2018, p. 21): "mimicking portfolios have many potential uses, including (though not limited to): (1) Evaluating active manager performance, (2) Substituting for a desired investment in illiquid assets, (3) Determining the true potential for improved diversification, (4) Understanding the sources of past return volatility, (5) Predicting the likely level of future return volatility."

theories is based on the equivalence between the beta pricing model and the mean-variance efficiency.

Least-mispricing theory suggests several methods for constructing FMPs. The first method is based on Lehmann and Modest (1988, henceforth LM), who apply the weighted-least-squares Fama-Macbeth (1973) cross-sectional regressions and then use the time series of estimated coefficients in the second-pass regression as FMPs. A simplified Ordinary Least Squares (OLS) cross-sectional method is also widely used. The second is the time-series approach, where non-traded factors are regressed on contemporaneous returns of basis assets. Lamont (2001) is one of the leading papers using this approach; the paper constructs FMPs from 13 basis assets, which are portfolios formed by sorting individual firm characteristics. The third approach is the sorting-by-beta approach, where stocks are sorted into portfolios by their factor loadings (betas). Then a long-short portfolio between top and bottom deciles is used as an FMP.

When the underlying factor contains measurement errors and when there are multiple correlated risk factors, empirically constructed FMPs are subject to several methodological difficulties. The first issue is factor contamination. Specifically, when there is a measurement error in the underlying factor, and the underlying factor is correlated with another risk factor, we can show that the constructed FMP is a linear combination of the underlying factor and the other risk factor. Hence, the FMP not only represents the underlying factor, but also the other factor. In other words, the risk associated with the FMP is contaminated by the other factor, which defeats the purpose for creating the FMP. If this issue is not resolved, then we can introduce a different source of error in the factor. We propose a methodology to avoid factor contamination.

The cross-sectional method also suffers an errors-in-variables (EIV) issue. To correct it, our first enhanced approach relies on instrumental variable (IV) estimation with individual equities, following Jegadeesh et al. (2019). We divide the entire sample into even- and odd-month subsamples and estimate betas in each subsample separately. Then betas from the even-month subsample act as instrumental variables for the betas from odd months, or vice versa, in cross-sectional IV regressions with individual stock returns as the dependent variable. Stein (1956) introduces a shrinkage method to reduce the root-mean-square error. We also examine this alternative approach to FMP construction.

Given that an FMP is an excess return, Shanken (1992) shows that its average value is the risk-premium estimate. Another method is to reapply cross-sectional regression to estimate the risk premium using FMPs as factors. The first method is called a one-stage method, since the FMP creation and risk-premium estimation occur in the same cross-sectional regression. We call the latter a two-stage method, since we create an FMP in first stage and reapply the cross-sectional method in the second stage. Creating FMPs using the cross-sectional method can mitigate the measurement errors in the underlying factor; thus, it is natural to believe that reapplying the cross-sectional method can further reduce the effect of measurement errors in risk-premium estimation. We show that this is indeed the case. The one-stage method results in noisier risk-premium estimates for factors with large measurement errors than does the two-stage method; thus, we propose using the two-stage approach when applying FMPs to test asset-pricing models.

Basis asset selection can be important in FMP construction. However, different research papers create FMPs from various candidate assets.[2] To avoid arbitrary basis asset selection, we propose to use a large number of test assets (preferably individual stocks/bonds). However, Gospodinov, Kan, and Robotti (2019) show that including uncorrelated assets can affect the inference of the asset-pricing test. To mitigate this issue, we further suggest variable selection criteria to exclude uncorrelated assets.

Our simulations show that an FMP constructed following the approach proposed above (we call it "IV approach") can correctly represent the underlying factor. We find that the correlation between the IV FMP and the return-related component of the underlying factor is nearly equal to 1.[3] However, the correlation is smaller than 0.8 for all existing methods. Moreover, the IV FMP of an underlying factor is not correlated with other uncorrelated (orthogonalized) risk

[2] Lamont (2001) proposes economic tracking portfolios using 13 basis assets that include eight industry-sorted stock portfolios, four bond portfolios, and a stock market return. Vassalou (2003) uses six equity portfolios sorted by size and book-to-market, term spread, and default spread. Kroencke et al. (2013) use equity portfolios sorted by size and book-to-market, and a momentum portfolio. Bianchi, Guidolin, and Ravazzolo (2017) use six size and book-to-market sorted portfolios, plus the default and term spreads. Barillas and Shanken (2017) use 15 traded factors as basis assets. Maio (2018) uses the excess market return, value spread, term spread, and S&P 500 price-to-earnings ratio. With the cross-sectional approach, Lehmann and Modest (1988) use size, dividend yield, and variance sorted portfolios; and Cooper and Priestley (2011) use 40 portfolios sorted by size, book-to-market, momentum, and asset growth. Pukthuanthong et al. (2019) use 50 portfolios sorted by size, book-to-market, momentum, investment and operating profitability. Roll and Srivastava (2018) use eight Exchange Traded Funds (ETF) portfolios.

[3] We define a non-traded factor $\tilde{f}$ as $\tilde{f} = f + \varepsilon_f$. See Section 2.1 for our theory and definition of variables.

factors, while FMPs constructed by other methods are correlated with them, suggesting that only the IV FMP is uncontaminated by other factors.

We evaluate existing and newly introduced approaches to constructing FMPs and test underlying factors. We propose that the examination should hinge on the following criteria:

(1) FMPs should be correlated with the underlying factors,

(2) FMPs should be correlated with the systematic risk of returns, and

(3) FMPs should explain the cross-section of mean returns.

Intuitively, an FMP should be a proxy for the risk of the underlying factor; therefore, (1) should be satisfied. In addition, following Pukthuanthong, Roll, and Subrahmanyam (2019), if the underlying factor is a true pricing factor that reflects asset risk, an FMP should represent systematic risk and price assets cross-sectionally (so that criteria (2) and (3) also should be satisfied).

The criteria are examined in the real data. For criterion (1), we estimate the correlations between FMPs and underlying factors. We find that the correlation between the FMP return and the underlying factor using the IV approach is smaller than that using the cross-sectional OLS approach. However, we show that the higher correlation of the OLS approach can be driven by the EIV issues. Therefore, the neoclassical criteria for creating the empirical maximum correlation portfolio, which is the FMP created by OLS method, can be misleading, especially for weak macroeconomic factors in which the EIV bias is relatively large. We also find that the above correlation for the IV approach is larger than that for the time-series approach. Across subperiods, the time-series approach cannot produce significant correlation, while the cross-sectional approaches can.

We also examine criterion (2) following Pukthuanthong, Roll, and Subrahmanyam (2019), in which they propose that a genuine risk factor must be related to the systematic risks (proxied by the covariance matrix of returns).[4] We find that for cross-sectional approaches, most of the FMPs

[4] Kozak, S., Nagel, S., & Santoshi, S. (2018) argue that the covariance of assets might not represent the risk. However, the risk factors should be associated with the covariance of the assets. Thus, examining whether covariance of assets is related to FMP can be viewed as a necessary condition to identify the risk factor.

are related with the covariance of stocks. However, for the time-series approach, the FMPs fail to consistently deliver a systematic risk.

For criterion (3), our empirical results reveal that IV-constructed FMPs yield more significant risk premiums. For example, the monthly risk premium for the FMP of consumption growth constructed with the IV approach is 0.164%, while the risk premium of the OLS method is 0.066%. The significance level increases from 5% to 1%. We also find that FMPs for the unemployment rate and CPI contain significant risk premiums when they are constructed with IV but not when obtained via other approaches. The Lehmann and Modest (1988) and the Stein (1956) methods yield more significant risk premiums than does the OLS method, but not as significant as the IV approach. Risk premiums estimated by the Lamont (2001) and sorting-by-beta approaches are in general insignificant. To study the robustness of consumption growth, we also include the CAY (the log of the consumption-to-wealth ratio) factor of Lettau and Ludvigson (2001) and the consumption volatility factor by Boguth and Kuehn (2013), but they deliver little incremental power to consumption growth.

We also construct FMPs from various approaches for non-traded factors to price individual corporate bonds. The only method of FMP construction that passes all three criteria is the IV approach. Using the IV, we find that consumption growth, industrial production, bond market factors, and the default spread are associated with significant and positive risk premiums.

Our contribution is fourfold. First, we provide a new economic interpretation for the factor-mimicking portfolio, in addition to the maximal correlation portfolio proposed by Breeden, Gibbons, and Litzenberger (1989), and Huberman, Kandel, and Stambaugh (1987). We also examine the connection between the two theories.

Second, we propose new methods for constructing FMPs. In simulations, we find that the IV approach yields a perfect proxy for the risk component of non-traded factors. In contrast, previous cross-sectional approaches—such as OLS, Lehmann and Modest (LM) (1988), Stein's approach, as well as sorting-by-beta and time-series approach—are subject to substantial factor contamination.

Third, we propose three selection criteria for examining factor-mimicking portfolios. For example, instead of maximizing the correlation between the FMP and the underlying factor, our analogous but revised criterion is that an effective FMP should have significant, but not maximum, correlation with its underlying risk factor. This is to avoid inflated correlation in the presence of estimation errors.

Fourth, we sponsor a horse race among the IV approach and other approaches in both the stock and bond markets. We find that IV is the winning horse for traded versions (FMPs) of macroeconomic factors, including consumption growth, the CPI, unemployment, and default spread. FMPs constructed with IV satisfy three criteria: they are correlated with the underlying factors; associated with the covariance in asset returns; and, unlike the alternative approaches, have large and significant risk premiums in both equity and bond markets.

Our paper is related to Balduzzi and Robotti (2008), who conclude that using the time-series formulation of FMPs performs better in term of estimating risk premiums than using the original factors with the one-step cross-sectional approach. We provide a possible explanation for their findings, since the finite sample error is much larger for the one-step cross-sectional approach. Kleibergen and Zhan (2018) propose a test of the risk premiums of FMPs constructed by a time-series approach that does not depend on the magnitude of betas. However, their approach focuses more on inference and suffers from information lost through their portfolio construction. Rather than constructing mimicking portfolios for factors, Roll and Srivastava (2018) construct mimicking portfolios for individual stock returns using a cross-sectional OLS approach. Fama and French (2018b) construct mimicking portfolios for characteristics using the cross-sectional approach and find that these characteristic-mimicking portfolios have better explanatory power to average return than sorting-characteristic-based factors (such as SMB, HML, etc.). We focus on non-traded factors, especially macroeconomic factors. Instead of using FMPs, Kleibergen, and Zhan (2019 forthcoming) extend the Gibbons-Ross-Shanken statistic to identify the risk premiums of macro-risk factors and also uncover the significant risk premium associated with consumption growth.

2. The least-mispriced portfolio and its relation to the maximum correlation portfolio

This section proposes a new economic theory of an FMP. We also connect this new theory to maximum correlation theory. Although the economic interpretations are different, the two theories lead to the same FMP formula. We discuss the economic reason for the equivalence between these two theories and derive three different methods to construct FMPs based on the new theory.

2.1. The least-mispriced portfolio theory

The FMP is constructed by minimizing the mispriced component of asset returns. Let N denote the number of test assets. Let $\boldsymbol{R} = [R^1, \cdots, R^N]$,[5] an N by 1 vector, denote their excess returns. Let a non-traded factor be $\tilde{f}$.[6] The factor can be decomposed as $\tilde{f} = f + \varepsilon_f$, where f is its projection into the excess return space (i.e., there is a linear combination of excess returns that is equal to f), and ε_f is the measurement error, with mean zero and $cov\left(\varepsilon_f, \boldsymbol{R}\right) = 0$. We assume that the excess returns depend linearly on the projected factor f,

$$\boldsymbol{R} = \boldsymbol{\alpha} + \boldsymbol{\beta} f + \boldsymbol{\varepsilon}\,. \tag{1}$$

Here, $\boldsymbol{\beta}$ (N by 1) is the factor loading, $\boldsymbol{\alpha}$ (N by 1) is the mispricing component, $\boldsymbol{\varepsilon}$ (N by 1) is the residual of the pricing model, and its variance is denoted by $\boldsymbol{\Omega}$. Also, we assume that $\tilde{f}$ contains only one factor; therefore, it is possible that the residual, $\boldsymbol{\varepsilon}$, can be correlated with other factors. With this assumption, we extract the largest component of the return that is correlated with the factor, regardless of the existence of other factors.

Our goal is to select a factor-mimicking portfolio (FMP) that can minimize the mispricing of the asset-pricing model. The minimization problem can be written as follows:

$$\min_{\boldsymbol{w}} \boldsymbol{\alpha}' \, \boldsymbol{\Sigma} \boldsymbol{\alpha} \text{ subject to } cov(\boldsymbol{w}'\boldsymbol{R}, \boldsymbol{R}) = cov(\tilde{f}, \boldsymbol{R}). \tag{2}$$

Here $\boldsymbol{w}$ (N by 1) represents the weight of the portfolio, and the weighting matrix, $\boldsymbol{\Sigma}$ (N by N), is used to control the relative importance of the mispricing components across assets.

We minimize the mispricing component with respect to a single factor and solve for the portfolio weights of the FMP. It is possible to solve the same problem to obtain portfolio weights with respect to multiple factors. However, although these FMPs together can jointly minimize the mispricing component of the multifactor model, each of them does not guarantee to minimize the mispricing component of a single-factor model with its underlying as the only factor. This is because, by controlling for other factors, the FMP obtained from the multi-factor optimization

[5] Note that each entry of $\boldsymbol{R}$ is a random variable, representing the returns of N assets. In a later section, we will use $\mathfrak{R}$, a T by N matrix, to represent the matrix of time-series realization of N assets.

[6] Note that $\tilde{f}$ is a random variable. In later sections, we use the notation $\tilde{\boldsymbol{f}} = [\tilde{f}_1, \cdots, \tilde{f}_T]'$, a T by 1 vector, to represent the time-series realization of the non-traded factors. This convention is applied to all random variables defined in this paper.

problem might not extract the largest correlated component of the return from the single-factor model. Evaluating the risk or pricing effect of the portfolio computed from such weights can be misleading, because they might not fully reflect the risk or pricing effect of the underlying factors.

Note that the constraint in the minimization problem requires that for any asset, the covariance between the mimicking portfolio and the asset's return is the same as the covariance between the factor and the asset's return. Since f is the projection of the non-traded factor, the constraint (2) also implies that the portfolio return is equal to f, that is,

$$\boldsymbol{w}'\boldsymbol{R} = f.$$ [8]

With the above equation, $\boldsymbol{\alpha} = E(\boldsymbol{R}) - \boldsymbol{\beta}E(f) = E(\boldsymbol{R}) - \boldsymbol{\beta}E\left(\boldsymbol{w}'\,\boldsymbol{R}\right)$. Therefore, the objective function can be written as follows:

$$\min_{\boldsymbol{w}} \boldsymbol{\alpha}'\,\boldsymbol{\Sigma}\boldsymbol{\alpha} = \left(E(\boldsymbol{R}) - \boldsymbol{\beta}E\left(\boldsymbol{w}'\,\boldsymbol{R}\right)\right)'\,\boldsymbol{\Sigma}(E(\boldsymbol{R}) - \boldsymbol{\beta}E\left(\boldsymbol{w}'\,\boldsymbol{R}\right)) \quad .$$

(3)

In the following proposition, we derive the solution to the above minimization problem.

Proposition 1: Let the scaled weight be $\boldsymbol{w_s} = \frac{\boldsymbol{w}}{var(\boldsymbol{w}'\,\boldsymbol{R})}$. The optimum weight solution to the minimization problem (3) can be written as $\boldsymbol{w_s^*} = \frac{(cov(f,\boldsymbol{R}))'\,\boldsymbol{\Sigma}E(\boldsymbol{R})}{(cov(f,\boldsymbol{R}))'\,\boldsymbol{\Sigma}(cov(f,\boldsymbol{R}))}E(\boldsymbol{R})(E(\boldsymbol{R})'\,E(\boldsymbol{R}))^{-1}$. Hence, the expected portfolio return is

$$E(\boldsymbol{R})'\,\boldsymbol{w_s^*} = \frac{(cov(f,\boldsymbol{R}))'\,\boldsymbol{\Sigma}E(\boldsymbol{R})}{(cov(f,\boldsymbol{R}))'\,\boldsymbol{\Sigma}(cov(f,\boldsymbol{R}))}. \tag{4}$$

---

[8] From $cov(\boldsymbol{w}'\boldsymbol{R},\boldsymbol{R}) = cov(\tilde{f},\boldsymbol{R}) = cov(f,\boldsymbol{R})$, we have that $cov(\boldsymbol{w}'\boldsymbol{R} - f,\boldsymbol{R}) = cov(\boldsymbol{w}'\boldsymbol{R},\boldsymbol{R}) - cov(f,\boldsymbol{R}) = 0$. Since $\boldsymbol{w}'\boldsymbol{R} - f$ is in the excess return space, there is a linear combination of returns (denoted by $\boldsymbol{\pi}'\boldsymbol{R}$) that is equal to $\boldsymbol{w}'\boldsymbol{R} - f$. Thus, the equation $cov(\boldsymbol{w}'\boldsymbol{R} - f,\boldsymbol{R}) = 0$ implies that $var(\boldsymbol{w}'\boldsymbol{R} - f) = cov(\boldsymbol{w}'\boldsymbol{R} - f,\boldsymbol{w}'\boldsymbol{R} - f) = cov(\boldsymbol{w}'\boldsymbol{R} - f,\boldsymbol{\pi}'\boldsymbol{R}) = 0$. Given that the only random variable in excess return space that has zero variance is zero (if there is another constant number in the excess return space, there will be two different risk-free rates, which is impossible), it follows that $\boldsymbol{w}'\boldsymbol{R} - f = 0$.

The proof is in Appendix C. Given that the portfolio can minimize the mispricing component of the test assets, we call it the least-mispriced portfolio. Since $\boldsymbol{\beta} = \frac{cov(f,\boldsymbol{R})}{var(f)}$, equation (4) is equivalent to

$$E(\boldsymbol{R})' \, \boldsymbol{w}_{\boldsymbol{s}}^{*} = \frac{1}{var(f)} \frac{\boldsymbol{\beta}' \, \boldsymbol{\Sigma} E(\boldsymbol{R})}{\boldsymbol{\beta}' \, \boldsymbol{\Sigma} \boldsymbol{\beta}}. \tag{5}$$

From equation (5), if the cross-sectional expected returns are more highly correlated with their $\boldsymbol{\beta}$, then the numerator in equation (5) is larger. Thus, the optimum portfolio has higher expected returns. Intuitively, $\boldsymbol{\beta}$ reflects the sensitivity of the asset to the factor risk. If the majority of the cross-sectional variation of asset returns can be explained by this factor, then the mispriced component is smaller. Hence, the expected return of the pricing component, which is the left-hand side of equation (5), should be larger.

Define a unit weight as $\boldsymbol{w}_{\boldsymbol{u}} = \frac{\boldsymbol{w}}{std(\boldsymbol{w}' \, \boldsymbol{R})}$, and a unit factor as $f_u = \frac{f}{std(\boldsymbol{w}' \, \boldsymbol{R})}$. Equation (4) can also be rewritten as

$$E(\boldsymbol{R})' \, \boldsymbol{w}_{\boldsymbol{u}}^{*} = \frac{\left(cov(f_u,\boldsymbol{R})\right)' \, \boldsymbol{\Sigma} E(\boldsymbol{R})}{\left(cov(f_u,\boldsymbol{R})\right)' \, \boldsymbol{\Sigma}\left(cov(f_u,\boldsymbol{R})\right)}. \tag{6}$$

It is easy to show that the standard deviation of the unit weight portfolio return is $std(\boldsymbol{R}' \, \boldsymbol{w}_{\boldsymbol{u}}^{*}) = \frac{std\left(\boldsymbol{R}' \, \boldsymbol{w}\right)}{std\left(\boldsymbol{R}' \, \boldsymbol{w}\right)} = 1$. Hence, equation (6) represents the expected return of a portfolio with unit risk.

Given that the expected return for a factor-mimicking portfolio is also the risk premium of the underlying non-traded factor, equation (6) also represents the risk premium of the factor with unit risk.

2.2. Relation to the maximum correlation portfolio

The classical factor-mimicking portfolio is constructed by maximizing its correlation with the non-traded factor. Breeden, Gibbons, and Litzenberger (1989) and Roll and Srivastava (2018) provide the theoretical framework for maximum correlation portfolio construction. We follow their approach. For a given value of the loading of the risk factor on its mimicking portfolio, β_{fmp} (Normally, β_{fmp} is set to be 1), the variance of the FMP (following the same notation, $\boldsymbol{w}'\boldsymbol{R}$) is the reciprocal of the correlation between FMP and $\tilde{f}$, because the variance for the factor $(var(\tilde{f}))$ is a constant. The relation can be illustrated by the following formula:

$$\beta_{fmp} = corr\left(\tilde{f}, \boldsymbol{w}'\ \boldsymbol{R}\right)\frac{\sqrt{Var(\boldsymbol{w}'\ \boldsymbol{R})}}{\sqrt{Var(\tilde{f})}}. \tag{7}$$

Therefore, maximizing the correlation between FMPs and their underlying factor is equivalent to minimizing the variance of the FMPs themselves. Specifically, we want to select the weight for the following minimization problem:

$$\min_{\boldsymbol{w}} \boldsymbol{w}'\ \boldsymbol{V}\boldsymbol{w} + 2(\beta_{fmp} - \boldsymbol{w}'\boldsymbol{\beta})\,\lambda, \tag{8}$$

where $\boldsymbol{V}$ is the covariance matrix of testing asset returns, and λ is the Lagrange multiplier.

By solving the first-order condition of expression (8) and setting β_{fmp} to 1, the optimal weight is

$$\boldsymbol{w}^* = \boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}. \tag{9}$$

Then the expected return of the maximum correlation portfolio can be computed as

$$E(\boldsymbol{R})'\boldsymbol{w}^* = E(\boldsymbol{R})'\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}. \tag{10}$$

If we replace $\boldsymbol{\Sigma}$ by $\boldsymbol{V}^{-1}$, equations (5) and (10) are identical, except for the scaling in weight $(\frac{1}{var(f)})$. The maximum correlation portfolio coincides with the least-mispriced portfolio. The coincidence is not surprising, given the equivalence between mean-variance minimization and the beta-pricing model. If the maximum correlation portfolio of a factor lies on the mean-variance frontier of the test assets (so that the constraint in expression (8) is binding), the covariance between FMP return ($\boldsymbol{w}'\boldsymbol{R}$) and any asset return is a linear function of its expected return (Cochrane 2009), that is,

$$cov(\boldsymbol{w}'\boldsymbol{R}, \boldsymbol{R}) = \gamma_0 + \gamma_1 E(\boldsymbol{R}).$$

Here, γ_0 and γ_1 are constants. Since the $\boldsymbol{w}'\boldsymbol{R}$ and zero beta rate R_0 should also be priced, we have $var(\boldsymbol{w}'\ \boldsymbol{R}) = \gamma_0 + \gamma_1 E(\boldsymbol{w}'\boldsymbol{R})$, and $0 = \gamma_0 + \gamma_1 E(R_0)$. Combining these equations, we obtain

$$E(\boldsymbol{R}) - E(R_0) = \frac{cov(\boldsymbol{w}'\boldsymbol{R},\boldsymbol{R})}{var(\boldsymbol{w}'\boldsymbol{R})}\left(E(\boldsymbol{w}'\boldsymbol{R}) - E(R_0)\right) = \boldsymbol{\beta}\left(E(\boldsymbol{w}'\boldsymbol{R}) - E(R_0)\right) \quad . \tag{11}$$

Equation (11) implies that the maximum correlation portfolio can correctly price all test assets. Hence, the portfolio is also the least-mispriced portfolio (its mispricing is zero). When the factor misprices some of the assets, its maximum correlation portfolio is not on the mean-variance frontier. Equations (5) and (10) show that the mimicking portfolio that minimizes its variance (i.e., is closest to the mean-variance frontier) is also the portfolio that minimizes the mispricing component of the beta-pricing model.

