Internet Appendix

Appendix A. Maximum correlation portfolio and its relation to the least mispriced portfolio

The classical factor-mimicking portfolio is constructed by maximizing its correlation with the nontraded factor. Huberman, Kandel, and Stambaugh (1987), Breeden, Gibbons, and Litzenberger (1989), and Roll and Srivastava (2018) provide the theoretical framework for maximum correlation portfolio construction. For a given value of the loading of the risk factor on its mimicking portfolio, β_{fmp} (normally, β_{fmp} is set equal to one), the variance of the FMP (following the same notation, $\boldsymbol{w}' \boldsymbol{R}$) is the reciprocal of the correlation between FMP and f. The following formula illustrates the relation:

$$\beta_{fmp} = corr\left(f, \boldsymbol{w}' \boldsymbol{R}\right) \frac{\sqrt{Var(\boldsymbol{w}' \boldsymbol{R})}}{\sqrt{Var(f)}}. \quad \text{(A1)}$$

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:

$$\min_{\boldsymbol{w}} \boldsymbol{w}' \boldsymbol{V} \boldsymbol{w} + 2(\beta^*_{fmp} - \boldsymbol{w}' \boldsymbol{\beta})\, \lambda, \quad \text{(A2)}$$

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 (A2) and setting β^*_{fmp} equal to one, we find that the optimal weight is

$$\boldsymbol{w} = \boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}\,. \quad \text{(A3)}$$

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

$$E(\boldsymbol{R})' \boldsymbol{w} = E(\boldsymbol{R})' \boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}. \quad \text{(A4)}$$

Recall that from Eq. (4) in the main text, the FMP constructed through the least mispriced portfolio can be written as

$$E(\boldsymbol{R})' \boldsymbol{w} = \frac{\boldsymbol{\beta}' \boldsymbol{\Sigma} E(\boldsymbol{R})}{\boldsymbol{\beta}' \boldsymbol{\Sigma} \boldsymbol{\beta}}.$$

If there is only one factor, and if we replace $\boldsymbol{\Sigma}$ by $\boldsymbol{V}^{-1}$, Eq. (4) in the main text and Eq. (A4) are identical. The maximum correlation portfolio coincides with the least mispriced portfolio. The coincidence is not surprising, given the equivalence between the 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 (A2) is binding), the covariance between the FMP return ($\boldsymbol{w}'\boldsymbol{R}$) and any asset excess return is a linear function of its expected return (Cochrane, 2009), that is,

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

Here, γ_1 is a constant. Since $\boldsymbol{w}'\boldsymbol{R}$ and the zero beta rate, R_0, also should be priced, we have $var(\boldsymbol{w}'\boldsymbol{R}) = \gamma_1 E(\boldsymbol{w}'\boldsymbol{R})$. Combining these equations, we obtain

$$E(\boldsymbol{R}) = \frac{cov(\boldsymbol{w}'\boldsymbol{R},\boldsymbol{R})}{var(\boldsymbol{w}'\boldsymbol{R})} E\left(\boldsymbol{w}'\boldsymbol{R}\right). \quad \text{(A5)}$$

On the other hand,

$$\frac{cov(\boldsymbol{w}'\boldsymbol{R},\boldsymbol{R})}{var(\boldsymbol{w}'\boldsymbol{R})} = \frac{\mathrm{cov}\left(\boldsymbol{R}',\boldsymbol{R}\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}\right)}{var\left(\boldsymbol{R}\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}\right)} = \frac{\mathrm{cov}(\boldsymbol{R}',\boldsymbol{R})\boldsymbol{R}\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}}{\left(\boldsymbol{R}\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}\right)' cov(\boldsymbol{R}',\boldsymbol{R})\boldsymbol{R}\boldsymbol{V}^{-1}\boldsymbol{\beta}[\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}]^{-1}} = \boldsymbol{\beta}. \quad \text{(A6)}$$

Thus,

$$E(\boldsymbol{R}) = \boldsymbol{\beta} E\left(\boldsymbol{w}'\boldsymbol{R}\right). \quad \text{(A7)}$$

Eq. (A7) 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, the factor's maximum correlation portfolio is not on the mean-variance frontier. Eq. (4) in the main text and Eq. (A4) in this appendix 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 when $\boldsymbol{\Sigma} = \boldsymbol{V}^{-1}$.