2.3. Implied methodology for FMP construction

Both maximum-correlation theory and least-mispricing theory imply three construction methods for FMPs. In this section, we describe these methods from the latter theory. From equation (5), the portfolio return is $E(\boldsymbol{R})'\boldsymbol{w}_s^* = \frac{1}{var(f)}\frac{\boldsymbol{\beta}'\boldsymbol{\Sigma}E(\boldsymbol{R})}{\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}}$. The expected return ($E(\boldsymbol{R})$) and the factor loading ($\boldsymbol{\beta}$) are known (or can be estimated). Therefore, the different methods depend on different choices of the weighting matrix $\boldsymbol{\Sigma}$.

Case 1: time-series method. In this case, the weighting matrix is the inverse of the covariance matrix of testing-asset returns (i.e., $\boldsymbol{\Sigma} = \boldsymbol{V}^{-1}$). Thus we have

$$\left(var(f)\right)^2\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}E(\boldsymbol{R})'\boldsymbol{w}_s^* = var(f)\boldsymbol{\beta}'\boldsymbol{\Sigma}E(\boldsymbol{R}) = cov(f,\boldsymbol{R})'\boldsymbol{V}^{-1}E(\boldsymbol{R}) \quad . \tag{12}$$

The left-hand side is a scaled expected return of the least mispricing portfolio. The right-hand side provides the estimation method to calculate the return. Specifically, it can be estimated by regressing the non-traded factor on returns of the test assets:

$$f = a + b\boldsymbol{R} + \boldsymbol{u}. \tag{13}$$

The fitted value of the above time-series regression is $cov(f,\boldsymbol{R})'\boldsymbol{V}^{-1}\boldsymbol{R}$. Take the expected value; it is the same as the right-hand side of the equation (12).

Case 2: cross-sectional method. The weighting matrix can be written as $\boldsymbol{\Sigma} = (\boldsymbol{I} - \mathbf{1}(\mathbf{1}'\ \mathbf{1})^{-1}\mathbf{1}'\ )\boldsymbol{V}^{-1}(\boldsymbol{I} - \mathbf{1}(\mathbf{1}'\ \mathbf{1})^{-1}\mathbf{1}')$, where, $\boldsymbol{I}$ is the N by N identity matrix, and $\mathbf{1}$ is an N by 1 vector, with each entry being 1. Replacing $\boldsymbol{\Sigma}$ in equation (5), we obtain:

$$var(f)E(\boldsymbol{R})'\boldsymbol{w}_{\boldsymbol{s}}^{*} = \frac{\boldsymbol{\beta}'\left(\boldsymbol{I}-\mathbf{1}(\mathbf{1}'\ \mathbf{1})^{-1}\mathbf{1}'\right)\boldsymbol{V}^{-1}\left(\boldsymbol{I}-\mathbf{1}(\mathbf{1}'\ \mathbf{1})^{-1}\mathbf{1}'\right)E(\boldsymbol{R})}{\boldsymbol{\beta}'\left(\boldsymbol{I}-\mathbf{1}(\mathbf{1}'\ \mathbf{1})^{-1}\mathbf{1}'\right)\boldsymbol{V}^{-1}\left(\boldsymbol{I}-\mathbf{1}(\mathbf{1}'\ \mathbf{1})^{-1}\mathbf{1}'\right)\boldsymbol{\beta}} = \frac{\overline{\boldsymbol{\beta}}'\ \boldsymbol{V}^{-1}\bar{E}(\boldsymbol{R})}{\overline{\boldsymbol{\beta}}'\ \boldsymbol{V}^{-1}\overline{\boldsymbol{\beta}}}. \quad (14)$$

Here, for any random variable $\boldsymbol{X}$, the notation $\overline{\boldsymbol{X}}$ is the demeaned $\boldsymbol{X}$, where the mean is taken across test assets. The left-hand side of equation (14) still represents the expected return of a scaled least-mispricing portfolio. The right-hand side is the coefficient of regressing the expected return on its factor loadings across test assets. Specifically, for the following cross-sectional regression,

$$E(\boldsymbol{R}) = \alpha + \gamma\boldsymbol{\beta} + \boldsymbol{v}, \quad (15)$$

when the estimated coefficient γ is based on Generalized Least Square (GLS) with weighting matrix $\boldsymbol{V}^{-1}$, equation (15) takes the same form as the right-hand side of equation (14).

In this case, the weighting matrix contains the covariance matrix of asset returns. This matrix is more difficult to estimate when the number of test assets is large. A special scenario, in this case, is to set $\boldsymbol{V}^{-1} = \boldsymbol{I}$. The right-hand side of the equation (14) becomes $\frac{\overline{\boldsymbol{\beta}}'\ \bar{E}(\boldsymbol{R})}{\overline{\boldsymbol{\beta}}'\ \overline{\boldsymbol{\beta}}}$. This corresponds to the coefficient of regression in equation (15) using OLS. However, Lehmann and Modest (1988) suggest using the diagonal matrix, which only consists of the variances of the residuals from the regression (1) to replace $\mathbf{V}^{-1}$.

Case 3: sorting-by-beta method. In this case, the weighting matrix is diagonal. For example, we divide the assets into five groups. For the assets in the group with the lowest factor loadings/betas, the corresponding diagonal element is the negative reciprocal of the beta. For the assets in the group with the highest factor loadings/betas, the corresponding diagonal element is the positive reciprocal of the beta. For assets of the other three groups, the diagonal elements are zero. Specifically, assume that asset 1 through asset $M = \frac{N}{5}$ are in the group with the smallest beta, and asset $4M + 1$ to asset $5M$ are in the group with the largest beta. Hence, the weighting matrix can be written as $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma_{11}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma_{22}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma_{33}} \end{pmatrix}$, where $\boldsymbol{\Sigma_{11}} = \begin{pmatrix} -\frac{1}{\beta_1} & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & -\frac{1}{\beta_M} \end{pmatrix}$, $\boldsymbol{\Sigma_{22}} = \begin{pmatrix} 0 & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & 0 \end{pmatrix}$, and $\boldsymbol{\Sigma_{33}} = \begin{pmatrix} \frac{1}{\beta_{4M+1}} & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & \frac{1}{\beta_{5M}} \end{pmatrix}$. Replacing $\boldsymbol{\Sigma}$ in equation (5) results in

$$\frac{1}{M} var(f)\boldsymbol{\beta}' \boldsymbol{\Sigma}\boldsymbol{\beta} E(\boldsymbol{R})'\boldsymbol{w}_{\boldsymbol{s}}^* = \frac{1}{M}\left(\sum_{i=4M+1}^{5M} E\left(R^i\right) - \sum_{i=1}^{M} E\left(R^i\right)\right). \qquad (16)$$

The right-hand side of the equation (16) is the difference between the average expected return of the high beta and low beta groups, which represents another method for calculating the scaled least-mispriced portfolio on the left-hand side.

3. Measurement error in factors and econometric issues concerning FMP construction

We derived methods for constructing the least-mispriced portfolio in the previous section. Empirically, the measurement error in the non-traded factor raises issues with these methods. In this section, we discuss these issues and propose an adjustment approach. The focus is on the cross-sectional method, but we will also discuss the bias in the time-series and the sorting-by-beta methods.

3.1. One-factor case

In this section, we examine the simple case when the true expected asset returns depend only on a single factor, which is non-traded. Following the notation in the previous section, the non-traded factor including measurement error is denoted by $\tilde{f}$, f is its component related to returns, and ε_f is the measurement error. The FMP is constructed by the two-pass cross-sectional method. For a sample period of length T periods, let R_t^i be the return on asset *i* at time *t*, and $\tilde{f}_t$ be the factor at time *t*. Define $\boldsymbol{R}^i = [R_T^i, \cdots, R_T^i]'$, and $\tilde{\boldsymbol{f}} = [\tilde{f}_1, \cdots, \tilde{f}_T]'$. The first-pass estimates betas by running a following time-series regression for each asset:

$$\boldsymbol{R}^i = \alpha^i + \beta^i \tilde{\boldsymbol{f}} + \boldsymbol{\varepsilon}^i, \qquad (17)$$

where α^i and β^i are regression coefficients, and $\boldsymbol{\varepsilon}^i = [\varepsilon_T^i, \cdots, \varepsilon_T^i]'$ is the regression residual. The second pass is a cross-sectional regression at each time point t (for $t \in \{1,2,\cdots,T\}$). Define $\boldsymbol{R_t} = [R_t^1, \cdots, R_t^N]$ and $\boldsymbol{\beta} = [\beta^{\mathbf{1}}, \cdots, \beta^{\boldsymbol{N}}]$. Then the regression can be written as

$$\boldsymbol{R_t}' = a_t + \lambda_t \widehat{\boldsymbol{\beta}}' + \boldsymbol{\eta_t}. \tag{18}$$

Here we use $\widehat{\boldsymbol{\beta}}$ to represent the estimated value of the factor loading β^i, and residual $\boldsymbol{\eta_t} = [\eta_t^1, \cdots, \eta_t^N]$. The estimated coefficient λ_t is the return for the FMP at time *t*.

The key difference between equations (17) and (18) and the cross-sectional regression on traded factors is that $\tilde{f}$ contains measurement errors. With some mild regularity conditions on the measurement errors ε_f, as well as the factor and regression residuals, the next proposition shows that the FMP constructed by the two-pass method can adjust for the measurement errors.

Proposition 2: Assume that (1) measurement errors ε_f are uncorrelated with R^i for any asset i, uncorrelated with f, and uncorrelated with regression residual ε^i, (2) regression residuals are uncorrelated with factors f, and (3) beta ($\boldsymbol{\beta}$) is uncorrelated with the cross-sectional regression errors ($\boldsymbol{\eta_t}$). As the sample period *T* converges to infinity, the estimated coefficient λ_t converges to $c(f_t - E(f) + \gamma)$, where γ is the factor risk premium, and $c = (var(f) + var(\varepsilon_f))/var(f)$ is a constant.

The proof is in Appendix C. This is also shown in Section 6.2 of Balduzzi and Robotti (2008). From the proposition, the estimated coefficient is a linear transformation of the factor without measurement error.[9] In particular, the FMP is scaled by a constant c. Intuitively, the measurement error in the non-traded factor leads to a scaling effect in the estimated beta coefficient in the first pass. Given that every stock's beta (which is the independent variable for the second-pass regression) is scaled up by the same constant, the estimated coefficient in the second-pass cross-sectional regression is scaled down by the same constant.

Moreover, from Proposition 2, the scaling effect on λ_t is homogeneous across time (by a constant number c). Thus, the scaled FMP can still represent the same factor in the asset-pricing test. To test risk premium, we can calculate the average value of the coefficients over time as the

[9] When there is no measurement error ($\varepsilon_f = 0$), the factor at time t is f_t. Hence, the estimated coefficient in Proposition 2 ($c(f_t - E(f) + \gamma)$) is a linear transformation of the factor without measurement errors.

risk-premium estimates and calculate the Fama-Macbeth standard deviation of the regression coefficients (i.e., the average value of the coefficients is $\frac{1}{T}\sum_{t=1}^{T} c(f_t - E(f) + \gamma)$), and the Fama-Macbeth standard deviation of the coefficients is the sample standard deviation of $c(f_t - E(f) + \gamma)$. Both the average and standard deviation are scaled by the same constant. Hence, the T-stats formed by the estimate and the standard deviation given above are not affected by the constant number and converge to a standard normal distribution. In Section 3.5, we introduce the two-stage regression method (rerun the Fama-Macbeth method on λ_t for another time) to estimate the risk premium. In both testing methods, T-stats are not affected by the constant c, because the estimated risk premium and its standard deviation are both multiplied by the same number.[10]

3.2. Bias in FMP when there is a correlated factor

Although the cross-sectional method for the one-factor model can remove the measurement error and create an FMP that represents the risk of the factor, the true model can contain several correlated factors. Theoretically, to separate the effect of the factor of interest (the non-traded factor) on asset returns and construct an FMP, it is natural to control for other factors. However, for the measurement error in the non-traded factor, even after controlling for other factors, the proposition below shows that the FMP still contains a component from these other factors, which is not a desirable property for an FMP.

To simplify without loss of generality, we assume that there are two factors (the two-factor assumptions apply throughout the analysis in sections 3.2–3.4), $\tilde{f}_1$ and f_2. Factor 1 has measurement error denoted by ε_{f_1}.[11] Moreover, assume that the two factors are correlated (i.e., $cov(f_1, f_2) \neq 0$). With the same vector notation as in Section 3.1, the true regression model can be written as

$$\boldsymbol{R^i} = \alpha^i + \beta_1^i \boldsymbol{f}_1 + \beta_2^i \boldsymbol{f}_2 + \boldsymbol{\varepsilon}^i. \tag{19}$$

[10] In the finite sample, an EIV issue arises from the error in the estimated beta. This EIV issue can lead to biased *t*-statistic. We will discuss this issue and the corresponding method to mitigate it in Section 3.6.

[11] We assume there is no measurement error for the second factor (i.e., it can be a traded factor) to simplify the analysis. We obtain a similar form to that of equation (22) when measurement error is included.

If factor 1 has a measurement error, we must replace $\boldsymbol{f}_1$ by $\tilde{\boldsymbol{f}}_1$, its observed version, before being able to run regression (19). Then with the resulting estimated beta coefficients, we compute the following cross-sectional regression:

$$\boldsymbol{R}_t{}' = \alpha_t + \lambda_{1t}\boldsymbol{\beta}_1{}' + \lambda_{2t}\boldsymbol{\beta}_2{}' + \boldsymbol{\eta}_t. \tag{20}$$

Proposition 3: In the first-pass regression (19), assume that the measurement error of the factor 1 is uncorrelated with asset returns, regression residuals, and both factors; the regression residuals are also uncorrelated with both factors; and the betas are uncorrelated with the cross-sectional regression errors.

(A) When the sample size T converges to infinity, we have

$$\hat{\beta}_1^i \to \frac{var(f_1)var(f_2)-cov(f_1,f_2)^2}{DET_1}\beta_1^i \equiv B_1^i, \text{ and } \hat{\beta}_2^i \to \beta_2^i + \frac{var(\varepsilon_{f_1})(\beta_1^i cov(f_1,f_2)+\beta_2^i var(f_2))}{DET_1} \equiv B_2^i, \tag{21}$$

where $DET_1 = \left(var(f_1) + var\left(\varepsilon_{f_1}\right)\right) var(f_2) - cov(f_1, f_2)^2$.

In the second-pass regression (20),

(B) when both T and N (the number of test assets) converge to infinity, then

$$\hat{\lambda}_{1t} \to w_1\gamma_{1t} + w_2\gamma_{2t}, \tag{22}$$

where

$$w_1 = \frac{1}{DET_2}\frac{var(f_1)var(f_2)-cov(f_1,f_2)^2}{DET_1}\left(\overline{var}\left(B_2^i\right)\overline{var}\left(\beta_1^i\right) - \overline{cov}\left(\beta_1^i, B_2^i\right)\right)$$

$$w_2 = \frac{1}{DET_2}\frac{var(f_1)var(f_2)-cov(f_1,f_2)^2}{DET_1}\left(\overline{var}\left(B_2^i\right)\overline{cov}\left(\beta_1^i, \beta_2^i\right) - \overline{cov}\left(\beta_1^i, B_2^i\right)\overline{cov}\left(\beta_2^i, B_2^i\right)\right)$$

$$DET_2 = \overline{var}\left(B_1^i\right)\overline{var}\left(B_2^i\right) - \overline{cov}\left(B_1^i, B_2^i\right)^2$$

$$\gamma_{1t} = f_{1t} - E(f_1) + \gamma_1$$

$$\gamma_{2t} = f_{2t} - E(f_2) + \gamma_2.$$

Here, $\overline{var}$ and $\overline{cov}$ are the cross-sectional variance and covariance, respectively. We use the upper bar to distinguish them from their time-series companions. Moreover, γ_1 and γ_2 are the risk premium of factors 1 and 2.

(C) If $cov(f_1, f_2) = 0,\ w_2 = 0.$

The proof is in Appendix C. From equation (21) in Proposition 3(A), the estimated beta for factor 2 is related to the beta of factor 1. Moreover, when we employ the estimated betas in the cross-sectional regression, equation (22) in Proposition 3(B) shows that the constructed factor-mimicking portfolio for factor 1 is also affected by factor 2. In this case, the FMP contains an additional component from factor 2; thus, its risk premium reflects the excess returns associated with a combination of the two risk factors. We call this "factor contamination." It is not a desirable property for an FMP. For example, suppose we want to construct the FMP for consumption growth. When there is factor contamination, the FMP contains the risk from consumption growth as well as from other factors. If we would like to test the risk premium of the consumption growth, we essentially test the risk premium of a combination of several factors. If we would like to examine whether the consumption growth is correlated with the stock returns, the correlation can come from value or size factors. We create FMPs to mitigate the measurement error, but instead we introduce another noise from other factors. In conclusion, we cannot construct a pristine FMP using the cross-sectional method by just controlling other correlated risk factors. In section 3.3, we propose a new way to construct an FMP, using a single-factor approach.

When the factors 1 and 2 are uncorrelated, then from Proposition 3(C) and equation (22), the FMP of factor 1 does not depend on factor 2. In the real world, factor 2 is correlated with factor 1. However, if one can create a modified factor 2 that is uncorrelated with factor 1 and use it in the FMP construction for factor 1 (i.e. run regression (19) and (20) with factor 1 and the modified factor 2 to create $\hat{\lambda}_{1t}$), the resulting FMP will have the desirable property and represent the risk for factor 1. A classical approach to constructing two uncorrelated factors is to remove the effect of factor 1 from factor 2 and to use the remaining component of factor 2 (which is uncorrelated with factor 1) as a control. Specifically, in a regression of factor 2 on factor 1, the regression residual, being orthogonal to factor 1, can become the modified factor 2, to be used to as a control in the cross-sectional method. When there is no measurement error in factor 1, this method works. However, given that factor 1 contains a measurement error, the following proposition shows that the residual of factor 2 is still correlated with factor 1. Thus, controlling this residual leads to the same factor-contamination problem as in Proposition 3.

Proposition 4: Assume that factors 1 and 2 are correlated. Let $u_{12} = f_2 - \frac{cov(\tilde{f}_1, f_2)}{var(\tilde{f}_1)}\tilde{f}_1$, then the residual value of regressing factor 2 on non-traded factor 1 with measurement error. Then the correlation between u_{12} and f_1, is non-zero.

The proof is in Appendix C. In sum, we have shown in this section that several classical cross-sectional methods to construct the FMP for a non-traded factor are not desirable, as long as there is another risk factor that is correlated with the non-traded factor. In Section 3.3, we propose a solution to this problem.

3.3. Proposed method

We propose to construct an FMP using a one-factor cross-sectional regression approach even if there are other correlated risk factors. For the cross-sectional approach, there are two passes. We present the rationale for using only one factor in each pass.

The rationale for the first pass is from equation (1). By running a single-factor regression, we extract the largest component of return that is correlated with the underlying factor. Even if the true model contains another risk factor, we show below that a two-factor time-series regression can not achieve the same goal. Specifically, with two factors f_1 and f_2, assume that the true regression model follows equation (19). Rewriting that equation, we obtain:

$$\boldsymbol{R^i} = \alpha^i + \beta_1^i \boldsymbol{f}_1 + \beta_2^i \boldsymbol{f}_2 + \boldsymbol{\varepsilon}^i = \alpha^i + \left(\beta_1^i + \beta_2^i \frac{cov(f_1, f_2)}{var(f_1)}\right)\boldsymbol{f_1} + \beta_2^i\left(\boldsymbol{f_2} - \frac{cov(f_1, f_2)}{var(f_1)}\boldsymbol{f_1}\right) + \boldsymbol{\varepsilon}^i.$$

Redefine $\boldsymbol{f_1^*} = \boldsymbol{f_1}$, $\boldsymbol{f_2^*} = \boldsymbol{f_2} - \frac{cov(f_1, f_2)}{var(f_1)}\boldsymbol{f_1}$, $\beta_1^{i*} = \beta_1^i + \beta_2^i \frac{cov(f_1, f_2)}{var(f_1)}$ and $\beta_2^{i*} = \beta_2^i$.[12] The regression becomes

$$\boldsymbol{R^i} = \alpha^i + \beta_1^{i*}\boldsymbol{f_1^*} + \beta_2^{i*}\boldsymbol{f_2^*} + \boldsymbol{\varepsilon}^i \tag{23}$$

From the construction, $\frac{cov(f_1, f_2)}{var(f_1)}\boldsymbol{f_1}$ is the largest component of factor 2 that is correlated with factor 1, and $\boldsymbol{f_2} - \frac{cov(f_1, f_2)}{var(f_1)}\boldsymbol{f_1}$ is the uncorrelated component. Hence, $\beta_1^{i*}\boldsymbol{f_1^*}$ (which is the same as $\beta_1^{i*}\boldsymbol{f_1}$) is the largest component of return that is correlated with the factor f_1, and the remaining parts $\alpha^i + \beta_2^{i*}\boldsymbol{f_2^*} + \boldsymbol{\varepsilon}^i$ characterize the mispricing component and the error in equation (1). It is well known

[12] Hence, $\boldsymbol{f_1^*}$ and $\boldsymbol{f_2^*}$ are uncorrelated.

that the loading of factor 1 in equation (23) is equivalent to the slope coefficient of the single factor regression,

$$\boldsymbol{R^i} = \alpha^i + \beta_1^{i*}\boldsymbol{f_1^*} + \boldsymbol{\varsigma}^i = \alpha^i + \beta_1^{i*}\boldsymbol{f_1} + \boldsymbol{\varsigma}^i, \tag{24}$$

with $\boldsymbol{\varsigma}^i = \beta_2^{i*}\boldsymbol{f_2^*} + \boldsymbol{\varepsilon}^i$.

From the above analysis, if two factors are correlated, the factor loading of running a single factor regression (β_1^{i*}) is different from the factor loading of running a two-factor regression (β_1^i). Hence, the largest component of return that is correlated with the factor f_1 ($\beta_1^{i*}\boldsymbol{f_1^*}$) can be extracted from by running a single-factor regression. This cannot be achieved when we run the two-factor regression (we obtain $\beta_1^i\boldsymbol{f}_1$instead). Moreover, the FMP are constructed using the factor loadings. Cross-sectional regression with different loadings as independent variables can lead to different FMPs. Thus, the FMP constructed by two-factor model might not extract the largest component of the underlying factor.

We depend on the following assumption to show the validity of the second-pass regression using the one-factor model.

Assumption: When there are large enough numbers of assets, the factor loadings of two uncorrelated factors are also uncorrelated in cross-section. That is, for factors 1 and 2 ($\boldsymbol{f_1^*}$ and $\boldsymbol{f_2^*}$) in equation (23), we have

$$\overline{cov}(\beta_1^{i*}, \beta_2^{i*}) \to 0, \tag{25}$$

when the number of test assets *N* converges to infinity.

In equation (25), β_1^{i*} and β_2^{i*} represent the sensitivity of the asset i on $\boldsymbol{f_1^*}$ and $\boldsymbol{f_2^*}$, respectively. The assumption implies that for any typical asset that is highly sensitive to factor 1, its sensitivity to factor 2 does not necessarily have to be high ($\overline{cov}\left(\beta_1^{i*}, \beta_2^{i*}\right) > 0$ in this case) or low ($\overline{cov}\left(\beta_1^{i*}, \beta_2^{i*}\right) < 0$ in this case). When the number of assets is large, the group of assets with high factor-1 sensitivity could contain both high factor-2 sensitive and low factor-2 sensitive assets. For example, suppose consumption growth ($\boldsymbol{f_1^*}$) and value ($\boldsymbol{f_2^*}$) are uncorrelated risk factors for stocks. We should observe both value and growth firms no matter whether the firm is cyclical or defensive (i.e., if two factors represent uncorrelated risks, there should be firms that can represent any combination of the risk factors, when number of firms is large). In contrast, if consumption growth and market excess return are highly correlated risk factors, then the firms with high/low loading on market return should also have high/low loadings on consumption growth. In Section 5, we find that the average absolute correlation among factor loadings of orthogonalized (uncorrelated) factors across all individual stocks is only 8%, and the maximum absolute correlation is about 20%. This seems to be consistent with our assumption.

From equations (23) and (24), when we run the first-pass regression only using consumption growth alone, we produce the same factor loadings by controlling the component of the value factor that is uncorrelated with the consumption growth. In the second pass, the true model should be

$$\boldsymbol{R_t}' = \alpha_t + \lambda_{1t}\widehat{\boldsymbol{\beta}}_1^{*\prime} + \lambda_{2t}\widehat{\boldsymbol{\beta}}_{\boldsymbol{2}}^{*\prime} + \boldsymbol{\nu_t},$$

where $\lambda_{1t} = f_{1t}^* - E(f_1^*) + \gamma_1^* = f_{1t} - E(f_1) + \gamma_1$ and $\lambda_{2t} = f_{2t}^* - E(f_2^*) + \gamma_2^*$.

When equation (25) is imposed on uncorrelated factors $\boldsymbol{f_1^*}$ and $\boldsymbol{f_2^*}$, and we run one-factor cross-sectional regression as follows,

$$\boldsymbol{R_t}' = \alpha_t + \lambda_{1t}\widehat{\boldsymbol{\beta}}_1^{*\prime} + \boldsymbol{\eta_t}, \tag{26}$$

then Proposition 5 shows that the estimated coefficient in this regression is a linear transformation of f_1.

Proposition 5: Assume that equation (25) holds. If the measurement error in factor 1 is uncorrelated with asset returns, regression residuals, and both factors, the regression residuals are also uncorrelated with both factors, the betas are uncorrelated with the cross-sectional regression errors, and the estimated coefficient from model (26) is

$$\hat{\lambda}_{1t} \to c(f_{1t} - E(f_1) + \gamma_1),$$

where $c = (var(f_1) + var(\varepsilon_{f_1}))/var(f_1)$, as sample period T and the number of test assets N converge to infinity.

The proof is in Appendix C. Following the same example above, the estimated parameter in the second-pass regression is a good proxy for the consumption growth risk as long as the factor loadings of the single-factor regression (with consumption growth) is uncorrelated with the factor loading of the value factor (i.e., the FMP is not contaminated by other factors).

Note that the method we propose is the feasible FMP construction approach that can exclude the effect of factor 2. If we intend to exclude the effect of factor 2 by controlling factor 2, as shown in Section 3.2, then the FMP is still contaminated by factor 2.

3.4. Issues with the time-series approach

Up to now, we have mainly discussed the cross-sectional method. We could also construct an FMP using the time-series approach. Following regression (13), the time-series method also constructs an FMP for each factor independently. However, this approach remains exposed to several issues.

The classical time-series method only requires a small number of assets. This can lead to a breakdown of the assumption in equation (25), since that equation is only reasonable for a large number of assets. Suppose the number of test assets (denoted by N, which is also the number of independent variables in regression (13)) is large. In real data, the sample size (T) in regression (13) is finite (for macroeconomic factors, the highest frequency is monthly, leading to roughly 600 months over 50 years. So $T = 600$ in this case). Thus, there is an overfitting or an overidentification issue if N is close to or larger than T.

Even if we have a large enough sample size, we show in the proposition below that an FMP created by the time-series method can represent a combination of two risk factors.

Proposition 6: Let the true model be equation (23). The regularity assumptions about measurement errors and regression residuals are also satisfied as in Proposition 5. We estimate the coefficient in the following regression:

$$\tilde{\boldsymbol{f}}_{\boldsymbol{1}} = a + \boldsymbol{b}\mathfrak{R} + \boldsymbol{u}, \tag{27}$$

where $\mathfrak{R} = [\boldsymbol{R}^1, \cdots, \boldsymbol{R}^N] = [\boldsymbol{R}_{\boldsymbol{1}}; \cdots; \boldsymbol{R}_{\boldsymbol{T}}]$, where ";" is the operator that stacks row vectors,[13] and construct the FMP as $\frac{1}{N}\widehat{\boldsymbol{b}}'\boldsymbol{R}_t$. When sample size and number of test assets both converge to infinity,

[13] Recall that $\boldsymbol{R}^i$ is a T by 1 column vector, $\boldsymbol{R}_{\boldsymbol{t}}$ is a 1 by N row vector, and $\mathfrak{R}$ is a T by N matrix. For the given definition of ";", the two expressions of $\mathfrak{R}$ are equivalent.

the FMP constructed by the time-series method converges to $\frac{1}{N}\boldsymbol{\beta}_1^* \boldsymbol{V}^{-1}(a+\boldsymbol{\beta}_1^* \boldsymbol{f}_{1t} + \boldsymbol{\beta}_2^* \boldsymbol{f}_{2t}) var(f_1)$.

The proof is in Appendix C. As we can see from the Proposition 6, the FMP still contains the other factor unless $\bar{E}(\frac{1}{N}\boldsymbol{\beta}_1^* \boldsymbol{V}^{-1}\boldsymbol{\beta}_2^*) \to 0$ as N goes to infinity. In a simplified scenario, $\boldsymbol{V}^{-1}$ is an identity matrix. The expected value becomes $\bar{E}(\beta_1^{i*}\beta_2^{i*}) \to 0$. Since $\bar{E}\left(\beta_1^{i*}\beta_2^{i*}\right) = \bar{E}\left(\beta_1^{i*}\right)\bar{E}\left(\beta_2^{i*}\right) + \overline{cov}(\beta_1^{i*}, \beta_2^{i*})$, even if the factor loadings are uncorrelated (equation (25)), we still require that the cross-sectional average of factor loadings have mean 0. This is unlikely to be satisfied for most of the factors.

3.5. Two-stage method for risk-premium estimation

A fundamental goal in constructing an FMP is to test whether the non-traded factor is associated with a risk premium. Since the FMP is an excess return, Shanken (1992) shows that its average value is the risk premium (we call it one-stage method). Another method is to refit a cross-sectional regression to estimate the risk premium (following Connor, Korajczyk, and Uhlaner 2015). . In this section, we show that, although these two methods are asymptotically the same in a large sample, the finite sample error that stems from the measurement error becomes smaller when we reapply a cross-sectional regression to test for a risk premium. Given that we apply the cross-sectional regression in the first stage to construct FMP, and apply the cross-sectional method again to estimate risk premium in the second stage, we call this the two-stage method.

Proposition 7: (1) As T converges to infinity, the average value of the coefficients in regression (26) converges to the true risk premium times a constant: $\frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_{1t} \to c\gamma_1$. (2) When T is finite, in the average value of the coefficients ($\frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_{1t}$), the finite sample estimation error that comes from factor measurement error (ε_f) is of order $O(\frac{1}{\sqrt{T}})$. (3) With the two-stage method, the finite sample estimation error that comes from the factor measurement error (ε_f) in the estimated risk premium is of order $O(\frac{1}{T})$.

The proof is in Appendix C. If the measurement error for some non-traded factor is very large, it can lead to a large and noisy component in $\frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_{1t}$, the estimated risk premium in the one-stage method. When we apply the two-stage method, the effect of the measurement error becomes much smaller, which can lead to a smaller estimation error in a finite sample. Balduzi and Robotti (2009) compare the one-stage method with the two-stage method (although the FMP is constructed by the time-series approach in their paper), and they find that the two-stage method is superior. We provide a possible explanation for this finding. Connor, Korajczyk, and Uhlaner (2015) propose a two-stage method to estimate the risk premium, because the method can reduce the EIV bias. In this section, we have shown that a similar two-stage method can also reduce the effect of the measurement error in the estimated risk premium.

Note that in the testing stage, we should incorporate all factors and FMPs to estimate risk premiums. In this proposition, we only prove the scenario with the single-factor model in the testing stage, but it is easy to extend it to the multifactor model. Also, note that two-stage method can only reduce the error from the measurement error. The error in the testing stage is still of order $O(\frac{1}{\sqrt{T}})$. Thus, the method can be particularly useful for those factors with large measurement errors, possibly including macroeconomic factors.

3.6. EIV issues and the IV approach

It is well known that cross-sectional regression analysis is subject to EIV bias in testing asset-pricing models. EIV also affects the correlation between factors and their corresponding FMPs. To see this, suppose that $\boldsymbol{\beta}$ is estimated with an error v, so that $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \boldsymbol{\nu}$. Then the true variance of the optimal FMP is larger than its estimated variance, that is,

$$\boldsymbol{w}' \boldsymbol{V}\boldsymbol{w} = \left[\boldsymbol{\beta}' \boldsymbol{V}^{-1}\boldsymbol{\beta}\right]^{-1} > \left[\widehat{\boldsymbol{\beta}}' \boldsymbol{V}^{-1}\widehat{\boldsymbol{\beta}}\right]^{-1}. \tag{28}$$

The first equality is established through plugging equation (9) into the objective function in expression (8). Because the variance of the FMP is reciprocal to the correlation between FMP and $\tilde{f}$ (equation (7)), the correlation between the estimated FMP and the factor is empirically higher than its true value. Therefore, a method (such as the OLS method) that delivers a maximally correlated portfolio may not be the optimal choice unless the EIV problem is corrected.

EIV is also presented in the sorting-by-beta method. To construct FMPs, we first estimate factor loadings (betas) for each asset, then sort assets by their betas and group the assets into portfolios by the sorted betas. FMPs are the difference between average returns of assets in the highest and the lowest beta groups. Because estimated betas contain estimation errors, the larger (smaller) betas are more likely to produce a positive (negative) estimation error. In an extreme case, where a major part of the estimated betas is an error, the sorting-by-beta approach is tantamount to sorting by error. Thus, even if the factor does price assets in the cross-section, the difference in the average returns between the assets in the highest and the lowest estimated beta groups may not represent the difference of their risk exposures to the factor. The FMP created by this approach might not be well correlated with the original factor. Hence, the sorting-by-beta method, like the cross-sectional approach, suffers from an EIV bias.

The IV approach can adjust for this issue. Assume that we want to test a K factor model or construct an FMP (in this case, $K = 1$). We divide the total sample into odd- and even-month subsamples. We run time-series regressions for the subsamples of odd and even months separately, thereby estimating independent odd- and even-month betas for each asset. With odd-month betas as IV betas and even-month betas as EV (evaluation variable) betas, we construct the matrices for betas of all assets: $\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}$ and $\widehat{\boldsymbol{B}}_{\boldsymbol{EV}}$, where $\boldsymbol{B}$ is the $N \times (K+1)$ matrix containing all the betas augmented by a vector of 1, that is, $\boldsymbol{B} = [\boldsymbol{1}, \boldsymbol{\beta}]$, where $\boldsymbol{1}$ is an N by 1 vector of 1s.

Then we calculate a second-pass cross-sectional IV regression. For each even month, we run a two-stage least-squares regression, and the estimated risk premium can be written as[14]

$$\widehat{\boldsymbol{\gamma}}_{\boldsymbol{t}} = (\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}{}' \, \widehat{\boldsymbol{B}}_{\boldsymbol{EV}})^{-\boldsymbol{1}} \widehat{\boldsymbol{B}}_{\boldsymbol{IV}}{}' \, \boldsymbol{r}_{\boldsymbol{t}}. \tag{29}$$

Here, $\boldsymbol{r}_{\boldsymbol{t}}$ is the excess return for even months and is an N by 1 vector. Correspondingly, for each odd month, we take the betas in the even months' subsample as the IVs and estimate the equation to obtain the risk premium.

When the error contains no factor structure (Section 3.1), Jegadeesh et al. (2019) shows that the IV approach can converge at the speed of $\sqrt{NT}$. When the error contains a factor structure (Section 3.2 and later), Jegadeesh and Noh (2014) shows that the IV approach can converge at the speed of e^{T}, while the classical OLS method can only converge at the speed of $\sqrt{T}$. Hence, in both cases, the IV approach can adjust for the EIV issue.

[14] We do not use the GLS-IV approach, because Roll and Ross (1994) find that only an OLS approach correctly and economically interprets the coefficient. In addition, the covariance matrix of individual assets is not invertible when the number of assets is much larger than the number of time periods.

3.7. Asset selection

We propose to use a large number of test assets (e.g., individual stocks/bonds) for the reason described in Section 3.3 (equation (24)). However, including more assets can create issues if they are not correlated with the underlying factor.[15] For example, Gospodinov, Kan, and Robotti (2019) find that the estimated coefficients in a cross-sectional regression can be large and significant even if the factor is not priced. The IV approach can be used to select well-correlated assets, because assets that are well correlated with the factor should have similarly estimated betas in the odd and the even samples. Thus, the sign of the IV and EV betas usually should be the same. However, assets that are not driven by the factors likely have very noisy returns, and thus the IV and EV betas would often have different signs. Therefore, we should select only assets that have $\hat{\beta}_{IV}{}' \hat{\beta}_{EV} > 0$, or retain only those assets with IV and EV betas having the same sign.

This selection criterion cannot escape the possibility that the IV and EV betas have the same signs due to the likelihood of random errors in the same direction. If this happens often, it will induce an attenuation bias in the estimated risk premium. However, our simulation results in Appendix A show that the bias in the IV method is small (similar to that of the IV approach without this adjustment).

For the time-series method, when we apply it to many test assets, it is also useful for selecting the best-correlated ones. In this paper, we incorporate a simple variable selection method—the least absolute shrinkage and selection operator (known as the "Lasso") and examine its effect on FMP construction empirically in Appendix B. Giglio and Xiu (2019) suggest the use of a large set of portfolio returns as test assets and apply PCA methods to construct FMP. In Appendix B, we also apply their approach.

3.8. A horse race among methods

Based on the previous section, we propose our method as follows. First, we apply a cross-sectional regression on a single-factor model to adjust for factor contamination. Second, we apply the IV method to mitigate the EIV issue. Third, we apply the factor-selection approach to select basis

[15] This is particularly important for the time-series approach, as basis assets are usually the characteristic sorted portfolio returns. If these characteristics are not correlated with the macroeconomic risk, the basis asset can be uncorrelated with the macro factor.

assets. Finally, we reapply the cross-sectional regression method for all factors and FMPs to estimate the risk premium. In the simulation in Section 5 and empirical work in Section 6, we will compare the proposed approach with various other existing methods. These methods include the classical time-series approach, the sorting by beta approach, and the cross-sectional OLS approach.

Lehmann and Modest (1988) introduce a weighted-least-squares method to construct the FMP. Instead of using the identity matrix as in the OLS approach, they suggest using the diagonal matrix consisting of the residual variances from the first-pass time-series regression as the weighting matrix.

Stein (1956) and James and Stein (1961) propose the shrinkage method to minimize Root Mean-Square Errors (RMSEs) when at least three parameters are being estimated. Their shrinkage beta can be written as

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{Stein}} = \left(1 - \frac{(N-3)}{\|\widehat{\boldsymbol{\beta}}_{\boldsymbol{\sigma}}^*\|}\right)\widehat{\boldsymbol{\beta}}_{\boldsymbol{new}} + \bar{\hat{\beta}}, \quad (30)$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\sigma}}^* = [\hat{\beta}_\sigma^1, \hat{\beta}_\sigma^2, \dots, \hat{\beta}_\sigma^N]$, and $\hat{\beta}_\sigma^i = \frac{\hat{\beta}^i - \bar{\hat{\beta}}}{\sigma^i}$, in which σ^i is the standard error of $\hat{\beta}^i$. Also, $\|\widehat{\boldsymbol{\beta}}_{\boldsymbol{\sigma}}^*\| = \sum_{i=1}^N (\hat{\beta}_\sigma^i)^2$, $\bar{\hat{\beta}} = \frac{1}{N}\sum_{i=1}^N \hat{\beta}^i$ or the mean of $\hat{\beta}^i$, and $\widehat{\boldsymbol{\beta}}_{\boldsymbol{new}} = [\hat{\beta}^1 - \bar{\hat{\beta}}, \hat{\beta}^2 - \bar{\hat{\beta}}, \dots, \hat{\beta}^N - \bar{\hat{\beta}}]$.

Then the Stein-adjusted risk-premium estimate is

$$\boldsymbol{\lambda_t} = (\boldsymbol{B}_{\boldsymbol{Stein}}^{'} \boldsymbol{B}_{\boldsymbol{Stein}})^{-1} \boldsymbol{B}_{\boldsymbol{Stein}}^{'} \boldsymbol{R_t}, \quad (31)$$

where $\boldsymbol{B}_{\boldsymbol{Stein}}$ is the $N \times (K+1)$ matrix containing all the Stein shrinkage betas $\widehat{\boldsymbol{\beta}}_{\boldsymbol{stein}}$, augmented by a vector of 1.

Stein's method can reduce the mean-squared error of the OLS estimator through the reduction of the standard error. Hence, the FMP constructed by this approach will be less volatile.

We will compare both the LM and Stein's approach with others in this paper. Note that for all these classical cross-sectional approaches, the convention is to run multifactor regression to create the FMP, which will lead to factor contamination.

4. Data

In this section, we describe the data and variables used in this paper. The summary of the descriptive statistics of these variables is reported in Table 1.

[Table 1 about here]

4.1. Stock return data

Monthly individual stock returns are from CRSP. The data start from January 1964 and run to March 2016 (627 months). Following the extant literature, we exclude stocks with prices less than 1 dollar or market capitalizations less than 6 million dollars. We also exclude stocks that have less than 60 continuous monthly returns. After these exclusions, 10,833 stocks remain in our sample; there are 2,850 stocks in an average month; the total observations are 1,784,351. The mean return of individual stocks over a risk-free rate (the one-month T-bill rate) is 1.012% per month, but the median is 0.44% per month, indicating that there are very large positive returns for individual stocks.

4.2. Explanatory variables

Four macroeconomic variables obtained from the Federal Reserve Bank of St. Louis Research Website (FRED) serve as our non-traded factors: (1) the growth in per capita consumption (DPCERAM1M225NBEA in FRED code), (2) the percentage change in the consumer price index (CPIAUCSL), (3) the percentage change in industrial production (INDPRO), and (4) the percentage change in the unemployment rate (UNRATE). Following Chen, Roll, and Ross (1986), we use innovations in these macroeconomic variables as factors. To measure innovations, we use the residuals from a first-order vector autoregression (VAR). It is also possible to use first differences as innovations. But as discussed by Boguth and Kuehn (2013), the first-difference method is a more conservative specification for risk exposures. Moreover, the results are robust to first differences (unreported). We also study factors for bonds, such as the default spread and term spread, downloaded from Robert Shiller's website. For traded factors, we downloaded from Kenneth R. French's website (excess market return, small-minus-large market capitalization, and high-minus-low book/market portfolio returns). To construct the time-series factor-mimicking portfolios, we obtained 25 portfolios formed on size and book/market, 10 industry portfolios from

French's website, and four bond returns, which include 1-year, 5-year, and 10-year treasury bond yields, and Moody's seasoned Baa corporate bond yield from FRED.

We also examine other consumption-related factors, including the CAY factor (the ratio of consumption to aggregate wealth proposed by Lettau and Ludvigson 2001) and the consumption volatility factor proposed by Boguth and Kuehn (2013). These factors are available from Martin Lettau's and Oliver Boguth's websites, respectively.

4.3. Corporate bond-return data

For corporate bonds, we use transaction records in the Trade Reporting and Compliance Engine (TRACE). TRACE provides corporate bond intraday trading prices, trading volumes, sell and buy indicators, and so forth. Our sample period is from August 2002 to June 2017. We follow the Bai, Bali, and Wen (2019) data-screening procedure and return-estimation approach. The monthly corporate bond returns are computed from the average quoted price at the end of the current month, accrued interest, and coupon payment for a month divided by the average quoted price at the end of the previous month or the beginning of the current month. A bond's excess return is the difference between its computed total return and the risk-free rate, where the latter is proxied by the 1-month Treasury bill rate. Our final sample consists of 331,728 observations, and the average cross-sectional excess return is 0.389%, which is comparable to the Bai, Bali, and Wen (2019) sample. We include only bonds that have at least 30 continuous monthly returns;[16] 6,421 bonds remain in our final sample.

As for the explanatory variables, we first consider the four non-traded macroeconomic factors. We subsequently add the default spread, the term spread, and the corporate bond market return. The default spread is the return difference between Moody's long-term corporate BAA-rated bonds and AAA-rated bonds. The term spread is the return difference between the 10-year

[16] The results are robust using different windows.

and 1-year treasury bonds. The monthly corporate bond market return is the equally weighted average of corporate bond returns in our sample.

5. Simulation

5.1. Simulation procedure

In this section, we examine, in the finite sample, the magnitude of the factor contamination of FMPs constructed using the methods described in Section 3 through simulations.[17] Following the same notation used in Sections 2 and 3, an observed factor ($\widetilde{f_t} = f_t + \varepsilon_{f,t}$) contains two parts. The first part, f_t, is the projection of the factor into the space of excess return (we call it the return-related component of the underlying factor, or the return-related factor), and the other part, ε_f , is the measurement-error term that is uncorrelated with excess return. The goal of the simulation is to examine the effectiveness of FMPs constructed by various methods to extract the factor in excess return (f_t) from an observed factor ($\widetilde{f_t}$). If it is effective, the FMP should be almost perfectly corrected with the underlying factor (f_t), and not be correlated with another uncorrelated (orthogonalized) factor ($f_t^{\perp}$).

We use the four macro factors and three Fama-French factors (FF, henceforth) in the return-generating process. For the macro factors, we construct FMPs by the method described in Section 3. Specifically, we use a single-factor IV approach with asset selection to construct FMP, and we call it IV approach for simplicity. We demonstrated in Section 3 (in the large-sample theories) and show in this section (in the finite sample) that this method delivers an FMP that does not contain measurement errors and does not suffer from factor contamination. Thus, these FMPs for macro factors are used to construct the return-related components of the macro factors in simulations. Since the three FF factors are traded factors, they contain no measurement errors. Their original factors are the same as the return-related factors.

[17] In addition to factor contamination, the existing method also suffers from EIV bias when there is an estimation error in factor loadings as well as in the basis asset selection. These issues are evaluated extensively in the literature of asset-pricing tests. Since the issues are similar for FMP construction, we defer the simulation on measurement errors of beta loadings to Appendix A. We find that the FMPs constructed by IV methods can yield risk premiums with a bias of less than 5%, but FMPs constructed by other methods suffer from severely biased estimations of the risk premium. See Table A1 in Appendix A.

For the return-related component of each macro factor, we orthogonalize the other factors to make them uncorrelated.[18] For example, if we want to construct the FMP for consumption growth, the other six factors are orthogonalized.[19] Orthogonalized factors are used for the data-generating process and for examining factor contamination. If the FMP of the consumption growth factor is correlated with orthogonalized control factors, there is factor contamination, because the FMPs contain the risk component that is in the orthogonalized control factors but not in the consumption growth factor. The vector of all factors (the return-related macro factor and other orthogonalized factors) is denoted by $f_t^{\perp}$. We calculate the mean, variance, and covariance (0 in this case) of the factors in $f_t^{\perp}$. These values are used for simulations.