If the weighting matrix is not equal to the inverse of the covariance matrix of asset returns, then the maximum correlation portfolio and the least mispricing portfolio are not the same. When $\boldsymbol{\Sigma}$ is more distant from $\boldsymbol{V}^{-1}$,[1] the two types of FMPs can be different. We discuss three methods in Section 2. The non-GLS cross-sectional method ($\frac{\boldsymbol{\beta}'\boldsymbol{\Sigma} E(\boldsymbol{R})}{\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}}$ with the weighting matrix $\boldsymbol{\Sigma} \neq \boldsymbol{V}^{-1}$) and the sorting-by-beta method are implied by the least mispriced portfolio theory, but not directly by the maximum correlation theory.

When the number of assets (N) is large, the different choice of the weighting matrix can imply the same expected return. More specifically, from Eq. (4), the expected portfolio return is

[1] The distance between two matrices is captured by the L2 norm of them. The L2 norm for the matrix $\boldsymbol{M} = (m_{i,j})$ is $\| \boldsymbol{M} \|_{L2} = (\sum_{i,j} m_{i,j}^2)^{1/2}$.

$E(\boldsymbol{R})'\ \boldsymbol{w}^{*} = \frac{\boldsymbol{\beta}'\ \boldsymbol{\Sigma}E(\boldsymbol{R})}{\boldsymbol{\beta}'\ \boldsymbol{\Sigma\beta}}$. If the number of assets (N) converges to infinity, for any invertible $\boldsymbol{\Sigma}$, the law of large numbers implies that $\frac{\boldsymbol{\beta}'\ \boldsymbol{\Sigma}E(\boldsymbol{R})}{\boldsymbol{\beta}'\ \boldsymbol{\Sigma\beta}}$ converges to the same number. The covariance of asset returns can be difficult to estimate when N is large; thus, a different choice of the weighting matrix is preferred. In this sense, the least mispriced portfolio theoretically provides a more flexible choice of weighting, while the maximum correlation portfolio, in theory, relies on the covariance matrix of asset returns.

A finite N is economically meaningful in considering a flexible weighting matrix. For example, if the weighing matrix is the inverse of the covariance matrix of idiosyncratic returns, Eq. (3) in the main text minimizes the information ratio of the underlying factor. Suppose the weighing matrix is the inverse of a diagonal matrix, with the diagonal element being the variance of the idiosyncratic returns. In that case, we minimize the summation of the mispricing relative to idiosyncratic risk.

Appendix B. Correlation between the FMP constructed by the four-procedure method and the underlying factor

This section shows that the correlation between FMPs constructed by the four-procedure method and the underlying factor is less than the true correlation.

Applying Eq. (19) in the main text, we can write the variance of the FMP return as

$$\left(\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}{}' \, \boldsymbol{V}^{-\mathbf{1}}\widehat{\boldsymbol{B}}_{\boldsymbol{EV}}\right)^{-\mathbf{1}} \widehat{\boldsymbol{B}}_{\boldsymbol{IV}}{}' \, \boldsymbol{V}^{-\mathbf{1}}\boldsymbol{var}(\boldsymbol{r_t})\boldsymbol{V}^{-\mathbf{1}}\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}(\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}{}' \, \boldsymbol{V}^{-\mathbf{1}}\widehat{\boldsymbol{B}}_{\boldsymbol{EV}})^{-\mathbf{1}}.$$

When the error contains no factor structure (Section 3.1), Jegadeesh et al. (2019) show that $\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}{}' \, \boldsymbol{V}^{-\mathbf{1}}\widehat{\boldsymbol{B}}_{\boldsymbol{EV}}$ can converge to $\boldsymbol{B}' \, \boldsymbol{V}^{-\mathbf{1}}\boldsymbol{B}$ at a speed of $\sqrt{NT}$. When the error contains a factor structure (Section 3.2 and later), Jegadeesh and Noh (2014) show that $\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}{}' \, \boldsymbol{V}^{-\mathbf{1}}\widehat{\boldsymbol{B}}_{\boldsymbol{EV}}$ can converge to $\boldsymbol{B}' \, \boldsymbol{V}^{-\mathbf{1}}\boldsymbol{B}$ at a speed of e^T, a rate that is much fast than the classical ordinary least squares (OLS) method (converge at the speed of $\sqrt{T}$). Thus, as T large,

$$(\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}{}' \, \boldsymbol{V}^{-\mathbf{1}}\widehat{\boldsymbol{B}}_{\boldsymbol{EV}})^{-\mathbf{1}}\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}{}' \, \boldsymbol{V}^{-\mathbf{1}}\boldsymbol{var}(\boldsymbol{r_t})\boldsymbol{V}^{-\mathbf{1}}\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}\left(\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}{}' \, \boldsymbol{V}^{-\mathbf{1}}\widehat{\boldsymbol{B}}_{\boldsymbol{EV}}\right)^{-\mathbf{1}} \rightarrow$$