We run the time-series regression for returns of each asset (R_t^i) on the orthogonal risk factors ($f_t^{\perp}$) to obtain beta loadings (B^i) and residual ε_{it}. Similar to the orthogonalized factors, we orthogonalize the beta loadings of the seven factors, which are expressed as $B_i^{\perp}$. We orthogonalize the beta to satisfy the assumption in Section 3.3 (equation (25)), in which we assume that the cross-sectional correlation between loadings of two factors converges to zero when N approaches infinity, if the two factors are uncorrelated. In the real data, we find that for each macro factor, the cross-sectional correlations between its factor loading and the loadings of other orthogonalized factors are small. On average, it is 0.08. Therefore, the assumption is close to being true in the real data. We still orthogonalize loadings to make the correlations exactly equal to 0. The mean, variance, and covariance (0 in this case) of the loadings $B_i^{\perp}$ are calculated for simulations.

With these orthogonalized factors and loadings, we proceed to the data-generating process of asset returns in simulation. We first simulate orthogonalized factors using Monte Carlo simulations and keep the mean, variance, and covariance of the factors the same as those from the data. For all simulated orthogonalized factors, we subtract the mean of these factors and then add pre-specified true premiums λ_0 set equal to the observed average risk premiums from Chen, Roll,

[18] Note that we orthogonalize only return-related factors (i.e., FMP constructed by the IV method for macro factors and the original three FF factors).

[19] We adopt Gram-Schmidt's orthogonalization process. Specifically, we regress the first control factor to consumption growth and use the residual as the proxy for the orthogonalized first control factor. Then we regress the second control factor on the orthogonalized first control factor and the consumption growth factor, and we use the residual term as the proxy for the orthogonalized second control factor. Following the same procedure for other control factors, the resulting factor matrix has seven columns, and every column is orthogonal to the other columns.

and Ross (1986) and Chen and Kan (2004). The resulting factor is $f_t^{true} = f_t^{\perp} - \overline{f^{\perp}} + \lambda_0$ (note that we create the ex post factor following Shanken (1992) and Jegadeesh et al. (2019)). We also generate the factor loadings ($B_i^{\perp}$) of all stocks for each factor from a multinomial normal distribution by keeping the same mean, variance, and covariance from the real data. The simulated return is computed as $R_{it}^s = f_t^{true} B_i^{\perp} + \varepsilon_{it}^s$, where ε_{it}^s is simulated from a normal distribution by keeping the same mean and variance of regression residual for stock i, ε_{it} from the real data. We have simulated the returns for each individual stock and the FF 25 size and book-to-market portfolios.

Note that the orthogonalized factors are only used for data-generating process, but these factors are not observable. Instead, the four macro factors and three FF factors are observed. The return-related components of these factors are correlated, and there are measurement errors for macro factors. Given that only observed factors are used to construct the FMP, we need to simulate them. Since the three FF factors have no measurement errors, they are observed factors. The four macro factors contain measurement errors. Hence, the simulated observed factor is constructed by adding to the simulated return-related factor an error term in simulation: $f_t^s = f_t + v_t$, where v_t is extracted from a normal distribution with variance equal to the variance of the return-related factor in the real data.[20] The return-related factors, f_t (factors not being orthogonalized), are simulated from a multinomial normal distribution with the same mean, variance, and covariance as those for the return-related factors in the real data.

After simulating returns and observed factors with measurement errors, we apply various methods to construct FMPs for each macro factor. For the IV method (denoted FMP_IV), we first run single-factor regressions. Factor contamination occurs if the FMPs capture components that are not in its underlying factor but are found in other factors. We iterate the simulation 1,000 times and report the summary statistics of these correlations in Table 2.

5.2. Simulation results

[20] From the real data, the correlation between FMP_IV (IV method to create FMP) and the observed factor is close to 0.5. If FMP_IV has no factor contamination, the correlation of 50% implies that the variance of the return-related factor and the measurement error is equal. Moreover, the results are robust if the variance of the measurement error is 5 times larger than that of the return-related factor.

Panel A in Table 2 reports the correlation between FMPs with their corresponding return-related factors (f_t). We do not examine the correlation between the FMP and the observed factor ($\widetilde{f_t}$), because the correlation depends on the variance of the measurement error. Instead, the correlation between the FMP and the return-related factor can be as large as 1. Taking consumption growth as an example, the averaged correlation between FMP_IV (IV method to create FMP, and we use this convention to denote the construction method of the FMPs for the rest of the paper) and the true return-related component of the consumption growth factor is 0.999. Hence, FMPs constructed by the single-factor IV method have a nearly perfect correlation with the return-related component of the underlying factors, thereby suggesting that the IV approach has almost no factor contamination. The correlation between FMP_OLS and the return-related component of the consumption growth factor is 0.798. FMP_Stein has the same correlation with the return related factor as FMP_OLS because the Stein method only adds a scaling effect to the FMP_OLS. The averaged correlation between the FMP_LM and the return-related factor is 0.841, slightly higher than that for the FMP_OLS. These cross-sectional methods (such as OLS, LM, and Stein) could not yield FMPs that have close-to-perfect correlations with return-related factors because of the factor contamination from using multiple regression in the cross-sectional approach. Using univariate regression with the asset selection method, the IV method alleviates these problems. The correlation between FMPs constructed by the sorting-by-beta method and the return-related component of the consumption growth factor is 0.704. The time-series approach with the 25 FF portfolios as basis assets also has low correlations with return related factors. These two methods suffer the same contamination issues as for the cross-sectional approach; thus the low correlations are expected.

[Table 2 near here]

Panels B and C of Table 2 present the correlations between the FMPs for each macro factor and other orthogonalized factors.[21] In the first data column, we present the correlations for

[21] The reason for not examining the correlation between the FMP of one macro factor and the return-related component of another macro factor (which is not orthogonalized) is discussed below. Since the return-related factors are correlated, for any macro factor, the correlation between its FMP and other return-related factors is not 0. But the nonzero correlation is capturing component of another factor that is also part of the underlying macro factors. For example, let consumption growth and market returns be correlated risk factors. If the FMP of the consumption growth factor is perfectly correlated with the return-related component of the factor (implying that the FMP is not contaminated by other factors), it is still correlated with market return. However, this nonzero correlation does not indicate that the FMP contains the risk that is not generated from the consumption growth. Suppose that the FMP

consumption growth factors. The results for other macro factors are shown in consecutive columns, and they are qualitatively similar. The FMPs of consumption growth (CG) have six correlation coefficients with six other orthogonalized factors. We report the maximum value and the average value across the six correlation coefficients in each simulation in Panels B and C, respectively. Panel B lists the average values of the maximum value across 1,000 simulations. Panel C lists the mean values of the recorded average value across 1,000 simulations. Consistent with the analysis in Section 3, we find that the IV method generates almost zero factor contamination. In comparison, all FMPs constructed by the other methods lead to factor contamination. For example, the maximum correlation between FMP_SB for consumption growth and the other six control risk factors can be 0.441.

Overall, these simulation results testify that the commonly used FMP construction methods (sorting-by-beta, time-series, and multivariate cross-sectional) suffer from factor contamination. The method we propose (the IV method in a univariate cross-sectional analysis with asset selection) yields almost perfect correlation with the return-related component of the underlying factor and contains the minimal factor contamination.

6. Examining the FMP in the data

In this section, we evaluate whether FMPs constructed by various methods empirically satisfy the three FMP selection criteria. We focus on the FMPs for the four macro factors and augment this by examining the FMPs for traded factors (the three FF factors) for comparison.[22] To examine these FMPs, we explore three criteria:

(1) FMPs are correlated with the underlying factors,

(2) FMPs are correlated with the systematic risk of returns, and

(3) FMPs explain the cross-sectional of mean returns.

of consumption growth is correlated with the orthogonalized market return, then the FMP captures the risk not belonging to the consumption growth. In this case, the FMP is contaminated.

[22] We also test the robustness by using alternative traded factors, such as Carhart 4 factors, FF 5 factors, and FF 6 factors. We find these results are robust to these specifications. In some cases, we must drop the market factor to achieve robust results due to the strong correlation between the FMP of consumption growth and the market factors. Theoretically, consumption growth and market factor should present similar risks.

Intuitively, an FMP should represent the risk of the underlying factor, as shown in the constraint of the least-mispriced theory. Besides, if the underlying factor is a true risk factor, the systematic risk should be correlated with FMP returns. If the factor can price assets in cross-section, the FMP should generate significant risk premiums. Our three criteria are studied sequentially in Sections 6.1 to 6.3. We provide some robustness checks by adding several other consumption-related factors in Section 6.4. We provide the rationale for these criteria in Appendix D.

6.1. First criterion: Correlation with underlying factors

An effective FMP should correlate significantly with its underlying non-traded factor, so that the FMP contains similar risk information. In unreported results where we estimate correlation for the whole sample period, all of the correlations between FMPs and their underlying factors are significant at the 1% level, thereby suggesting that all the FMPs satisfy this criterion.[23] However, the observed significance may be driven by the large sample size and possibly by non-stationarity. Correlations between FMPs and the underlying risk factors could be spuriously inflated by time variations in the mean returns. To examine this possibility, we divide the full sample into five subsamples roughly by decades and compute subsample correlations. Table 3 lists the average value of the correlation coefficients in the five subsamples and the number of significant correlation coefficients in these subsamples at the 1% significance level.

Although the correlations for cross-sectional and sorting-by-beta methods are all statistically significant, their magnitudes vary across FMPs. Notably, the correlation between FMPs constructed by the IV method (FMP_IV) and the underlying factors are smaller than those from the OLS method. The correlation between FMP_IV for consumption growth and the underlying consumption growth factor is 0.411, but the correlation between FMP_OLS for consumption growth and the underlying consumption growth factor is 0.636. However, as discussed in Section 3.6, OLS-based FMPs are subject to substantial EIV issues, overstating sample correlations with its factors.

[23] Although there is no well-accepted threshold of correlations that an effective FMP should satisfy, the correlation should be significantly different from 0 to avoid the "useless factor" problem (Barillas and Shanken, 2018). The requirement of significant correlation avoids selecting the wrong model setting for the time-series approach. For example, if the R square from a time-series is very low (e.g., 0.05), then the basis assets have virtually no relation to the factors.

For the time-series approaches, we find that most of them are not significantly correlated with the factors in most subperiods. For instance, the correlation is only significant in one subperiod for the consumption-growth FMP. Since the different sets of basis assets are likely to be different in their correlation with underlying macro factors, the choice of the basis assets is essential. For example, unexpected inflation can be related to bond returns; therefore, basis assets that have bonds as a component (the Lamont approach) are likely to be correlated with this factor. Our result confirms this conjecture: for four out of five decades, the correlations between FMP_time-series and CPI are significant. Correlations between time-series FMPs and the underlying factors are also lower than those for cross-sectional FMPs. For example, for the industrial production factor, the correlation of a time-series approach is 0.293, whereas it more than doubles to 0.789 for a cross-sectional approach (OLS). Such a striking difference emphasizes again how FMPs can be method dependent. Compared with FMPs for non-traded factors, FMPs for the FF factors have strong correlations with their underlying risk factors in most cases. Due to similarities across various approaches, FMPs are likely to be less advantageous for traded factors.

[Table 3 around here]

6.2. Second criterion: Relation with the Covariance of asset returns

If an FMP represents a risk factor, it should be correlated with the systematic risk of returns. Following Pukthuanthong et al. (2019), we test whether the FMP is related with the covariance of asset returns. We apply the asymptotic approach of Connor and Korajczyk (1988) (CK) to extract 10 principal components from the equities return series. The principal components of the covariance matrix of returns represent the systematic part of asset returns. We then compute canonical correlations between the 10 CK principal components and the factor candidates and test the significance of these canonical correlations by the chi-squared statistic.

We examine this criterion for the FMPs that are constructed by various methods and also for their corresponding original factors as a comparison. The four original macroeconomic factors are those we have already considered above: CG, CPI, IP, and UE. The original traded factors are the three FF factors (MKT, SMB, and HML). To examine this criterion, two conditions must be satisfied. We assume that an FMP strongly satisfies this criterion if it (1) is significantly related to any canonical variate in all decades or (2) has a mean *t*-statistic in the second row of each panel in

Table 4 exceeding the one-tailed, 2.5% cutoff based on the chi-squared value, and (3) also has an average number of significant decade *t*-statistics exceeding 1.75 (bottom row of each panel.)[24]

[Table 4 about here]

Notably, the four original macro factors do not pass, whereas the three FF factors pass this second criterion. FMPs constructed by all the cross-sectional methods and the sorted beta method satisfy this criterion. For FMP_time-series, none of the macro factors pass the second criterion.

6.3. Third criterion: Risk-premium estimation using FMPs

If an FMP can price assets, it should imply a significant risk premium. This section compares the extent to which various construction methods for FMPs produce different risk-premium estimates.

Cross-sectional approaches

Table 5 reports risk-premium estimates using IVs and other cross-sectional methods for obtaining FMPs. We select stocks whose betas in odd and even months have the same signs in order to select assets that are well correlated. Only 463 (about 4.6%) of more than 10,000 stocks are not used to construct any FMPs, with each factor using about 6,000 stocks.

The results show that risk premiums estimated by different methods often have dramatic disagreements in both magnitude and significance. For example, the risk premium for consumption growth is 0.066 (*t*-value = 2.202) using OLS, while it is 0.164 (*t*-value=3.238) using IV. Even though LM and Stein partly resolve the OLS downward bias, the risk premiums for consumption growth for these two methods are still much smaller than those for IV. With OLS, LM, and Stein, the risk-premium estimates for industrial production and unemployment are negative and positive, respectively, which are opposite to theoretical prediction[25] and contrast markedly with the IV estimates. The risk premium for the unemployment rate should be negative, because stocks with a positive unemployment beta can be viewed as hedging to economic downturns. Stocks with a

[24] Pukthuanthong et al. (2019) require an average number of significant decade *t*-statistics exceeding 2.5 from 10 factor candidates. We have seven factor candidates, thus 1.75 is from using the same proportion as theirs. The reason is as follows: "This is a conservative threshold to ensure we do not miss a true factor at our necessary condition stage. We focus on the significant canonical correlations, rather than all canonical correlations, because insignificant CCs imply that none of the factors matter, so using them would be over-fitting."

[25] A similar result is obtained by Giglio and Xiu (2019). As a robustness check, we drop the FMP for the industrial production factor since it is not significant, and we find that the results for the other FMPs are virtually unaltered.

negative unemployment beta are riskier, because the returns for these stocks decrease during periods of high unemployment.

For comparison, we estimate risk premiums for the three FF traded factors. The signs and significance levels for risk-premium estimates do not vary much across methods. Hence, EIV seems to be less of a problem for traded factors.

As revealed in Table 5, the IV-based risk premium of macro factors are still significant, even after including the three FF factors; this is not true for the other FMP construction methods.[26] Moreover, the risk premiums for the FF factors are virtually the same with any of the FMPs added into the estimation. This is somewhat curious, because it suggests that the FF factors do not contain the same information as the macroeconomic factors. One might very well wonder what risks the FF factors do represent.

Time-series approach

Under the time-series approach, we perform risk-premium estimation with FMPs constructed by the Lamont (2001) method. Compared with the IV and OLS approaches, the estimated risk premium for consumption growth is negative and insignificant, while the risk premium for industrial production is significantly negative. However, the time-series method works well for the three FF factors. Specifically, it produces a similar statistical significance for the risk-premium estimates of the three factors.

As discussed earlier, a time-series construction method for FMPs can be effective if the basis assets are correlated with the factors (Lewellen, Nagel, and Shanken 2010). The basis portfolios in Lamont (2001) include a market return, which is one of the three FF factors. Lamont's industry-sorted portfolio returns might very well be correlated with the MKT, SMB, and HML factors. Thus, the Lamont method can be an effective FMP construction method for Fama-French risk factors, although FMP is not normally applied to traded factors. Correlations between macro factors and Lamont basis assets are low; thus the Lamont method is probably not very effective for FMP construction of macro factors.

[26] Note that a negative risk premium for unexpected inflation is to be expected (see Boudoukh and Richardson 1993); high unexpected inflation has a downward impact on stock prices. Stocks with a positive inflation beta hedge inflation risk; they have higher returns in periods of high inflation, while stocks with a negative inflation beta are riskier, because they decrease in value during periods of high inflation.

Sorting-by-beta approaches

For the sorting-by-beta method, we estimate betas from time-series multivariate regressions and then sort the betas for each factor into 10 equally weighted deciles. We then construct FMPs as the average return in the highest decile minus the average return in the lowest percentile (high-minus-low, or HML). The FMPs are used as factors to estimate risk premiums, which are reported in the table. For macro factors, the estimated risk premiums are insignificant, especially when the three FF factors are included. This is consistent with our conjecture that the FMP constructed by the sorting-by-beta method can attenuate the variation in returns among stocks with different sensitivities to the factor when the measurement error in beta is large.

[Table 5 about here]

6.4. Robustness with respect to other consumption-related factors

Previous literature finds little relation between asset returns and consumption-based factors, probably because of the large noise in consumption growth.[27] However, we find that the risk premium for consumption growth is significant when using its associated FMP. Rather than focusing on the measurement problem of consumption growth, Lettau and Ludvigson (2001) derive a conditional consumption CAPM, which can explain average stock returns in the cross-section, using the consumption-to-wealth ratio (CAY) as a conditioning variable.

Table 6 reports estimated risk premiums when adding the CAY factor.[28] In all but one specification, the CAY factor is associated with a negative risk premium and is always insignificant. The consumption growth factor has a significantly positive risk premium for the IV FMP and the OLS FMP and is weakly significant at 10% for sorting beta. This is consistent with our earlier results that use four macroeconomic factors. Following Lettau and Ludvigson (2001), we include an interaction term between CAY and consumption growth; its risk premium is insignificant. CG remains significant at 5% for IV and at 10% for OLS and sorting beta. When we

[27] Indeed, we also find that the risk premium for consumption growth is insignificant if we use raw consumption growth rather than its FMP (unreported).

[28] In Appendix E, Table E1, we report analogous results using all four macro factors (CG, CPI, IP, and UE). The IV results are robust to this alternate specification. As shown there, results for time-series-based approaches are not robust. Correlations with underlying non-traded factors differ dramatically and so do their risk-premium estimates.

include the three FF factors, CG is significant at 5% for IV but is not significant for other approaches.

[Table 6 about here]

Boguth and Kuehn (2013) find that consumption volatility, supposedly a proxy for macroeconomic uncertainty, is also a source of risk and has a negative risk premium. In unreported results, we also add consumption volatility as a control variable. We find that the risk premium of the consumption volatility factor is negative but insignificant in all specifications across all FMP construction methods. In contrast, the risk premium for consumption growth is still significant. Therefore, using FMPs of consumption growth, our results confirm that consumption growth is a robust risk factor that can explain the cross-sectional stock returns conditional on other consumption-related factors.

Our overall conclusion for US equities is that FMPs constructed by the IV approach satisfy all our criteria. Moreover, the IV-based FMPs dominate FMPs constructed by other methods in producing larger and more significant risk-premium estimates. With the IV method, consumption growth, inflation, unemployment rate, and the three FF factors can explain cross-sectional stock returns.

7. Test risk premiums in the corporate bonds market by using FMPs

Bond returns are associated with macroeconomic factors, because bonds are related to firms' fundamentals that are affected by the business cycle (Ludvigson and Ng, 2009). Fama and French (1993) propose two non-traded bond factors, the default and term spread. Gebhardt, Hvidkjaer, and Swaminathan (2005) find that the default spread significantly explains cross-sectional bond returns even after controlling for bond characteristics, such as duration and rating. In contrast, Bai, Bali, and Wen (2019) find that such attributes as value-at-risk and rating dominate default and term. Bessembinder et al. (2008) suggest that a broad bond market return, unexpected GDP growth, and unexpected inflation explain excess abnormal bond returns. Following these papers, we evaluate the FMPs for four macroeconomic factors, a broad bond market return (MKT_B), the default spread (DS) and term spread (TS).

We use various FMP construction methods: IV, OLS, LM, Stein, time-series, and sorting-by-beta. Following the criteria, we first analyze correlations between FMPs with their underlying factors (first criterion) and with the principal components of the covariance matrix of individual corporate bond returns (second criterion). The results are similar to those for equities. Panel A in Table 7 shows that all the FMPs have strong correlations with their underlying factors; thus, all approaches satisfy the first criterion. Unlike the previous test of the first criterion, where we presented correlations for each decade, we present the correlation of FMPs and underlying factors for the whole sample period due to the shorter time series (from 2002 to 2017).[29] Panel B examines whether the FMP is related with the covariance of bond returns. FMPs constructed by all approaches except the time-series approach pass this criterion. We focus only on one criterion (the *t*-statistic of significant canonical correlation), because our sample period for bonds is only a decade. For the time-series approach, only consumption growth and shock in CPI pass.

Panel C presents the results of estimating the risk premium for corporate bond returns in a similar vein as that for equity returns. The time-series method (Lamont approach) produces significant risk premiums. For IV-based FMPs, consumption growth, industrial production, bond market return, and default spread pass and have the signs of risk premiums consistent with the results from stock returns. Risk premiums associated with Stein, LM, or OLS FMPs are insignificant. Interestingly, the time-series method (the Lamont approach) produces significant risk premiums but with counterintuitive signs for consumption growth, industrial production, and unemployment. Theoretically, positive consumption growth and industry production shocks (unemployment shocks) are associated with strong (weak) firm fundamentals, suggesting positive (negative) bond returns. Since Lamont (2001) uses stock portfolio returns as basis assets to create time-series FMPs, the negative sign might reflect the negative correlation between stock returns and bond returns. In an untabulated result, we use portfolios constructed by bond returns as the basis assets with the time-series approach. The risk premiums of the macro factors above are positive but insignificant when we control for other traded factors.

[Table 7 about here]

[29] Our sample period is consistent with that in Bai, Bali, and Wen (2019).

8. Conclusion

A voluminous literature applies FMPs to convert non-traded factors into their traded versions and then to estimate risk premium. However, no studies summarize and demonstrate that indeed, there are quite a few ways to construct FMPs, and each methodology has pros and cons for testing the underlying factors. The performance of FMPs depends on the way we construct FMPs and the basis assets we use to construct them. Our paper proposes a new economic explanation of FMPs, delves into the issues for existing methods, and offers a new method for FMP construction. We also examine FMPs using three necessary conditions.

Simulation results show that all other existing methods suffer factor contamination, while our method has almost no factor contamination. Empirically, we apply our IV method to estimate risk premiums for a series of non-traded factors. We first construct FMPs from four classical macroeconomic factors using the IV approach and find that the non-traded versions of consumption growth (CG), CPI, and the unemployment rate (UE) have significant risk premiums in the stock market. The conclusion cannot be obtained from the same factors constructed by the extant mimicking-portfolios approaches. We also apply our method to estimate risk premiums in the corporate bond market and find that consumption growth and industrial production price corporate bond returns with positive risk premiums.

References

Adrian, T., Etula, E., and Muir, T. (2014). Financial intermediaries and the cross-section of asset returns. *Journal of Finance*, 69 (6), 2557-2596.

Bai, J., Bali, T. G., and Wen, Q. (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, *131*(3), 619–642.

Balduzzi, P., and Robotti, C. (2008). Mimicking portfolios, economic risk premia, and tests of multi-beta models. *Journal of Business & Economic Statistics*, *26*(3), 354–368.

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

Barillas, F., and Shanken, J. (2018). Comparing asset pricing models. *Journal of Finance*, 73(2), 715–754.

Barillas, F., Kan, R., Robotti, C., and Shanken, J. A. (2019). Model comparison with Sharpe ratios. *Journal of Financial and Quantitative Analysis*, forthcoming.

Bessembinder, H., Kahle, K. M., Maxwell, W. F., and Xu, D. (2008). Measuring abnormal bond performance. *Review of Financial Studies*, 22(10), 4219–4258.

Bianchi, D., Guidolin, M., and Ravazzolo, F. (2017). Macroeconomic factors strike back: A Bayesian change-point model of time-varying risk exposures and premia in the US cross-section. *Journal of Business & Economic Statistics*, 35(1), 110–129.

Boguth, O., and Kuehn, L. A. (2013). Consumption volatility risk. *Journal of Finance*, *68*(6), 2589–2615.

Boudoukh, J., and Richardson, M. (1993). Stocks Returns and Inflation: A Long Term Perspective. *American Economic Review*, 83, 1346-1355.

Breeden, D. T., Gibbons, M. R., and Litzenberger, R. H. (1989). Empirical tests of the consumption-oriented CAPM. *Journal of Finance*, 44(2), 231–262.

Chen, R., and Kan, R. (2004). Finite Sample Analysis of Two-Pass Cross-Sectional Regressions, working paper University of Toronto.

Chen, N. F., Roll, R., and Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, 59, 383-403.

Cochrane, J. H. (2009). *Asset Pricing*, revised edition. Princeton, NJ: Princeton University Press.

Connor, G., and Korajczyk, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, 21(2), 255–289.

Connor, G. Korajczyk, R. A., and Uhlaner, R. (2015). A synthesis of two factor estimation method. *Journal of Financial and Quantitative Analysis*, 50(4), 825–842.

Cooper, I., and Priestley, R. (2011). Real investment and risk dynamics. *Journal of Financial Economics*, 101(1), 182–205.

Fama, E. F., and 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., and French, K. R. (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies*, 29(1), 69–103.

Fama, E. F., and French, K. R. (2018a). Choosing factors. *Journal of Financial Economics*, 128(2), 234–252.

Fama, E. F., and French, K. R. (2018b). Comparing Cross-Section and Time-Series Factor Models. Working paper.

Fama, E. F., and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3), 607–636.

Ferson, W., Siegel, A. F., and Xu, P. T. (2006). Mimicking portfolios with conditioning information. *Journal of Financial and Quantitative Analysis*, 41(3), 607–635.

Gebhardt, W. R., Hvidkjaer, S., and Swaminathan, B. (2005). The cross-section of expected corporate bond returns: Betas or characteristics? *Journal of Financial Economics*, 75(1), 85–114.

Giglio, Stefano., and Xiu, Dacheng (2019). Asset Pricing with Omitted Factors. Working paper, University of Chicago.

Gospodinov, N., Kan, R., and Robotti, C. (2019). Too good to be true? Fallacies in evaluating risk factor models. *Journal of Financial Economics*, 132(2), 451–471.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650-705.

Huberman, G., Kandel, S., and Stambaugh, R. F. (1987). Mimicking portfolios and exact arbitrage pricing. *Journal of Finance*, 42(1), 1–9.

James, W., and Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 361–379, University of California Press, Berkeley, Calif., *1961*.

Jegadeesh, N., and Noh, J. (2014). Empirical Tests of Asset Pricing Models with Individual Stocks. Working paper, Emory University.

Jegadeesh, N., Noh, J., Pukthuanthong, K., Roll, R., and Wang, J. (2019). Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation. *Journal of Financial Economics*, 133(2), 273–298.

Kleibergen, F. (2009). Tests of risk premia in linear factor models. *Journal of Econometrics*, *149*(2), 149-173.

Kleibergen, F., and Zhan, Z. (2018). Identification-robust inference on risk premia of mimicking portfolios of non-traded factors. *Journal of Financial Econometrics*, 16(2), 155–190.

Kleibergen, F. R., and Zhan, Z. (2019). Robust inference for consumption-based asset pricing. *Journal of Finance*, Forthcoming.

Kroencke, T. A., Schindler, F., Sebastian, S. P., and Theissen, E. (2013). GDP mimicking portfolios and the cross-section of stock returns. Working paper, University of Neuchâtel.

Lamont, O. A. (2001). Economic tracking portfolios. *Journal of Econometrics*, 105(1), 161–184.

Lehmann, B. N., and Modest, D. M. (1988). The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics*, 21(2), 213–254.

Lettau, M., and 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., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, 96(2), 175–194.

Maio, P. F. (2018). Comparing asset pricing models with traded and macro risk factors. Working paper, Hanken School of Economics.

Kozak, S., Nagel, S., & Santoshi, S. (2018). Interpreting Factor Models. *Journal of Finance,* 73(3), 1183-1223.

Pukthuanthong, K., Roll, R., and Subrahmanyam, A. (2018). A protocol for factor identification. *Review of Financial Studies*, 32(4), 1573–1607.

Roll, R., and Ross, S. A. (1994). On the cross-sectional relation between expected returns and betas. *Journal of Finance*, 49(1), 101–121.

Roll, R., and Srivastava, A. (2018). Mimicking portfolios. *Journal of Portfolio Management*, 44(5), 21–35.

Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial Studies*, 5(1), 1–33.

Stein, E. M. (1956). Interpolation of linear operators. *Transactions of the American Mathematical Society*, 83(2), 482–492.

Vassalou, M. (2003). News related to future GDP growth as a risk factor in equity returns. *Journal of Financial Economics*, 68(1), 47–73.

Table 1. Descriptive Statistics

This table reports the summary of statistics on the main variables, in which we list the number of observations, mean, median, standard deviation, and percentiles (1st, 5th, 25th, 75th, 95th, and 99th). Panel A reports the statistics for excess stock returns and its explanatory variables. For the stock return, we have 10,833 stocks in total and 626 months of data. The stock returns are over a risk-free rate (1-month T-bill rate). The explanatory variables include consumption growth rate (CG), consumer price index (CPI), industrial production (IP), unemployment rate (UE), excess stock market return (MKT), small-minus-big size portfolio (SMB), high-minus-low book/market portfolio (HML) and the consumption to wealth ratio (CAY). Panel B lists the statistic for corporate bond returns and its explanatory variables. For corporate bond returns, we have 6,421 bonds and 179 months of data. The bond returns are over a risk-free rate (one-month T-bill rate). In addition to the four macro variables (CG, CPI, IP, UE), bond market excess return (MKT_Bond), default spread (DS), and term spread (TS) are taken as explanatory variables. MKT_B is the equally weighted return of all corporate bond return in our sample in excess of the risk-free rate. DS is a default spread, measured by the return difference between Moody's long-term corporate BAA-rated bonds and AAA-rated bonds. TS is a term spread, measured by the return difference between 10-year treasury bonds and 1-year treasury bonds. The sources of these data are described in detail in Section 3.

Panel A: Statistics for stock returns and its explanatory variables

	N	Mean	Median	SD	1st	25th	75th	99th
Stock return	1,784,351	1.012	0.440	13.341	–31.764	–5.194	6.209	42.179
CG	626	0.018	0.006	0.528	–1.557	–0.303	0.305	1.344
CPI	626	0.007	–0.006	0.248	–0.695	–0.124	0.145	0.594
IP	626	0.004	0.018	0.699	–1.985	–0.374	0.373	1.932
UE	626	0.003	0.001	0.161	–0.408	–0.097	0.107	0.403
MKT	626	0.490	0.785	4.466	–11.804	–2.100	3.450	11.178
SMB	626	0.229	0.130	3.108	–6.695	–1.520	2.050	8.435
HML	626	0.349	0.310	2.819	–8.097	–1.160	1.710	7.930
CAY	626	–0.002	–0.002	0.021	–0.046	–0.015	0.015	0.034

Panel B: Statistics for bond returns and its explanatory variables

	N	Mean	Median	SD	1st	25th	75th	99th
Bond return	331,728	0.389	0.236	2.544	–5.764	–0.349	1.089	7.248
CG	179	0.021	0.021	0.362	–1.248	–0.164	0.238	0.880
CPI	179	0.004	0.020	0.278	–0.875	–0.127	0.132	0.678
IP	179	–0.027	–0.003	0.648	–1.909	–0.359	0.330	1.355
UE	179	0.002	–0.003	0.154	–0.378	–0.098	0.105	0.376
MKT_Bond	179	0.380	0.390	1.875	–6.031	–0.367	1.094	7.276
DS	179	1.093	0.960	0.469	0.579	0.853	1.218	3.090
TS	179	1.805	1.870	1.003	–0.374	1.235	2.590	3.368

Table 2. Simulation

This table shows the simulation results of the effectiveness of FMPs constructed by various approaches as a proxy of true risk factors. We generate simulated factors using a return-related factor added by a normally distributed measurement error. We generate simulated returns using orthogonalized true risk factors multiplied by orthogonalized true beta loadings. With the simulated return and simulated risk factors, we construct FMPs using the six methods described in Section 5. The details of the simulations are discussed in Section 5. Panel A shows the correlations between FMPs and their underlying factors. Panel B presents the maximum correlations between FMPs for a macro factor with other factors. Panel C presents the averaged correlations between FMPs for a macro factor with other factors. The values in each table are mean values across 1,000 simulations. The sample period is January 1964 to March 2016. To be included, individual stocks must have at least 60 continuous monthly returns on CRSP. The macro factors include unexpected consumption growth (CG), unexpected changes in the CPI (CPI), unexpected changes in industrial production (IP), and unexpected changes in the unemployment rate (UE).

Panel A: Correlation between FMPs and their underlying factors

	CG	CPI	IP	UE
FMP_IV	0.999	0.996	0.986	0.988
FMP_OLS	0.798	0.792	0.801	0.811
FMP_Stein	0.798	0.792	0.801	0.822
FMP_LM	0.841	0.894	0.827	0.831
FMP_SB	0.704	0.795	0.614	0.782
FMP_TS	0.264	0.186	0.173	0.197

Panel B: Maximal correlation between FMPs of a target factor and other true risk factors

	CG	CPI	IP	UE
FMP_IV	0.007	0.008	0.013	0.013
FMP_OLS	0.119	0.097	0.099	0.077
FMP_Stein	0.119	0.097	0.099	0.077
FMP_LM	0.118	0.094	0.098	0.075
FMP_SB	0.441	0.555	0.360	0.131
FMP_TS	0.087	0.075	0.079	0.092

Panel C: Average correlation between FMPs for a target factor and other true risk factors

	CG	CPI	IP	UE
FMP_IV	0.003	0.004	0.006	0.006
FMP_OLS	0.051	0.045	0.050	0.038
FMP_Stein	0.051	0.045	0.050	0.038
FMP_LM	0.052	0.045	0.048	0.037
FMP_SB	0.106	0.126	0.158	0.059
FMP_TS	0.040	0.035	0.040	0.043

Table 3. Correlations with the Underlying Factors

This table presents correlations between FMPs and their the underlying factors. Specifically, this table reports correlations between FMPs and their underlying factors across five subsamples, one of which spans a decade. The first row is the mean correlation, and the second row (# Sig) contains the number of subsample correlations that is significant at the 1% level (maximum of 5.) We compare FMPs constructed by different methods: the FMP constructed by the IV approach in cross-sectional regression (FMP_IV), the FMP constructed by the OLS approach in cross-sectional regression (FMP_OLS), the FMP constructed by the Lehmann and Modest method (FMP_LM), the FMP constructed by the Stein method (FMP_Stein), the FMP constructed by time-series regression with Lamont portfolios as basis assets (FMP_Time-series), and the FMPs constructed with the sorting by beta approach (FMP_SB).

		CG	CPI	IP	UE	MKT	SMB	HML
					Cross-sectional approach			
FMP_IV		0.4105	0.4853	0.6627	0.5652	0.7413	0.8038	0.7097
	# Sig	5	5	5	5	5	5	5
FMP_OLS		0.6356	0.7401	0.7889	0.7233	0.9033	0.9176	0.8749
	# Sig	5	5	5	5	5	5	5
FMP_LM		0.4975	0.6897	0.7166	0.7254	0.9095	0.9565	0.886
	# Sig	5	5	5	5	5	5	5
FMP_Stein		0.6356	0.7401	0.7889	0.7233	0.9033	0.9176	0.8749
	# Sig	5	5	5	5	5	5	5
					Time-series approach			
FMP_Time-series		0.2129	0.3416	0.2932	0.2218	1	0.4195	0.3012
	# Sig	1	4	3	2	5	5	3
					Sorting-by-beta approach			
FMP_SB		0.7283	0.7671	0.725	0.6907	0.877	0.9142	0.8557
	# Sig	5	5	5	5	5	5	5

Table 4. Relation with Covariance of Asset Returns

This table reports the canonical correlations between FMPs and the principal components (PCs) of the covariance matrix of individual stocks. The factor candidates include FMPs constructed using the methods of IV, OLS, Lehmann and Modest (1999), Stein (1956), and Lamont (2001). As a comparison, we also include the canonical correlation results for the original risk factors. The PCs are extracted as explained in Pukthuanthong et al. (2019) using the Connor and Korajczyk (henceforth CK, 1988) cross-sectional method. We summarize significance levels for factor candidates. The following procedure is implemented to derive the significance levels of each factor candidate: First, for each canonical pair, the eigenvector weights for the 10 PCs are taken and the weighted average PC (which is the canonical variate for the 10 PCs that produced the canonical correlation for this particular pair) is constructed. Then a regression using each CK PC canonical variate as the dependent variable and the candidate factor realizations as the seven independent variables is run over the sample months. The *t*-statistics from the regression then give the significance level of each candidate factor. There are 10 pairs of canonical variates in each decade and a canonical correlation for each one; thus, there is a total of 50 such regressions (10 regression per decade). The first row (Avg *t*) presents the mean *t*-statistic over all canonical correlations. The second row (Avg *t* Sig. CC) reports the mean *t*-statistic when the canonical correlation itself is statistically significant. The third row (# decades) reports the average number of significant canonical correlations over the five decades. Critical rejection levels for the *t*-statistic are 1.65 (10%), 1.96 (5%), and 2.59 (1%). We assume that an FMP satisfies this criterion if (1) it is significantly related to any canonical variate in all decades or has a mean *t*-statistic in the second row that exceeds the one-tailed, 2.5% cutoff based on the chi-squared value, and (2) has an average number of significant *t*-statistics exceeding 1.75 (the third row of each panel). *t*-statistics breaching the 5% (1%) critical level are in boldface. The factors that pass necessary conditions are highlighted in gray.

	FMP				Equity factors		
	CG	CPI	IP	UE	Rm.Rf	SMB	HML
				Original			
Avg *t*	1.14	1.15	1.10	1.01	10.66	6.70	3.31
Avg *t* (Sig. CC)	1.25	1.37	1.04	1.14	**22.80**	**14.19**	**6.87**
# decades	1.40	1.40	1.40	0.60	2.80	2.80	2.60
				FMP_IV			
Avg *t*	2.43	2.57	1.89	2.22	9.71	5.00	3.30
Avg *t* (Sig. CC)	**3.00**	**3.27**	**2.10**	**2.57**	**13.41**	**6.49**	**4.25**
# decades	3.00	3.20	2.20	2.80	3.40	3.80	3.20
				FMP_OLS			
Avg *t*	2.50	2.53	2.05	2.66	10.07	5.45	3.42
Avg *t* (Sig. CC)	**3.01**	**3.12**	**2.41**	**3.39**	**13.80**	**7.18**	**4.46**
# decades	2.80	3.00	2.80	3.20	3.20	4.00	4.20
				FMP_LM			
Avg *t*	2.28	2.52	2.07	1.93	3.35	3.77	3.03
Avg *t* (Sig. CC)	**2.72**	**3.13**	**2.28**	**2.30**	**4.21**	**4.79**	**3.97**
# decades	3.20	3.20	3.20	1.80	3.00	4.40	3.20
				FMP_Stein			
Avg *t*	2.50	2.53	2.05	2.66	10.07	5.45	3.42
Avg *t* (Sig. CC)	**3.01**	**3.12**	**2.41**	**3.39**	**13.80**	**7.18**	**4.46**
# decades	3.40	3.40	3.60	3.00	3.00	4.00	3.00
				FMP_Time-series			
Avg *t*	1.39	1.57	1.37	1.40	9.79	6.74	3.25
Avg *t* (Sig. CC)	**2.01**	**2.09**	1.63	1.70	**19.45**	**13.17**	**6.09**
# decades	2.00	1.80	1.80	1.80	3.40	3.20	3.60
				FMP_SB			
Avg *t*	2.01	2.80	2.16	2.15	6.54	5.42	5.23
Avg *t* (Sig. CC)	**2.21**	**3.20**	**2.38**	**2.36**	**8.07**	**6.68**	**6.46**
# decades	2.40	4.00	3.20	3.20	3.40	3.60	3.40

Table 5. Risk Premiums in Equity Market By Using Factor-Mimicking Portfolios

This table reports risk-premium estimates using the Fama-MacBeth regression. For cross-sectional methods, we assume betas are constant. For the IV method, we use betas in even months as instruments for betas in odd months and apply the IV method with sample adjustment to obtain FMPs (assets that have $\hat{\beta}_{IV}{}' \hat{\beta}_{EV} > 0$). Then we apply the IV method for a second time to test the risk premiums of these FMPs (IV). For other methods, such as OLS, Lehmann, and Modest (LM), and Stein, we apply the corresponding approach (OLS, LM, Stein) to obtain factor-mimicking portfolios and use these methods to run Fama-MacBeth regressions again to test risk premiums. For the time-series approach, we use time series to create FMPs and then estimate the risk premiums for these FMPs using the OLS method in the second-pass regression. For the sorting-by-beta method, we estimate betas from time-series multivariate regressions and then sort the betas for each factor into ten deciles. We then construct FMPs as the arithmetic average returns in the highest decile minus the average in the lowest percentile (HML). The FMPs are used as factors to estimate the risk premium, which is reported in the table. The sample period is January 1964 to March 2016. We use individual stocks that have at least 60 continuous month returns in CRSP. The risk factors include four macroeconomics variables, the consumption growth rate (CG), unexpected CPI changes (CPI), unexpected changes in industrial production (IP), and unexpected changes in the unemployment rate (UE). MKT is the excess market return (proxied by the value-weighted return of all CRSP firms in the United States), SMB is the FF small-minus-big size factor, and HML is the FF high-minus-low book-to-market factor. The values in parentheses are *t*-statistics. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Intercept	CG	CPI	IP	UE	MKT	SMB	HML
				Cross-sectional approach				
IV	0.464***	0.164***	–0.017	0.009	–0.022**			
	(3.980)	(3.238)	(–0.920)	(0.192)	(–2.000)			
OLS	0.549***	0.066**	–0.004	–0.067**	0.006			
	(4.180)	(2.202)	(–0.366)	(–2.117)	(0.791)			
LM	0.306***	0.095**	–0.017	–0.079**	0.005			
	(3.140)	(2.506)	(–1.291)	(–2.312)	(0.544)			
Stein	0.527***	0.083**	–0.008	–0.094**	0.026			
	(4.009)	(2.202)	(–0.366)	(–2.117)	(0.791)			
IV	0.515***					0.553**	0.281*	–0.474***
	(4.805)					(2.281)	(1.816)	(–3.411)
OLS	0.495***					0.485**	0.194	–0.296**
	(5.177)					(2.516)	(1.423)	(–2.278)
LM	0.277***					0.590***	0.216*	–0.208
	(3.163)					(3.124)	(1.655)	(–1.583)
Stein	0.494***					0.492**	0.206	–0.394**
	(5.153)					(2.516)	(1.423)	(–2.278)
IV	0.653***	0.136**	–0.048**	0.094	–0.028**	0.545**	0.262*	–0.487***
	(5.985)	(2.070)	(–2.132)	(1.606)	(–2.134)	(2.269)	(1.835)	(–3.766)
OLS	0.457***	–0.015	–0.007	–0.046	0.003	0.490***	0.228*	–0.282**
	(5.132)	(–0.618)	(–0.619)	(–1.357)	(0.303)	(2.599)	(1.703)	(–2.241)
LM	0.259***	–0.012	–0.019	–0.044	0.004	0.584***	0.233*	–0.223*
	(3.237)	(–0.496)	(–1.487)	(–1.339)	(0.478)	(3.133)	(1.797)	(–1.745)
Stein	0.452***	–0.139	–0.025	–0.252	0.009	0.497***	0.241*	–0.367**
	(5.054)	(–0.618)	(–0.619)	(–1.357)	(0.303)	(2.599)	(1.703)	(–2.241)

	Intercept	CG	CPI	IP	UE	MKT	SMB	HML
				Time-series approach				
Model 1	0.544***	–0.003	–0.003	–0.039***	0.005***			
	(4.864)	(–0.496)	(–0.763)	(–3.974)	(2.664)			
Model 2	0.326***					0.582***	0.126**	–0.129***
	(3.065)					(2.833)	(2.077)	(–2.910)
Model 3	0.352***	–0.004	–0.001	–0.030***	0.005***	0.527***	0.107*	–0.138***
	(3.510)	(–0.605)	(–0.295)	(–3.211)	(2.725)	(2.617)	(1.798)	(–3.274)
				Sorting-by beta (SB) approach				
Model 1	0.688***	0.738*	0.096	–0.595**	0.184			
	(4.039)	(1.772)	(0.296)	(–2.032)	(0.618)			
Model 2	0.516***					0.922**	0.569	–0.704*
	(5.122)					(2.266)	(1.415)	(–1.677)
Model 3	0.491***	0.054	–0.103	–0.235	–0.039	0.928**	0.594	–0.742*
	(4.929)	(0.209)	(–0.328)	(–0.950)	(–0.128)	(2.335)	(1.511)	(–1.837)

Table 6. Estimated Risk Premiums with FMPs for Consumption Growth and CAY

This table shows the estimated risk premiums for consumption growth and the CAY factor of Lettau and Ludvigson (2001) and with and without the three FF factors. CAY is the log ratio of consumption to aggregate wealth. FMPs are constructed for each factor using four methods (IV, OLS, time-series, and sorting-by-beta (OLS-SB)). CG is the unexpected consumption growth rate. MKT is the excess market return, SMB is the FF small-minus-big size factor, and HML is the FF high-minus-low book-to-market factor. We obtain the monthly CAY from Martin Lettau's websites. The monthly sample is from January 1964 to March 2016. The t-values in parentheses are based on Newey-West standard errors. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Intercept	CAY	CG	CG*CAY	MKT	SMB	HML
IV	0.751***	–0.171*					
	(4.002)	(–1.757)					
OLS	0.781***	–0.123					
	(3.974)	(–1.616)					
Time-series	0.928***	–0.035					
	(4.610)	(–0.948)					
SB	0.858***	–0.313					
	(4.424)	(–1.586)					
IV	0.566***		0.125**	–0.003			
	(3.761)		(2.017)	(–0.022)			
OLS	0.600***		0.074**	–0.008			
	(3.772)		(2.077)	(–0.110)			
Time-series	0.818***		0.007	–0.010			
	(5.601)		(0.909)	(–1.235)			
SB	0.722***		0.613*	0.129			
	(4.145)		(1.938)	(0.559)			
IV	0.498***	–0.08	0.150**	–0.03			
	(3.384)	(–0.920)	(2.391)	(–0.229)			
OLS	0.592***	–0.094	0.061*	–0.014			
	(4.090)	(–1.351)	(1.808)	(–0.190)			
Time-series	0.811***	–0.023	0.006	–0.011			
	(5.652)	(–0.774)	(0.787)	(–1.496)			
SB	0.739***	–0.155	0.533*	0.127			
	(4.502)	(–0.670)	(1.833)	(0.567)			
IV	0.584***	0.135	0.238***	–0.197	0.635**	0.361**	–0.504***
	(5.452)	(1.145)	(2.692)	(–1.165)	(2.494)	(2.166)	(–3.169)
OLS	0.463***	–0.032	–0.016	–0.006	0.520**	0.187	–0.296**
	(5.017)	(–0.381)	(–0.614)	(–0.084)	(2.574)	(1.419)	(–2.061)
Times-series	0.322***	0.044	–0.002	–0.011	0.588***	0.105*	–0.134***
	(3.117)	(1.421)	(–0.390)	(–1.448)	(2.881)	(1.767)	(–3.102)
SB	0.602***	–0.047	0.129	0.005	0.673**	0.531	–0.632*
	(5.869)	(–0.183)	(0.691)	(0.020)	(1.978)	(1.567)	(–1.669)

Table 7. Testing FMPs to Explain Individual Corporate Bond Returns

This table shows the four criterion tests for FMPs constructed by various approaches and with individual corporate bond returns as basis assets. Panel A reports correlations between FMPs and the underlying factors for explaining corporate bonds. Panel B presents canonical correlations of FMPs with asymptotic PCs using the same approach as that reported in Table 4 for equities. Panel C shows risk premiums for corporate bond returns estimated with FMPs as factors. The factors are described in Table 1. The sample period is August 2002 to June 2017. The numbers in parentheses are *t*-statistics. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Correlations with underlying factors

	CG	CPI	IP	UE	MKT_B	DS	TS
				Cross-sectional approach			
IV	0.466	0.438	0.560	0.278	0.867	0.619	0.211
OLS	0.633	0.674	0.703	0.630	0.949	0.849	0.358
Stein	0.716	0.724	0.762	0.742	0.978	0.875	0.451
LM	0.633	0.674	0.703	0.630	0.949	0.849	0.358
				Time-series approach			
Time-series	0.423	0.417	0.461	0.473	0.693	0.973	1.000
				Sorting-by-beta approach			
Sorting by beta	0.426	0.584	0.534	0.259	0.934	0.468	0.137

Panel B: Canonical correlations with asymptotic PCs and significance levels of factor candidates

	CG	CPI	IP	UE	MKT	DS	TS
				Original			
Avg *t*	1.18	1.04	1.96	1.03	4.98	1.55	1.09
Avg *t* (Sig. CC)	1.51	0.98	**4.90**	0.72	**16.09**	**2.34**	0.98
				Cross-sectional approach			
				FMP_IV			
Avg *t*	5.35	5.02	4.75	2.12	4.54	4.51	3.52
Avg *t* (Sig. CC)	**5.35**	**5.02**	**4.75**	**2.12**	**4.54**	**4.51**	**3.52**
				FMP_OLS			
Avg *t*	4.48	4.43	6.90	3.54	10.12	3.10	3.14
Avg *t* (Sig. CC)	**5.07**	**5.01**	**8.03**	**3.98**	**11.67**	**3.23**	**3.63**
				FMP_Stein			
Avg *t*	4.48	4.43	6.90	3.54	10.12	3.10	3.14
Avg *t* (Sig. CC)	**5.07**	**5.01**	**8.03**	**3.98**	**11.67**	**3.23**	**3.63**
				FMP_LM			
Avg *t*	3.41	3.55	5.44	2.02	7.00	3.27	2.26
Avg *t* (Sig. CC)	**4.42**	**4.63**	**7.38**	**2.30**	**9.54**	**3.96**	**2.54**
				Time-series approach			
Avg. *t*	1.99	1.88	1.33	0.98	2.60	2.26	1.55
Avg *t* (Sig. CC)	**2.39**	**2.53**	1.75	0.69	**3.53**	**2.77**	1.79
# of sig	2	2	2	1	4	5	3
				Sorting-by-beta approach			
Avg. *t*	5.25	4.64	4.56	4.88	6.99	5.10	3.84

Avg *t* (Sig. CC)	**6.00**	**5.04**	**5.07**	**5.68**	**8.05**	**5.94**	**4.33**
# of sig	4	6	4	6	4	6	3

Panel C: Estimated risk premiums using FMPs

	Intercept	CG	CPI	IP	UE	MKT_B	DS	TS
				Cross-sectional approach				
IV	0.892***	0.291***	–0.049	0.586***	–0.090	0.376*	0.176**	–0.355
	(3.838)	(3.082)	(–0.589)	(2.838)	(–1.174)	(1.870)	(2.275)	(–1.389)
OLS	0.095**	–0.003	–0.028	–0.023	0.009	0.305*	0.115***	0.036
	(2.310)	(–0.088)	(–1.092)	(–0.370)	(0.558)	(1.723)	(3.317)	(0.770)
LM	0.01	–0.004	–0.017	–0.009	0.013	–0.477	0.106***	0.051
	(0.445)	(–0.113)	(–0.665)	(–0.150)	(0.908)	(–1.140)	(2.695)	(1.136)
Stein	0.121**	–0.003	0.012	–0.03	0.022	0.215	0.035	0.152
	(2.161)	(–0.067)	(0.405)	(–0.315)	(1.171)	(1.431)	(1.125)	(1.441)
				Time-series approach				
Time-series	0.129**	–0.020*	–0.019*	–0.049**	0.019***	0.299**	0.119***	–0.007
	(2.303)	(–1.790)	(–1.759)	(–2.059)	(3.345)	(2.411)	(3.692)	(–0.213)
				Sorting by beta approach				
Sorting by beta	0.115***	0.048	–0.137	0.016	–0.04	0.401	0.277*	0.014
	(2.766)	(0.244)	(–0.801)	(0.082)	(–0.220)	(1.329)	(1.775)	(0.087)

Appendix A: Simulation of the risk-premium estimation by using FMPs

A.1. Simulation procedure

This appendix presents the simulations that compare various FMP construction methods. The purpose is to study the magnitude of biases in risk-premium estimations due to the measurement error of beta loadings, as well as other statistical properties in finite samples. The finite-sample properties of risk-premium estimation for traded factors are studied by Jegadeesh et al. (2019). However, the non-traded factors are possibly associated with higher estimation errors in factor loadings, which elicit larger finite sample errors-in-variables bias (Kleibergen 2009). Therefore, it is necessary to reexamine these properties for non-traded factors.

Our simulation parameters match attributes of the real data. We use individual stock-returns data from February 1964 to March 2016, covering 626 months and 10,833 stocks. For instance, most stocks do not have data spanning the entire sample period; hence, our simulated stocks have data only in the same periods as their corresponding stocks in the real data.[1]

The simulation procedure is as follows.

Step 1: Regress excess stock returns on factors and obtain estimated betas $(\boldsymbol{\beta})$ and residuals $(\boldsymbol{\varepsilon})$ for each stock. $\boldsymbol{\beta}$ is an N by K matrix, and $\boldsymbol{\varepsilon}$ is an N by T matrix, where N is the number of stocks, and T is the number of time periods. Since the existing periods for most stocks are less than T, the matrix $\boldsymbol{\varepsilon}$ is not a balanced panel, in which we define the value as missing if one stock does not have return data on the corresponding period.

Step 2: In each simulation, create a $T \times 1$ vector $\mathbf{S}$ by randomly selecting T numbers with replacement from 1 to T, where T is the maximum month number. Then create simulated factors by rearranging observed factors to match the randomly chosen observation number in the vector $\mathbf{S}$. Finally, augment the simulated factors by adding prespecified true premiums λ_0 set equal to the observed average risk premiums from Chen, Roll, and Ross (1986) and Chen and Kan (2003).

[1] We also conduct simulations in which all simulated stocks have data in every period (unreported). This leads to less overall bias, because the number of months in first-pass time-series regressions and the number of stocks in second-pass cross-sectional regressions are much larger. However, the overall conclusions comparing FMP methods remain the same.

Step 3: Generate simulated residuals that are randomly selected and normally distributed, for which the mean and variance are equal to the sample mean and variance of the observed actual residuals for the stock.

Step 4: Construct simulated returns for each stock in each month as estimated betas multiplied by simulated factors plus simulated residuals.

Step 5: Apply the methods in Section 2 of the paper to construct FMPs using simulated returns and simulated factors.

Step 6: After construction of the FMP, reapply various Fama-Macbeth two-pass methods (OLS, IV, and Stein with individual stocks) to estimate risk premiums using simulated returns and the FMPs, thereby obtaining simulated estimates of FMP-based risk premiums.

Notably, all simulations are multivariate, except IV. For the IV, we use a univariate regression to create FMP due to the condition of $\hat{\beta}_{IV}{}'\hat{\beta}_{EV} > 0$. To estimate risk premiums, like Panel C in Table 1 in the main paper, we use multivariate regression. That is, we create FMPs for each factor independently. The second step is to test risk premiums for the FMP, where we use multivariate regression to estimate risk premiums of all FMPs.

Repeat these six steps 1,000 times. Then calculate the mean difference between simulated estimates of FMP-based risk premiums and the true simulated risk premiums (which is the ex-ante bias). This mean difference is the predicted bias of each FMP method.[2]

A.2. Simulation results: Bias

Table A1 presents the simulation results. Following Chen, Roll, and Ross (1986), we use four macro factors as risk factors: changes in the consumption growth rate (CG), changes in the Consumer Price Index (CPI), changes in industrial production (IP), and changes in the unemployment rate (UE). We assume that beta is constant for each stock. In the first stage, we

---

[2] In unreported results, we also compute an ex-post bias as the differences between estimated risk premiums and the sample mean of the corresponding factor realizations in that particular simulation. We find the difference between ex-post bias and ex-ante bias is minimal; thus we provide only ex-ante bias in Table 2 in the paper.

apply the univariate method to construct each FMP; then in the second stage, we apply a multivariate approach to estimate the risk premium.

Consistent with our expectations, the IV approach resolves the EIV bias in a two-pass regression and produces nearly unbiased risk premiums. The alpha in the IV estimation is 0.0055%, which is close to 0. The differences between the estimated risk premiums and true risk premiums are minimal. For instance, the estimated risk premium for consumption growth is just 0.2% larger than its true risk premium.

[Table A1 around here]

The OLS method produces the estimated risk premiums that are much smaller than the true risk premiums, which is consistent with our conjecture that the OLS is subject to downward bias caused by measurement errors in estimated betas. The bias ranges from 28.7% for consumption growth to 50% for the unemployment rate. The mispricing part (intercept term) is 0.1412%, which is much larger than for IV. The alpha from the sorting-by-beta method (OLS_SB) is even larger than OLS, and the risk premium bias is large as well. For example, the bias for consumption growth is 31%. The biases for the Lehmann and Modest (LM) method is as large as that for OLS. Furthermore, the mispricing in LM method is 32%, even larger than the mispricing of OLS. Stein's method yields less bias than then the OLS and LM methods, but the biases in Stein's method are still large, over 20%.

We also consider the time-series approach. Specifically, we present the time series approach using Lamont's (2001) portfolios as basis assets, which is constructed by regressing risk factors on a series of basis portfolio returns, and then calculating predicted values as FMPs. Then we estimate the risk premiums for the FMPs constructed by time-series approach (TS_FMPs). Our simulation in Panel B of Table A1 shows that the average bias for time-series approach is larger than 40%.

A.3 Simulation results: Root-mean-square errors

We also calculate the root-mean-square errors (RMSEs) for the simulations. The error is simply the difference between the true risk premium and the risk premium estimated for each simulated replication, of which there are 1,000. We call this the ex ante RMSE. We also calculate the difference between the estimated risk premium in each replication and the sample mean of the corresponding risk-factor realizations and then compute its root mean square across 1,000 replications. This is the ex post RMSE. These RMSEs are reported in the Panel A of Table A2. Since the ex-ante and ex-post RMSEs are quite similar, we focus on the former.

The ex-ante RMSEs with IV are uniformly smaller than with OLS; i.e., IV is considerably more accurate. For example, the CPI ex-ante RMSE is 0.082 for OLS, but just 0.026 for IV. Stein has smaller RMSE than OLS, but it is still larger than that of IV. OLS has the largest RMSE. LM has marginally smaller RMSE than OLS. Overall, the IV method dominates all other methods by the RMSE criterion.

[Table A2 about here]

A.4. The size and power of *t*-tests

To estimate the size and power of *t*-tests for risk premiums with the IV approach, we first consider the probability of rejecting the null hypotheses falsely (i.e., the type-I error). In the simulations, we set the true risk premium equal to 0 for all factors. Then we follow the simulation procedure to estimate risk premiums and their corresponding *t*-statistic (i.e., the mean estimated risk premium divided by its corresponding standard error). We use a 5% significance level (the critical value is 1.96) and calculate the frequency of the absolute value of the *t*-statistic that is larger than 1.96 in 1,000 replications. Panel B of Table A2 reports that the size of each macro factor is around 5% or slightly below.

To examine the *t*-test power (the probability of rejecting the null hypothesis when the alternative hypothesis is true), we set the true premium to the mean-risk premium estimated from a Fama-MacBeth two-pass regression with OLS. This is a relatively small risk premium (e.g., compared to that obtained with IV), thus implying a more conservative threshold for power. Panel B in Table A2 presents the results. The frequency of rejecting an incorrect null hypothesis that the risk premium is 70.9% for consumption growth and a bit higher for the other three macro factors. Overall, size and power tests indicate that a normal *t*-statistic delivers effective inferences about macro factor premiums.

Table A1. Simulation on Biases of Mispricing and Risk Premiums of FMPs

This table reports biases in the estimated risk premiums from Monte Carlo simulations for FMPs. We simulate the stock returns and factors as described in Appendix A. We create FMPs following the methods described Section 3 of the paper. With these FMPs, we run cross-sectional regressions again to estimate risk premiums. The first rows in the table show the true risk premiums for four macro variables when alpha (mispricing) is set to 0. The table reports estimated risk premiums along with their mean biases across 1,000 replications, expressed as a percentage of the true value. Panel A presents risk premiums and biases on the IV method that only retains stocks whose betas in even and odd subsamples have the same sign in the IV method, OLS, sorting-by-beta (OLS-SB), Lehmann and Modest (1988), (LM), and Stein (1956). In Panel B, we present the time-series FMP methods by using Lamont (2001)'s basis assets. Panel C presents the risk premiums of FMPs computed from the estimated time-series coefficients in Panel A as a proxy for the FMPs. We list IV, OLS, and sorting-by-beta as examples of cross-sectional approaches. The sample period is January 1964 to March 2016. To be included, individual stocks must have at least 60 continuous monthly returns on CRSP. The macro factors include unexpected consumption growth (CG), unexpected changes in the CPI (CPI), unexpected changes in industrial production (IP), and unexpected changes in the unemployment rate (UE).

	Alpha	CG	CPI	IP	UE
True risk premium	0	0.2	–0.2	1.2	0.3
Panel A: Cross-sectional methods					
IV	–0.0003	0.2072	–0.2011	1.1732	0.3035
		3.60%	0.55%	–2.23%	1.17%
OLS	0.1412	0.1426	–0.1147	0.6124	0.1499
		–28.7%	–42.7%	–49.0%	–50.0%
OLS-SB	–0.0569	0.1379	–0.1774	1.1362	0.2685
		–31.05%	–11.30%	–5.32%	–10.50%
LM	0.2000	0.1367	–0.1150	0.6738	0.1555
		–31.7%	–42.5%	–43.9%	–48.2%
Stein	0.0533	0.153	–0.1506	0.7485	0.2441
		–23.5%	–24.7%	–37.6%	–18.6%
Panel B: Time-series method					
Time-series	–5.0599	0.0861	–0.0803	0.7098	0.1624
		–56.95%	–59.85%	–40.85%	–45.87%

Table A2. Simulation on RMSE, Size, and Power *t*-Test

Panel A reports the root-mean-square error (RMSE.) The ex-ante RMSE measures the mean difference between the estimated risk premium and the true risk premium. The ex-post RMSE measures the difference between the estimated risk premium and the risk factor's realization. The RMSEs are computed across 1,000 replications in each simulation. Panel B shows the size and power of *t*-statistics for the IV method. Size is based on the 1.96 critical value (a 5% significance level.). It measures the probability of improperly rejecting a true null hypothesis (simulated here as truly zero-risk premiums.) Power is the probability of rejecting a false null hypotheses; in this case, the alternative (true) hypothesis consists of risk premiums obtained from the OLS method, which are generally smaller than those of the other methods.

Panel A: RMSE

		CG	CPI	IP	UE
OLS	Ex-ante	0.0700	0.0820	0.5906	0.1509
	Ex-post	0.0690	0.0821	0.5901	0.1509
IV	Ex-ante	0.0549	0.0260	0.1318	0.0224
	Ex-post	0.0520	0.0239	0.1287	0.0217
LM	Ex-ante	0.0850	0.0793	0.5312	0.1480
	Ex-post	0.0812	0.0791	0.5400	0.1474
Stein	Ex-ante	0.0622	0.0473	0.4577	0.0585
	Ex-post	0.0592	0.0456	0.4553	0.0581

Panel B: Size and power of *t*-test in the IV approach

	CG	CPI	IP	UE
Size	4.9%	5.2%	4.3%	3.8%
Power	70.9%	86.0%	83.1%	71.2%

Appendix B: FMPs constructed by time-series approach with a different set of predictors

In this section, we evaluate the FMPs constructed using the time-series approach with different sets of predictors. In addition to constructing the time-series FMP (TS_FMP) and following Lamont (2001) in our main text, we examine three more methods. Giglio and Xiu (2019) create FMPs with the time-series approach using principal components (PCs) of a large set of portfolios as predictors. First, we apply their method and use the first 10 PCs (The 10 PCs are constructed by 202 portfolios[3]) as predictors, and we construct TS_FMPs by using the fitted value from the regression. The FMPs are denoted by FMP_GX10PC. Second, we use 202 portfolios from Giglio and Xiu (2019) (these portfolios capture most of the cross-section anomalies) as the predictor to construct FMP (denoted by FMP_GX202). The time-series approach has a limitation because a large set of predictors in OLS regressions suffers the curse of the dimensionality and overfitting problem. We mitigate this problem by applying the time-series approach with a variable selection method to a large set of portfolios. Specifically, we adopt the Lasso method to solve the overfitting problem and select predictors from the 202 portfolios, which is the third method, denoted by "FMP_GX202_LASSO".

Table B1 summarizes the empirical results of the FMPs constructed by three methods above. Panel A lists the FMP correlation with original factors. FMPs created by FMP_GX202 have larger correlations with original factors than the other methods; however, FMP_GX202 suffers an overfitting problem, which boosts the correlation. Except for FMP_GX202, the other methods yield correlations lower than that by Lamont (2001) in Table 3 in the main text.

[Table B1 about here]

Panel B shows the correlations between FMPs and the systematic risk of stock returns (relation with covariance matrix). All the factors from FMP_GX10PC and FMP_GX202_LASSO have significant correlation with systematic risk, indicating that the methods can effectively extract risk components from noisy non-traded factors. However, CG and IP factors from FMP_GX202 are not significantly correlated with systematic risks.

[3] These portfolios include 25 portfolios sorted by size and book-to-market ratio, 17 industry portfolios, 25 portfolios sorted by operating profitability and investment, 25 portfolios sorted by size and variance, 35 portfolios sorted by size and net issuance, 25 portfolios sorted by size and accruals, 25 portfolios sorted by size and beta, and 25 portfolios sorted by size and momentum.

Panel C shows the risk premiums of these FMPs. Of the three methods, the FMP_GX202_LASSO dominates the other two in terms of the magnitude and significance of the risk premium, in which consumption growth and industrial production have significant risk premiums and are robust when controlling the FMPs of the three Fama-French factors. However, the significance of the consumption growth of FMP_GX202_LASSO are still smaller than those of FMP_IV in Table 3 (in the main text). The FMP_GX202 method leads to the smallest risk premium. It is noteworthy that even though IP is significant, it has the opposite sign of what is predicted theoretically. Overall, these results show that FMPs constructed by the time-series approach using a large set of portfolios and their principal components do not pass all criteria.

Table B1. FMPs Constructed by the Time-Series Approach with Different Sets of Predictors

This table presents the FMP examination when using different sets of predictors in the time-series approach to constructing FMPs. FMP_GX10PC includes the FMPs that are constructed by the time-series approach using 10 principal components (from 202 Giglio and Xiu (2019) portfolios) as predictors. FMP_GX202 consists of the FMPs constructed by the time-series approach with 202 portfolios as predictors. FMP_GX202_LASSO consists of the FMPs constructed by the time-series approach using the LASSO method to select predictors from 202 portfolios. Panel A displays the average correlations in five subsamples between FMPs and their underlying factors and the number of significant correlations (# Sig) at the 1% significance level in five subsamples, similar to Panel A in Table 3 (first criteria). Panel B shows the correlations between FMPs and the systematic risk factors extracted from individual stock returns (the second criteria). Panel C presents the estimated risk premiums (the third criteria). *t*-statistics breaching the 5% (1%) critical level are in boldface

Panel A: FMP correlation with original factors

	CG	CPI	IP	UE	MKT	SMB	HML
FMP_GX10PC	0.213	0.256	0.093	0.092	0.997	0.986	0.926
#Sig	2	4	1	1	5	5	5
FMP_GX202	0.636	0.600	0.588	0.552	0.999	0.997	0.993
#Sig	5	5	5	5	5	5	5
FMP_GX202_LASSO	0.241	0.291	0.266	0.037	0.999	0.996	0.991
#Sig	3	4	3	1	5	5	5

Panel B: FMP correlation with covariance matrix

	FMP				Equity factors		
	CG	CPI	IP	UE	MKT	SMB	HML
				FMP_GX10PC			
Avg *t*	**3.71**	**3.57**	**3.63**	**2.81**	**9.62**	**2.65**	**5.55**
Avg *t* (Sig. CC)	**4.44**	**4.2**	**4.36**	**3.13**	**11.31**	**3.16**	**6.32**
# decades	**4.40**	**4.00**	**4.80**	**3.60**	**4.60**	**3.20**	**4.40**
				FMP_GX202			
Avg *t*	1	1.41	1.24	1.30	**11.15**	**6.65**	**3.75**
Avg t (Sig. CC)	1.08	1.68	1.26	1.44	**20.93**	**12.48**	**6.76**
# decades	1.40	**2.20**	1.40	**1.80**	**3.00**	**2.80**	**3.00**
				FMP_GX202_LASSO			
Avg *t*	1.38	**2.96**	**2.18**	1.68	**11.49**	**6.25**	**4.47**
Avg *t* (Sig. CC)	1.56	**3.98**	**2.66**	1.81	**17.07**	**9.23**	**6.40**
# decades	1.80	3.60	3.20	2.20	**3.80**	**2.80**	**3.80**

Panel C: Risk premium estimation of FMPs

	Intercept	CG	CPI	IP	UE	MKT	SMB	HML
FMP_GX10PC	0.541***	0.010*	–0.001	–0.005*	0.001			
	(4.572)	(1.934)	(–0.370)	(–1.696)	(1.273)			
FMP_GX202	0.742***	0.022	0.001	–0.03	0.007			
	(4.790)	(1.130)	(0.082)	(–1.600)	(1.607)			
FMP_GX202_LASSO	0.409***	0.009**	–0.002	–0.008**	0.007			
	(3.590)	(2.378)	(–1.056)	(–2.369)	(0.898)			
FMP_GX10PC	0.356***	0.004	–0.003	–0.008**	0.000	0.575***	0.227*	–0.336***
	(5.253)	(0.785)	(–1.076)	(–2.525)	(0.638)	(3.051)	(1.779)	(–2.931)
FMP_GX202	0.357***	-0.011	–0.004	–0.026	0.003	0.558***	0.193	–0.288**
	(4.680)	(-0.691)	(–0.502)	(–1.365)	(0.661)	(2.956)	(1.462)	(–2.344)
FMP_GX202_LASSO	0.337***	0.007*	–0.003	–0.008**	0.006	0.559***	0.219*	–0.308**
	(4.502)	(1.835)	(–1.538)	(–2.221)	(0.784)	(2.954)	(1.680)	(–2.537)

Appendix C: Proofs of Propositions

Proof of Proposition 1: The objective function can be written as

$$\left(E(\boldsymbol{R})-\boldsymbol{\beta}E(\boldsymbol{w}'\boldsymbol{R})\right)'\boldsymbol{\Sigma}\left(E(\boldsymbol{R})-\boldsymbol{\beta}E(\boldsymbol{w}'\boldsymbol{R})\right)=\left(E(\boldsymbol{R})-\frac{cov(f,\boldsymbol{R})}{var(\boldsymbol{w}'\boldsymbol{R})}E(\boldsymbol{w}'\boldsymbol{R})\right)'\boldsymbol{\Sigma}\left(E(\boldsymbol{R})-\frac{cov(f,\boldsymbol{R})}{var(\boldsymbol{w}'\boldsymbol{R})}E(\boldsymbol{w}'\boldsymbol{R})\right).$$

Here, $\boldsymbol{\beta}=\frac{cov(f,\boldsymbol{R})}{var(f)}=\frac{cov(f,\boldsymbol{R})}{var(\boldsymbol{w}'\boldsymbol{R})}$. Recall that $\boldsymbol{w_s}=\frac{\boldsymbol{w}}{var(\boldsymbol{w}'\boldsymbol{R})}$. The above equation can be expanded to

$$\left(E(\boldsymbol{R})-\frac{cov(f,\boldsymbol{R})}{var(\boldsymbol{w}'\boldsymbol{R})}E(\boldsymbol{w}'\boldsymbol{R})\right)'\boldsymbol{\Sigma}\left(E(\boldsymbol{R})-\frac{cov(f,\boldsymbol{R})}{var(\boldsymbol{w}'\boldsymbol{R})}E(\boldsymbol{w}'\boldsymbol{R})\right)=E(\boldsymbol{R})'\boldsymbol{\Sigma}E(\boldsymbol{R})-2(cov(f,\boldsymbol{R})E(\boldsymbol{w_s}'\boldsymbol{R}))'\boldsymbol{\Sigma}E(\boldsymbol{R})+(cov(f,\boldsymbol{R})E(\boldsymbol{w_s}'\boldsymbol{R}))'\boldsymbol{\Sigma}(cov(f,\boldsymbol{R})E(\boldsymbol{w_s}'\boldsymbol{R})).$$

Taking the first-order condition with respect to $\boldsymbol{w_s}$ and noting that $E(\boldsymbol{w_s}'\boldsymbol{R})$ is a scalar, we obtain:

$$0=-2cov(f,\boldsymbol{R})'\boldsymbol{\Sigma}E(\boldsymbol{R})E(\boldsymbol{R})+2\left(cov(f,\boldsymbol{R})\right)'\boldsymbol{\Sigma}\left(cov(f,\boldsymbol{R})\right)E(\boldsymbol{R})E(\boldsymbol{R})'\boldsymbol{w_s}.$$

By rearranging the above equation, we obtain the equation (4).

Proof of Proposition 2: From equation (17) (the first-pass regression) and using the assumption in the proposition, the estimated coefficient is

$$\hat{\beta}^i\to\frac{cov(\tilde{f},R^i)}{var(\tilde{f})}=\frac{cov(\tilde{f},R^i)}{var(f)c}=\frac{\beta}{c}.$$

The second equality is satisfied, because the commensurable component f and the measurement error ε_f are uncorrelated.

In the second pass, following Shanken (1992), the true model becomes

$$\boldsymbol{R}_t'=(f_t-E(f)+\gamma)\boldsymbol{\beta}'+\boldsymbol{\eta_t}.$$

The estimated coefficient in equation (18), therefore, becomes

$$\hat{\lambda}_t\to\frac{cov(\hat{\beta},R_t)}{var(\hat{\beta})}\to\frac{cov(\frac{\beta}{c},R_t)}{var(\frac{\beta}{c})}=c(f_t-E(f)+\gamma).$$

Proof of proposition 3: (1) In the first-pass regression, when T converges to infinity, we have

$$\hat{\beta}_1^i \to \frac{1}{DET_1}\left(var(f_2)cov\left(\tilde{f}_1, R^i\right) - cov(f_1, f_2)cov\left(f_2, R^i\right)\right).$$

Use $\boldsymbol{R^i} = \boldsymbol{\alpha^i} + \boldsymbol{\beta_1^i f_1} + \boldsymbol{\beta_2^i f_2} + \boldsymbol{\varepsilon^i}$ to replace $\boldsymbol{R^i}$ and note that $\varepsilon_{f_1}$is uncorrelated with factors and returns, resulting in

$$\hat{\beta}_1^i \to \frac{1}{DET_1}\beta_1^i(var(f_1)var(f_2) - cov(f_1, f_2)^2).$$

Similarly,

$$\hat{\beta}_2^i \to \frac{1}{DET_1}\left((var(f_1) + var\left(\varepsilon_{f_1}\right))cov\left(f_2, R^i\right) - cov(f_1, f_2)cov\left(\tilde{f}_1, R^i\right)\right)$$

$$\to \beta_2^i + \frac{1}{DET_1}\left(var\left(\varepsilon_{f_1}\right)(\beta_2^1 cov(f_1, f_2) + \beta_2^i var(f_2))\right).$$

(2) In the second pass, regress returns on B_1^i and B_2^i. Following Shanken (1992), the true model becomes

$$\boldsymbol{R}_t' = (f_{1t} - E(f_1) + \gamma_1)\boldsymbol{\beta_1}' + (f_{2t} - E(f_2) + \gamma_2)\boldsymbol{\beta_2}' + \boldsymbol{\eta_t}.$$

The estimated coefficient for the factor 1 when both *N* and T go to infinity converges to:

$$\hat{\lambda}_{1t} \to \frac{1}{DET_2}\frac{var(f_1)var(f_2) - cov(f_1, f_2)^2}{DET_1}(\overline{var}\left(B_2^i\right)\overline{cov}\left(\beta_1^i, R^i\right) - \overline{cov}(\beta_1^i, B_2^i)\overline{cov}(B_2^i, R^i))$$

$$\to \frac{1}{DET_2}\frac{var(f_1)var(f_2) - cov(f_1, f_2)^2}{DET_1}\left(\left(\overline{var}\left(B_2^i\right)\overline{var}\left(\beta_1^i\right) - \overline{cov}\left(\beta_1^i, B_2^i\right)^2\right)\gamma_{1t} + (\overline{var}\left(B_2^i\right)\overline{cov}\left(\beta_1^i, \beta_2^i\right) - \right.$$
$$\left.\overline{cov}(\beta_1^i, B_2^i)\overline{cov}(\beta_2^i, B_2^i))\gamma_{2t}\right).$$

(3) When $cov(f_1, f_2) = 0$, $B_2^i = \beta_2^i\left(1 + \frac{var(\varepsilon_{f_1})var(f_2)}{DET_1}\right)$, then

$$\overline{var}\left(B_2^i\right)\overline{cov}\left(\beta_1^i, \beta_2^i\right) - \overline{cov}\left(\beta_1^i, B_2^i\right)\overline{cov}\left(\beta_2^i, B_2^i\right) = \left(1 + \frac{var(\varepsilon_{f_1})var(f_2)}{DET_1}\right)^2\left(\overline{var}\left(\beta_2^i\right)\overline{cov}\left(\beta_1^i, \beta_2^i\right) - \right.$$
$$\left.\overline{cov}\left(\beta_1^i, \beta_2^i\right)\overline{var}\left(\beta_2^i\right)\right) = 0$$. Hence, $w_2 = 0$.

Proof of proposition 4: $cov(u_{12}, f_1) = cov\left(f_2 - \frac{cov(\tilde{f}_1, f_2)}{var(\tilde{f}_1)}\tilde{f}_1, f_1\right) = cov(f_2, f_1)(1 - \frac{var(f_1)}{var(f_1) + var(\varepsilon_{f_1})}) \neq 0$.

Proof of proposition 5: In the first pass, we regress the return only on factor 1 with regression (24), even if the true model follows regression (23). Hence, as T converges to infinity, we get

$$\hat{\beta}_1^{i*} \to \frac{cov(\tilde{f}_1, R^i)}{var(\tilde{f}_1)} = \frac{cov(\tilde{f}_1, \alpha^i + \beta_1^{i*} f_1^* + \beta_2^{i*} f_2^* + \varepsilon^i)}{var(f_1) + var(\varepsilon_{f_1})}.$$

By definition, $f_1^* = f_1, cov(f_1^*, f_2^*) = 0$. Applying regularity assumptions about the measurement error and regression residuals results in

$$\frac{cov(\tilde{f}_1, \alpha^i + \beta_1^{i*} f_1^* + \beta_2^{i*} f_2^* + \varepsilon^i)}{var(f_1) + var(\varepsilon_{f_1})} \to \frac{var(f_1)}{var(f_1) + var(\varepsilon_{f_1})} \beta_1^{i*} = \frac{\beta_1^{i*}}{c}.$$

In the second pass, we run regression (26), but the true model is

$$\boldsymbol{R_t}' = \alpha_t + \lambda_{1t} \widehat{\boldsymbol{\beta}}_1^{*\prime} + \lambda_{2t} \widehat{\boldsymbol{\beta}}_{\boldsymbol{2}}^{*\prime} + \boldsymbol{\nu_t},$$

where $\lambda_{1t} = f_{1t}^* - E(f_1^*) + \gamma_1^* = f_{1t} - E(f_1) + \gamma_1,$ and $\lambda_{2t} = f_{2t}^* - E(f_2^*) + \gamma_2^*$.

With the regularity assumptions, as both N and T go to infinity, the estimated coefficient for factor 1 in regression (26) becomes

$$\hat{\lambda}_{1t} \to \frac{\overline{cov}(\frac{\beta_1^{i*}}{c}, R^i)}{\overline{var}(\frac{\beta_1^{i*}}{c})} = c \frac{\overline{cov}(\beta_1^{i*}, \alpha_t + \lambda_{1t} \widehat{\beta}_1^{i*\prime} + \lambda_{2t} \widehat{\beta}_2^{i*\prime} + \nu_t)}{\overline{var}(\beta_1^{i*})}.$$

Since $\overline{cov}(\beta_1^{i*}, \beta_2^{i*}) \to 0$ as N goes to infinity, ν_t is uncorrelated with β_1^{i*}:

$$\frac{\overline{cov}(\beta_1^{i*}, \alpha_t + \lambda_{1t} \widehat{\beta}_1^{i*\prime} + \lambda_{2t} \widehat{\beta}_2^{i*\prime} + \nu_t)}{\overline{var}(\beta_1^{i*})} \to \lambda_{1t}.$$

Proof of proposition 6: From regression (27) and the regularity assumptions, when T converges to infinity, we get

$$\widehat{\boldsymbol{b}} \to \boldsymbol{V}^{-1} cov(\tilde{f}_1, \boldsymbol{R}) = \boldsymbol{V}^{-1} \boldsymbol{\beta}_1^* var(f_1).$$

Therefore, when N converges to infinity, since beta and the regression residuals are uncorrelated, we find that

$$\frac{1}{N} \widehat{\boldsymbol{b}}' \boldsymbol{R}_t \to \frac{1}{N} \boldsymbol{\beta}_1^* \boldsymbol{V}^{-1} (a + \boldsymbol{\beta}_1^* \boldsymbol{f}_{1t} + \boldsymbol{\beta}_2^* \boldsymbol{f}_{2t}) var(f_1).$$

Proof of proposition 7: (1) When T converges to infinity, using proposition 5 yields the following:

$$\frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_{1t} \to c\frac{1}{T}\sum_{t=1}^{T}(f_{1t} - E(f_1) + \gamma_1) \to c\gamma_1.$$

(2) When T is finite, in the first-pass regression (24), for any asset i,

$$\hat{\beta}_1^i = \frac{\frac{1}{T}\sum_{t=1}^{T}((\tilde{f}_{1t} - \frac{1}{T}\sum_{s=1}^{T}\tilde{f}_{1s})(R_t^i - \frac{1}{T}\sum_{s=1}^{T}R_s^i))}{\frac{1}{T}\sum_{t=1}^{T}((\tilde{f}_{1t} - \frac{1}{T}\sum_{s=1}^{T}\tilde{f}_{1s})^2)}.$$

Recalling that $\tilde{f}_{1t} = f_{1t} + \varepsilon_{f_{1,t}}$, the goal is to prove that any term related to measurement error $\varepsilon_{f_{1,t}}$, is of order $O(\frac{1}{\sqrt{T}})$. To this end, we define $O(\frac{1}{\sqrt{T}}, \varepsilon_{f_1})$ as the error that comes from the measurement error of order $O(\frac{1}{\sqrt{T}})$. And $O(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i)$ is the error that comes from measurement error of order $O(\frac{1}{\sqrt{T}})$, which also depends on beta or regression residuals of the i'th testing asset. Similarly, we define $O\left(\frac{1}{\sqrt{T}}, Other\right)$ as the error that does not come from the measurement error of order $O(\frac{1}{\sqrt{T}})$.

The denominator of $\hat{\beta}_1^i$ can be written as

$$\frac{1}{T}\sum_{t=1}^{T}\left((\tilde{f}_{1t} - \frac{1}{T}\sum_{s=1}^{T}\tilde{f}_{1s})^2\right) = \frac{1}{T}\sum_{t=1}^{T}\left((f_{1t} - \frac{1}{T}\sum_{s=1}^{T}f_{1s})^2\right) + \frac{1}{T}\sum_{t=1}^{T}\left((\varepsilon_{f_1t} - \frac{1}{T}\sum_{s=1}^{T}\varepsilon_{f_1s})^2\right) + 2\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - \frac{1}{T}\sum_{s=1}^{T}f_{1s})(\varepsilon_{f_1t} - \frac{1}{T}\sum_{s=1}^{T}\varepsilon_{f_1s})) = var(f_1) + var(\varepsilon_{f_1}) + \left(\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))^2) - var(f_1)\right) + \left(\frac{1}{T}\sum_{t=1}^{T}((\varepsilon_{f_1t})^2) - var(\varepsilon_{f_1})\right) + 2\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))(\varepsilon_{f_1t})) + O(\frac{1}{T}).$$

Here

$\left(\frac{1}{T}\sum_{t=1}^{T}((\varepsilon_{f_1t})^2) - var(\varepsilon_{f_1})\right) + 2\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))(\varepsilon_{f_1t}))$ depends on ε_{f_1t}, and is in the order of $O(\frac{1}{\sqrt{T}})$, so it is $O(\frac{1}{\sqrt{T}}, \varepsilon_{f_1})$. $\left(\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))^2) - var(f_1)\right)$ is also of order $O(\frac{1}{\sqrt{T}})$, but does not depend on ε_{f_1t}, so it is $O\left(\frac{1}{\sqrt{T}}, Other\right)$. Hence, the denominator can be written as

$$var(f_1) + O(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}) + O(\frac{1}{\sqrt{T}}, Other).$$

The numerator can be written as

$$\frac{1}{T}\sum_{t=1}^{T}((\tilde{f}_{1t} - \frac{1}{T}\sum_{s=1}^{T}\tilde{f}_{1s})(R_t^i - \frac{1}{T}\sum_{s=1}^{T}R_s^i)) = \frac{1}{T}\sum_{t=1}^{T}((f_{1t} - \frac{1}{T}\sum_{s=1}^{T}f_{1s} + \varepsilon_{f_1t} - \frac{1}{T}\sum_{s=1}^{T}\varepsilon_{f_1s})(\alpha^i + \beta_1^{i*}f_{1t}^* + \varsigma_t^i - \frac{1}{T}\sum_{s=1}^{T}(\alpha^i + \beta_1^{i*}f_{1t}^* + \varsigma_t^i))).$$

Since $f_{1t}^* = f_{1t}$, the above equation can be written as

$\beta_1^{i*}var(f_1) + \beta_1^{i*}\left(\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))^2) - var(f_1)\right) + \beta_1^{i*}\left(\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))(\varepsilon_{f_1 t}))\right) +$
$\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))\varsigma_t^i) + \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{f_1 t}\varsigma_t^i) + O(\frac{1}{T})$.

Therefore, in the numerator,

$$O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) = \beta_1^{i*}\left(\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))(\varepsilon_{f_1 t}))\right) + \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{f_1 t}\varsigma_t^i), \text{ and}$$

$$O\left(\frac{1}{\sqrt{T}}, Other\right) = \beta_1^{i*}\left(\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))^2) - var(f_1)\right) + \frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))\varsigma_t^i).$$

The numerator can be written as

$$\beta_1^{i*}var(f_1) + O(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i) + O(\frac{1}{\sqrt{T}}, Other) + O(\frac{1}{T}).$$

Using the formulas of numerator and denominator,

$$\hat{\beta}_1^i = \frac{\beta_1^{i*}var(f_1) + O(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i) + O(\frac{1}{\sqrt{T}}, Other) + O(\frac{1}{T})}{var(f_1) + var(\varepsilon_{f_1}) + O(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}) + O(\frac{1}{\sqrt{T}}, Other) + O(\frac{1}{T})}.$$

With the Taylor approximation up to the first order,

$$\hat{\beta}_1^i = \frac{1}{c}\beta_1^{i*} + \frac{1}{var(f_1) + var(\varepsilon_{f_1})}\left(\beta_1^{i*}\left(\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))(\varepsilon_{f_1 t}))\right) + \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_{f_1 t}\varsigma_t^i\right)\right) +$$
$$\frac{\beta_1^{i*}var(f_1)}{(var(f_1) + var(\varepsilon_{f_1}))^2}(\left(\frac{1}{T}\sum_{t=1}^{T}\left((\varepsilon_{f_1 t})^2\right) - var\left(\varepsilon_{f_1}\right)\right) + 2\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))(\varepsilon_{f_1 t}))) + O(\frac{1}{\sqrt{T}}, Other) +$$
$$O(\frac{1}{T}).$$

Hence, the estimated coefficient $\hat{\beta}_1^i = \frac{1}{c}\beta_1^{i*} + O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O(\frac{1}{T})$. The error term comes from the factor measurement error ($O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right)$) is of order $O(\frac{1}{\sqrt{T}})$.

In the second-pass regression (26), the estimated coefficient is

$$\hat{\lambda}_{1t} = \frac{\frac{1}{N}(\sum_{i=1}^{N}(\hat{\beta}_1^i - \frac{1}{N}\sum_{j=1}^{N}\hat{\beta}_1^j)(R_t^i - \frac{1}{N}\sum_{j=1}^{N}R_t^j))}{\frac{1}{N}(\sum_{i=1}^{N}(\hat{\beta}_1^i - \frac{1}{N}\sum_{j=1}^{N}\hat{\beta}_1^j)^2)}.$$

Replacing $\hat{\beta}_1^i = \frac{1}{c}\beta_1^{i*} + O(\frac{1}{\sqrt{T}}, \varepsilon_{f_1 t}, i) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O\left(\frac{1}{T}\right)$ in denominator, we obtain:

$$\frac{1}{N}\left(\sum_{i=1}^{N}\left(\hat{\beta}_1^i - \frac{1}{N}\sum_{j=1}^{N}\hat{\beta}_1^j\right)^2\right) = \frac{1}{c^2}\overline{var}(\beta_1^*) + \frac{1}{N}\left(\sum_{i=1}^{N}(\beta_1^{i*} - \frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*})(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) -\right.$$
$$\left.\frac{1}{N}\sum_{j=1}^{N}O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right))\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O\left(\frac{1}{T}\right),$$

where

$O(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i) = \frac{\beta_1^{i*}\left(\frac{1}{T}\sum_{t=1}^{T}((f_{1t}-E(f_1))(\varepsilon_{f_1 t}))\right)}{var(f_1)+var(\varepsilon_{f_1})} + \frac{\beta_1^{i*}var(f_1)}{(var(f_1)+var(\varepsilon_{f_1}))^2}(\left(\frac{1}{T}\sum_{t=1}^{T}\left((\varepsilon_{f_1 t})^2\right) - var\left(\varepsilon_{f_1}\right)\right) +$
$2\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))(\varepsilon_{f_1 t}))).$

Hence, $\frac{1}{N}\left(\sum_{i=1}^{N}\left(\hat{\beta}_1^i - \frac{1}{N}\sum_{j=1}^{N}\hat{\beta}_1^j\right)^2\right) = \frac{1}{c^2}\overline{var}(\beta_1^*) + O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O\left(\frac{1}{T}\right).$ [4]

Similarly, the numerator is

$\frac{1}{N}\left(\sum_{i=1}^{N}\left(\hat{\beta}_1^i - \frac{1}{N}\sum_{j=1}^{N}\hat{\beta}_1^j\right)\left(R_t^i - \frac{1}{N}\sum_{j=1}^{N}R_t^j\right)\right) = \frac{1}{c}\overline{var}(\beta_1^*)(f_{1t} - E(f_1) + \gamma_1) +$
$\frac{1}{N}\left(\sum_{i=1}^{N}\left(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N}O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right)\right)\left(R_t^i - \frac{1}{N}\sum_{j=1}^{N}R_t^j\right)\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O\left(\frac{1}{T}\right) =$
$\frac{1}{c}\overline{var}(\beta_1^*)(f_{1t} - E(f_1) + \gamma_1) + O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O\left(\frac{1}{T}\right).$

With Taylor expansion,

$\hat{\lambda}_{1t} = c(f_{1t} - E(f_1) + \gamma_1) + \frac{1}{\frac{1}{c^2}\overline{var}(\beta_1^*)}(\frac{1}{N}\left(\sum_{i=1}^{N}\left(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N}O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right)\right)\left(R_t^i -\right.\right.$
$\left.\left.\frac{1}{N}\sum_{j=1}^{N}R_t^j\right)\right)) + \frac{\frac{1}{c}\overline{var}(\beta_1^*)(f_{1t}-E(f_1)+\gamma_1)}{\left(\frac{1}{c^2}\overline{var}(\beta_1^*)\right)^2}\frac{1}{N}\left(\sum_{i=1}^{N}\left(\beta_1^{i*} - \frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*}\right)\left(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) -\right.\right.$
$\left.\left.\frac{1}{N}\sum_{j=1}^{N}O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right)\right)\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O\left(\frac{1}{T}\right).$

Hence, $\hat{\lambda}_{1t} = c(f_{1t} - E(f_1) + \gamma_1) + O(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, t) + O(\frac{1}{\sqrt{T}}, Other) + O(\frac{1}{T}).$

[4] Here, we treat N as a finite number. Hence, $\frac{1}{N}\left(\sum_{i=1}^{N}(\beta_1^{i*} - \frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*})(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right) - \frac{1}{N}\sum_{j=1}^{N}O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right))\right)$ is $O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}\right)$. When N is large, $\frac{1}{N}\left(\sum_{i=1}^{N}(\beta_1^{i*} - \frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*})(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right) - \frac{1}{N}\sum_{j=1}^{N}O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right))\right) = \frac{1}{N}\left(\sum_{i=1}^{N}(\beta_1^{i*} - \frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*})^2(O^1\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right) - \frac{1}{N}\sum_{j=1}^{N}O^1\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right))\right) \to \overline{var}(\beta_1^*)(O^1\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right) - \frac{1}{N}\sum_{j=1}^{N}O^1\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right))$, where $O^1(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i) = \frac{\left(\frac{1}{T}\sum_{t=1}^{T}((f_{1t}-E(f_1))(\varepsilon_{f_1 t}))\right)}{var(f_1)+var(\varepsilon_{f_1})} + \frac{var(f_1)}{(var(f_1)+var(\varepsilon_{f_1}))^2}(\left(\frac{1}{T}\sum_{t=1}^{T}\left((\varepsilon_{f_1 t})^2\right) - var\left(\varepsilon_{f_1}\right)\right) + 2\frac{1}{T}\sum_{t=1}^{T}((f_{1t} - E(f_1))(\varepsilon_{f_1 t})))$. The proof will be the same since $O(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i)$ is $O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}\right)$.

Here, the term $O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, t\right)$ is $\frac{1}{\frac{1}{c^2}\overline{var}(\beta_1^*)}(\frac{1}{N}\left(\sum_{i=1}^{N}\left(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N} O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right)\right)\left(R_t^i - \frac{1}{N}\sum_{j=1}^{N} R_t^j\right)\right)) + \frac{\frac{1}{c}\overline{var}(\beta_1^*)(f_{1t}-E(f_1)+\gamma_1)}{\left(\frac{1}{c^2}\overline{var}(\beta_1^*)\right)^2}\frac{1}{N}\left(\sum_{i=1}^{N}\left(\beta_1^{i*} - \frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*}\right)\left(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N} O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right)\right)\right)$.

Take the average of the $O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, t\right)$ over time,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\frac{1}{c^2}\overline{var}(\beta_1^*)}(\frac{1}{N}\left(\sum_{i=1}^{N}\left(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N} O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right)\right)\left(R_t^i - \frac{1}{N}\sum_{j=1}^{N} R_t^j\right)\right)) + \frac{\frac{1}{c}\overline{var}(\beta_1^*)(f_{1t}-E(f_1)+\gamma_1)}{\left(\frac{1}{c^2}\overline{var}(\beta_1^*)\right)^2}\frac{1}{N}\left(\sum_{i=1}^{N}\left(\beta_1^{i*} - \frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*}\right)\left(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N} O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right)\right)\right) =$$

$$\frac{1}{\frac{1}{c^2}\overline{var}(\beta_1^*)}(\frac{1}{N}\left(\sum_{i=1}^{N}\left(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N} O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right)\right)\frac{1}{T}\sum_{t=1}^{T}(R_t^i - \frac{1}{N}\sum_{j=1}^{N} R_t^j)\right)) + \frac{\frac{1}{c}\overline{var}(\beta_1^*)\gamma_1}{\left(\frac{1}{c^2}\overline{var}(\beta_1^*)\right)^2}\frac{1}{N}\left(\sum_{i=1}^{N}\left(\beta_1^{i*} - \frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*}\right)\left(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N} O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, j\right)\right)\right) + O(\frac{1}{T}).$$

The first two terms are both of order $O(\frac{1}{\sqrt{T}})$ and depends on $\varepsilon_{f_1 t}$, the average of $O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, t\right) = O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}\right) + O(\frac{1}{\sqrt{T}}, Other) + O(\frac{1}{T})$. Hence,

$\frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_{1t} = c\gamma_1 + O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}\right) + O(\frac{1}{\sqrt{T}}, Other) + O(\frac{1}{T})$. We have shown that, in the average value of the coefficients ($\frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_{1t}$), the finite sample error that comes from factor measurement error (ε_{f_1}) is of order $O(\frac{1}{\sqrt{T}})$.

(3) When we use $\hat{\lambda}_{1t}$ as the FMP to repeat the analysis, in the first-pass regression (24), we have

$$\hat{\beta}_1^i = \frac{\frac{1}{T}\sum_{t=1}^{T}((\hat{\lambda}_{1t} - \frac{1}{T}\sum_{s=1}^{T}\hat{\lambda}_{1s})(R_t^i - \frac{1}{T}\sum_{s=1}^{T}R_s^i))}{\frac{1}{T}\sum_{t=1}^{T}(\hat{\lambda}_{1t} - \frac{1}{T}\sum_{s=1}^{T}\hat{\lambda}_{1s})^2}.$$

Following a similar derivation to the one carried out before, we can show that the denominator

$$\frac{1}{T}\sum_{t=1}^{T}\left((\hat{\lambda}_{1t} - \frac{1}{T}\sum_{s=1}^{T}\hat{\lambda}_{1s})^2\right) = \frac{1}{T}\sum_{t=1}^{T}\left((cf_{1t} - \frac{1}{T}\sum_{s=1}^{T}cf_{1s})^2\right) + \frac{1}{T}\sum_{t=1}^{T}((cf_{1t} - \frac{1}{T}\sum_{s=1}^{T}cf_{1s})(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) + O(\frac{1}{\sqrt{T}}, Other) + O(\frac{1}{T}))) + \frac{1}{T}\sum_{t=1}^{T}(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) + O(\frac{1}{\sqrt{T}}, Other) + O(\frac{1}{T}))^2.$$

Given that measure errors $\varepsilon_{f_1,t}$ are uncorrelated with factors f_{1t} and the returns, $\frac{1}{T}\sum_{t=1}^{T}((cf_{1t} - \frac{1}{T}\sum_{s=1}^{T}cf_{1s})(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) + O(\frac{1}{\sqrt{T}}, Other) + O(\frac{1}{T}))) + \frac{1}{T}\sum_{t=1}^{T}(O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right) + O(\frac{1}{\sqrt{T}}, Other) + O(\frac{1}{T}))^2 = O\left(\frac{1}{\sqrt{T}}, Other\right) + O(\frac{1}{T}, \varepsilon_{f_1}) + O(\frac{1}{T}, Other)$, where $O(\frac{1}{T}, \varepsilon_{f_1}, i)$ and $O(\frac{1}{T}, \varepsilon_{f_1})$ are the errors that comes from the measurement error of order $O(\frac{1}{T})$, and $O\left(\frac{1}{T}, Other\right)$ as the error that does not come from the measurement error of order $O(\frac{1}{T})$. Hence,

$$\frac{1}{T}\sum_{t=1}^{T}\left((\hat{\lambda}_{1t} - \frac{1}{T}\sum_{s=1}^{T}\hat{\lambda}_{1s})^2\right) = \frac{1}{T}\sum_{t=1}^{T}\left((cf_{1t} - \frac{1}{T}\sum_{s=1}^{T}cf_{1s})^2\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O(\frac{1}{T}, \varepsilon_{f_1}, i) + O(\frac{1}{T}, Other).$$

Similarly, the numerator becomes

$$\frac{1}{T}\sum_{t=1}^{T}((\hat{\lambda}_{1t} - \frac{1}{T}\sum_{s=1}^{T}\hat{\lambda}_{1s})(R_t^i - \frac{1}{T}\sum_{s=1}^{T}R_s^i)) = \beta_1^i\frac{1}{T}\sum_{t=1}^{T}\left((cf_{1t} - \frac{1}{T}\sum_{s=1}^{T}cf_{1s})^2\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O\left(\frac{1}{T}, \varepsilon_{f_1}, i\right) + O(\frac{1}{T}).$$

Applying the Taylor expansion, the estimated error that comes from the measurement error ε_{f_1}, is only of order $O(\frac{1}{T})$.

In the second-pass regression (26), when we use the estimated beta above as the independent variable, the error of the estimated coefficient that comes from the measurement error ε_{f_1}, is also of order $O(\frac{1}{T})$. Specifically, the estimated coefficient

$$\hat{\lambda}_{1t} = \frac{\frac{1}{N}(\sum_{i=1}^{N}(\hat{\beta}_1^i - \frac{1}{N}\sum_{j=1}^{N}\hat{\beta}_1^j)(R_t^i - \frac{1}{N}\sum_{j=1}^{N}R_t^j))}{\frac{1}{N}(\sum_{i=1}^{N}(\hat{\beta}_1^i - \frac{1}{N}\sum_{j=1}^{N}\hat{\beta}_1^j)^2)}.$$

Replacing $\hat{\beta}_1^i = \frac{1}{c}\beta_1^{i*} + O\left(\frac{1}{T}, \varepsilon_{f_1t}, i\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O(\frac{1}{T}, Other)$, following the same analysis as before,

$$\frac{1}{N}\left(\sum_{i=1}^{N}\left(\hat{\beta}_1^i - \frac{1}{N}\sum_{j=1}^{N}\hat{\beta}_1^j\right)^2\right) = \frac{1}{c^2}\overline{var}(\beta_1^*) + \frac{1}{N}\left(\sum_{i=1}^{N}(\beta_1^{i*} - \frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*})(O\left(\frac{1}{T}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N}O\left(\frac{1}{T}, \varepsilon_{f_1}, j\right))\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O\left(\frac{1}{T}, Other\right).$$

Hence, $\frac{1}{N}\left(\sum_{i=1}^{N}\left(\hat{\beta}_1^i - \frac{1}{N}\sum_{j=1}^{N}\hat{\beta}_1^j\right)^2\right) = \frac{1}{c^2}\overline{var}(\beta_1^*) + O\left(\frac{1}{T}, \varepsilon_{f_1}, i\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O\left(\frac{1}{T}, Other\right)$.

Similarly, we can show that

$$\frac{1}{N}\left(\sum_{i=1}^{N}\left(\hat{\beta}_1^i - \frac{1}{N}\sum_{j=1}^{N}\hat{\beta}_1^j\right)\left(R_t^i - \frac{1}{N}\sum_{j=1}^{N}R_t^j\right)\right) = \frac{1}{c}\overline{var}(\beta_1^*)(f_{1t} - E(f_1) + \gamma_1) + \frac{1}{N}\left(\sum_{i=1}^{N}\left(O\left(\frac{1}{T}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N}O\left(\frac{1}{T}, \varepsilon_{f_1}, j\right)\right)\left(R_t^i - \frac{1}{N}\sum_{j=1}^{N}R_t^j\right)\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O\left(\frac{1}{T}\right) = \frac{1}{c}\overline{var}(\beta_1^*)(f_{1t} - E(f_1) + \gamma_1) + O\left(\frac{1}{T}, \varepsilon_{f_1}\right) + O\left(\frac{1}{\sqrt{T}}, Other\right) + O\left(\frac{1}{T}, Other\right).$$

Hence, $\hat{\lambda}_{1t} = c(f_{1t} - E(f_1) + \gamma_1) + O(\frac{1}{T}, \varepsilon_{f_1}, t) + O(\frac{1}{\sqrt{T}}, Other) + O(\frac{1}{T}, Other)$.

Here, the term $O\left(\frac{1}{T}, \varepsilon_{f_1}, t\right)$ is $\frac{1}{\frac{1}{c^2}\overline{var}(\beta_1^*)}(\frac{1}{N}\left(\sum_{i=1}^{N}\left(O\left(\frac{1}{T}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N} O\left(\frac{1}{T}, \varepsilon_{f_1}, j\right)\right)\left(R_t^i - \frac{1}{N}\sum_{j=1}^{N} R_t^j\right)\right)) + \frac{\frac{1}{c}\overline{var}(\beta_1^*)(f_{1t} - E(f_1) + \gamma_1)}{\left(\frac{1}{c^2}\overline{var}(\beta_1^*)\right)^2}\frac{1}{N}\left(\sum_{i=1}^{N}\left(\beta_1^{i*} - \frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*}\right)\left(O\left(\frac{1}{T}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N} O\left(\frac{1}{T}, \varepsilon_{f_1}, j\right)\right)\right)$.

Take the average of the $O\left(\frac{1}{T}, \varepsilon_{f_1}, t\right)$ over time,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\frac{1}{c^2}\overline{var}(\beta_1^*)}(\frac{1}{N}\left(\sum_{i=1}^{N}\left(O\left(\frac{1}{T}, \varepsilon_{f_1}, j\right) - \frac{1}{N}\sum_{j=1}^{N} O\left(\frac{1}{T}, \varepsilon_{f_1}, j\right)\right)\left(R_t^i - \frac{1}{N}\sum_{j=1}^{N} R_t^j\right)\right)) + \frac{\frac{1}{c}\overline{var}(\beta_1^*)(f_{1t} - E(f_1) + \gamma_1)}{\left(\frac{1}{c^2}\overline{var}(\beta_1^*)\right)^2}\frac{1}{N}\left(\sum_{i=1}^{N}\left(\beta_1^{i*} - \frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*}\right)\left(O\left(\frac{1}{T}, \varepsilon_{f_1}, i\right) - \frac{1}{N}\sum_{j=1}^{N} O\left(\frac{1}{T}, \varepsilon_{f_1}, j\right)\right)\right) =$$

$$\frac{1}{\frac{1}{c^2}\overline{var}(\beta_1^*)}\left(\frac{1}{N}\left(\sum_{i=1}^{N}\left(O\left(\frac{1}{T}, \varepsilon_{f_1}, j\right) - \frac{1}{N}\sum_{j=1}^{N} O\left(\frac{1}{T}, \varepsilon_{f_1}, j\right)\right)\frac{1}{T}\sum_{t=1}^{T}(R_t^i - \frac{1}{N}\sum_{j=1}^{N} R_t^j)\right)\right) + \frac{\frac{1}{c}\overline{var}(\beta_1^*)\gamma_1}{\left(\frac{1}{c^2}\overline{var}(\beta_1^*)\right)^2}\frac{1}{N}\left(\sum_{i=1}^{N}\left(\beta_1^{i*} - \frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*}\right)\left(O\left(\frac{1}{T}, \varepsilon_{f_1}, j\right) - \frac{1}{N}\sum_{j=1}^{N} O\left(\frac{1}{T}, \varepsilon_{f_1}, j\right)\right)\right) + O\left(\frac{1}{T}, Other\right) =$$

$$O(\frac{1}{T}, \varepsilon_{f_1}) + O(\frac{1}{T}, Other).$$

Then the estimated risk premium $\frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_{1t} = c\gamma_1 + O\left(\frac{1}{T}, \varepsilon_{f_1}\right) + O(\frac{1}{\sqrt{T}}, Other) + O(\frac{1}{T}, Other)$.

Thus, the measurement-error component in the estimation risk premium in the two-stage approach is also of order $O(\frac{1}{T})$.

Appendix D: Rationale for the three criteria

The first criterion can be derived directly following the calculation of covariance between the FMP return from equation (10)[5] and the corresponding factor:

$$\mathrm{cov}(f, \boldsymbol{R}\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1})$$

$$= \mathrm{cov}(f, (\boldsymbol{\alpha} + \boldsymbol{\beta} f + \boldsymbol{\varepsilon})\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}) = \mathrm{cov}(f, f)$$

The second equation follows the decomposition of returns from Equation (1). The above covariance is not zero as long as the factor has nonzero variance.

Therefore, the first criterion is: An FMP should be correlated with its underlying factor.

The second criterion is: An FMP should be correlated to the systematic risk in returns. To implement this criterion empirically, we apply the same necessary condition as that proposed by Pukthuanthong et al. (2019). This avoids the construction of FMPs that are driven only by idiosyncratic but not by systematic risk.

To examine why the above criterion is necessary, assume that with systematic risk factors $\mathbf{F}^{\boldsymbol{s}}$, the returns can be written as $\boldsymbol{R} = \boldsymbol{a}^{\boldsymbol{s}} + \boldsymbol{\beta}^{\boldsymbol{s}}\boldsymbol{F}^{\boldsymbol{s}} + \boldsymbol{\zeta}$, where $\boldsymbol{\beta}^{\boldsymbol{s}}$ is the loading of the returns on the systematic factor, $\boldsymbol{a}^{\boldsymbol{s}}$ is the mispricing component, and $\boldsymbol{\zeta}$ is the idiosyncratic risk. Placing the returns representation above in Equation (10), we obtain:

$$\boldsymbol{R}\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1} = (\boldsymbol{a}^{\boldsymbol{s}} + \boldsymbol{\beta}^{\boldsymbol{s}}\boldsymbol{F}^{\boldsymbol{s}} + \boldsymbol{\zeta})\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}$$

$$= \boldsymbol{a}^{\boldsymbol{s}}\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1} + \boldsymbol{\beta}^{\boldsymbol{s}}\boldsymbol{F}^{\boldsymbol{s}}\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1} + \boldsymbol{\zeta}\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}.$$

From the above equation, the first term is a constant. The third term is a linear combination of idiosyncratic errors, which is the idiosyncratic risk of the FMP. The second term represents the correlated part of the FMP on the systematic risk. Because this term is a linear combination of systematic risk factors $\boldsymbol{F}^{\boldsymbol{s}}$, if the FMP is not correlated with the systematic risk of stock returns, the second term is a constant. Assume that the idiosyncratic risks can be diversified, $\boldsymbol{\zeta}\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1} \to \mathbf{0}$ (For example, basis assets are individual stocks or the basis asset are portfolios and the idiosyncratic risk for portfolios converges to zero when number of individual

[5] Equation (10) represents the expected value of the FMP return. In this appendix, we instead use the FMP return, which is $\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}$**.**

stocks in each portfolio is large). Hence, the above FMP degenerates to a constant number if the second term is zero, and it cannot represent the risk of the underlying factor. Therefore, only if second term is correlated with the risk factors $\boldsymbol{F^s}$, FMP can present a genuine risk factor.

The third criterion is that factor-mimicking portfolios must price testing assets; that is, they must be associated with risk premiums.

To illustrate this criterion, regress the returns ($\boldsymbol{R}$) on the FMP return. Then the slope coefficient is

$$\frac{\operatorname{cov}\left(\boldsymbol{R'},\boldsymbol{RV^{-1}\beta}\left[\boldsymbol{\beta' V^{-1}\beta}\right]^{-1}\right)}{\operatorname{var}\left(\boldsymbol{RV^{-1}\beta}\left[\boldsymbol{\beta' V^{-1}\beta}\right]^{-1}\right)} = \frac{\operatorname{cov}(\boldsymbol{R'},\boldsymbol{R})\boldsymbol{RV^{-1}\beta}\left[\boldsymbol{\beta' V^{-1}\beta}\right]^{-1}}{\left(\boldsymbol{RV^{-1}\beta}\left[\boldsymbol{\beta' V^{-1}\beta}\right]^{-1}\right)' \operatorname{cov}(\boldsymbol{R'},\boldsymbol{R})\boldsymbol{RV^{-1}\beta}\left[\boldsymbol{\beta' V^{-1}\beta}\right]^{-1}} = \boldsymbol{\beta},$$

which is the beta coefficient of the returns on the factor. Hence, with the non-arbitrage assumption of the APT model, there is a linear relationship between the expected returns and the beta coefficient. That is, $\mathrm{E}(\boldsymbol{R}) = \mu_0 + \mu_1\boldsymbol{\beta}$.

Appendix E

Table E1: Combining Macro Factors and CAY

This table shows the FMP identification when combining the four macro factors with the CAY factor, which supplements the analysis in Table 6 in the text. Panel A shows the average correlations in five subsamples between FMPs and their underlying factors, and the number of significant correlations (# Sig) at the 1% significance level in five subsamples (the first criterion, similar to Table 3). Panel B shows the canonical correlations between FMPs and the principal component of the covariance matrix of asset returns (the second criterion, similar to Table 4). Panel C shows the estimated risk premiums by using FMPs (the third criterion, similar to Table 5). The details of each method are described in Table 1 in the text.

Panel A: Subsample correlations between FMPs and underlying factors

	CG	CPI	IP	UE	CAY	CG*CAY	MKT	SMB	HML
				Cross-sectional approach					
FMP_IV	0.411	0.485	0.663	0.565	0.505	0.518	0.741	0.804	0.710
# Sig	5	5	5	5	4	5	5	5	5
FMP_OLS	0.880	0.846	0.860	0.816	0.746	0.846	0.914	0.932	0.887
# Sig	5	5	5	5	5	5	5	5	5
FMP_LM	0.879	0.822	0.879	0.813	0.733	0.844	0.918	0.960	0.897
# Sig	5	5	5	5	5	5	5	5	5
FMP_Stein	0.880	0.846	0.860	0.816	0.746	0.846	0.914	0.932	0.887
# Sig	5	5	5	5	5	5	5	5	5
				Time-series approach					
FMP_Time-series	0.213	0.342	0.293	0.222	0.213	0.140	1.000	0.420	0.301
# Sig	1	4	3	2	4	1	5	5	3
				Sorting-by-beta approach					
FMP_SB	0.546	0.740	0.733	0.697	0.617	0.529	0.877	0.916	0.854
# Sig	5	5	5	5	5	5	5	5	5

Panel B: Canonical correlations with asymptotic PCs and significance levels of factor candidates

	FMP							
	CG	CPI	IP	UE	CAY	MKT	SMB	HML
Original								
Avg *t*	1.06	1.05	0.97	0.86	1.08	9.03	5.54	3.23
Avg *t* (Sig. CC)	1.21	1.29	0.97	1.00	1.34	22.03	13.36	7.62
# of sig *t*-stat	1.20	1.20	1.00	0.60	1.20	2.80	2.80	2.60
Cross-sectional approach								
FMP_IV								
Avg *t*	1.57	2.21	1.64	1.76	1.63	2.85	2.86	2.98
Avg *t* (Sig. CC)	1.80	2.90	2.02	2.13	1.92	3.95	3.92	4.21
# of sig *t*-stat	3.00	3.40	2.20	2.20	2.20	3.40	3.20	4.40
FMP_OLS								
Avg *t*	2.00	2.64	1.72	2.05	1.81	6.81	6.14	5.03
Avg *t* (Sig. CC)	2.39	3.26	2.00	2.46	2.16	8.97	8.07	6.65
# of sig *t*-stat	3.60	4.20	3.20	2.80	3.00	3.80	3.60	3.20
FMP_Stein								
Avg *t*	2.00	2.64	1.72	2.05	1.81	6.81	6.14	5.03
Avg *t* (Sig. CC)	2.39	3.26	2.00	2.46	2.16	8.97	8.07	6.65
# of sig *t*-stat	3.60	4.20	3.20	2.80	3.00	3.80	3.60	3.20
FMP_LM								
Avg *t*	1.93	2.71	1.71	2.05	1.85	6.47	5.70	4.43
Avg *t* (Sig. CC)	2.32	3.40	2.01	2.53	2.33	8.79	7.71	6.01
# of sig *t*-stat	3.60	4.40	3.40	2.40	2.80	3.20	3.40	3.40
Time-series approach								
Avg. *t*	1.15	1.05	1.18	1.10	0.93	2.88	1.30	1.25
Avg *t* (Sig. CC)	1.41	1.43	1.32	1.32	0.94	7.25	1.99	2.04
# of sig *t*-stat	1.40	1.20	1.60	1.40	0.40	1.80	2.00	1.60
Sorting-by-beta approach								
Avg. *t*	1.45	1.35	1.03	1.31	1.59	5.97	2.02	1.64
Avg *t* (Sig. CC)	1.91	2.00	1.21	1.69	2.27	11.79	3.39	2.32
# of sig *t*-stat	2.60	1.40	0.80	2.00	2.80	3.20	2.80	2.60

Panel C: Test risk premium by using FMPs

	Intercept	CG	CPI	IP	UE	CAY	CG*CAY	MKT	SMB	HML
	Cross-sectional approach									
IV	0.624***	0.227***	-0.078**	-0.016	0.005	-0.197	0.349	0.533*	0.130	-0.550***
	(4.282)	(2.849)	(-2.057)	(-0.180)	(0.114)	(-0.934)	(1.521)	(1.911)	(0.724)	(-3.609)
OLS	0.459***	-0.010	-0.011	-0.034	0.001	-0.025	-0.006	0.492***	0.219	-0.276**
	(5.193)	(-0.441)	(-0.918)	(-1.033)	(0.106)	(-0.487)	(-0.118)	(2.602)	(1.639)	(-2.219)
LM	0.167**	-0.005	-0.015	-0.05	0.004	0.005	-0.023	0.600***	0.222*	-0.194
	(2.202)	(-0.242)	(-1.247)	(-1.604)	(0.517)	(0.104)	(-0.502)	(3.208)	(1.717)	(-1.565)
Stein	0.371***	-0.02	-0.023	-0.173	0.004	-0.177	0.019	0.538***	0.227	-0.324**
	(4.238)	(-0.260)	(-0.776)	(-1.571)	(0.195)	(-0.771)	(0.194)	(2.813)	(1.524)	(-2.255)
	Time-series approach									
Time-series	0.327***	-0.003	-0.003	-0.024***	0.004**	0.027	-0.011	0.548***	0.100*	-0.127***
	(3.290)	(-0.435)	(-0.793)	(-2.702)	(2.551)	(0.951)	(-1.497)	(2.739)	(1.700)	(-3.050)
	Sorting-by-beta approach									
SB	0.588***	0.192	-0.229	-0.147	-0.083	-0.038	-0.053	0.671**	0.568*	-0.637**
	(6.155)	(1.043)	(-1.123)	(-0.849)	(-0.420)	(-0.239)	(-0.263)	(2.096)	(1.698)	(-2.019)