$$\left(\boldsymbol{B}' \, \boldsymbol{V}^{-\mathbf{1}}\boldsymbol{B}\right)^{-\mathbf{1}}\left(\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}{}' \, \boldsymbol{V}^{-\mathbf{1}}\widehat{\boldsymbol{B}}_{\boldsymbol{IV}}\right)\left(\boldsymbol{B}' \, \boldsymbol{V}^{-\mathbf{1}}\boldsymbol{B}\right)^{-\mathbf{1}} > \left(\boldsymbol{B}' \, \boldsymbol{V}^{-\mathbf{1}}\boldsymbol{B}\right)^{-\mathbf{1}}\left(\boldsymbol{B}' \, \boldsymbol{V}^{-\mathbf{1}}\boldsymbol{B}\right)\left(\boldsymbol{B}' \, \boldsymbol{V}^{-\mathbf{1}}\boldsymbol{B}\right)^{-\mathbf{1}} = \left(\boldsymbol{B}' \, \boldsymbol{V}^{-\mathbf{1}}\boldsymbol{B}\right)^{-\mathbf{1}}.$$

Note that the above equation is true since $\boldsymbol{var}(\boldsymbol{r_t}) = \boldsymbol{V}$. From the inequality above, we show that the variance of the FMP return is larger than the true value. Hence, the correlation between the FMPs and the underlying factor is less than the true value, given Eq. (A1).

Appendix C. Correlations between underlying factors and FMPs

Table C1. Correlations between underlying factors and FMPs.

This table shows the correlations between FMPs and the underlying factors using our four-procedure (FP) method. In panel A, we show correlations between first-stage FMPs and underlying factors. We apply our proposed method that encompasses single-factor model, the instrumental variables approach, and stock selection, to measure first-stage FMPs. In panel B, we show correlations between second-stage FMPs and underlying factors. We take first-stage FMPs as factors and reapply our proposed method to obtain second-stage FMPs. 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 sample period is January 1964 to March 2016. We use individual stocks that have at least 60 continuous months of returns on CRSP.

Panel A: Correlation between first-stage FMPs and underlying factors

	CG	CPI	IP	UE	MKT	SMB	HML
FP	0.411	0.485	0.663	0.565	0.741	0.804	0.710

Panel B: Correlation between second-stage FMPs and underlying factors

	CG	CPI	IP	UE	MKT	SMB	HML
FP	0.188	0.200	0.163	0.209	0.636	0.728	0.487

Appendix D. Additional simulation results

Table D1. Impact of measurement error on the relation between FMPs and their true factors.

As a complement to Table 2, this table shows additional simulation results for the effectiveness of the 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 the 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 true 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 last column of each table is the average value of the values of the four factors. In each table, we report two cases in which the correlation between the true and observed factors, $corr(\tilde{f}, f)$, is 30% and 10%, respectively. The values in each table are the mean values across 1,000 simulations. The sample period is 630 months. To be included in the simulations, individual stocks must have at least 60 continuous months of 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 true factors

	$corr(\tilde{f}, f)$ =30%				
	CG	CPI	IP	UE	Average
FMP_FP	0.988	0.985	0.852	0.836	0.915
FMP_OLS	0.752	0.826	0.729	0.902	0.803
FMP_Stein	0.752	0.826	0.729	0.902	0.803
FMP_LM	0.778	0.844	0.688	0.854	0.791
FMP_SB	0.997	0.707	0.139	0.106	0.487
FMP_TS	0.058	0.056	0.040	0.034	0.047
	$corr(\tilde{f}, f)$ =10%				
	CG	CPI	IP	UE	Average
FMP_FP	0.914	0.888	0.581	0.548	0.733
FMP_OLS	0.787	0.776	0.326	0.423	0.578
FMP_Stein	0.787	0.776	0.326	0.423	0.578
FMP_LM	0.795	0.769	0.299	0.356	0.555
FMP_SB	0.958	0.654	0.103	0.105	0.455
FMP_TS	0.018	0.019	0.013	0.009	0.015

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

	$corr(\tilde{f}, f)$ =30%				
	CG	CPI	IP	UE	Average
FMP_FP	0.004	0.007	0.040	0.008	0.015
FMP_OLS	0.298	0.448	0.384	0.157	0.322
FMP_Stein	0.298	0.448	0.384	0.157	0.322
FMP_LM	0.279	0.400	0.470	0.216	0.341
FMP_SB	0.024	0.677	0.879	0.900	0.620
FMP_TS	0.002	0.022	0.044	0.030	0.025
	$corr(\tilde{f}, f)$ =10%				
	CG	CPI	IP	UE	Average
FMP_FP	0.008	0.004	0.020	0.014	0.012
FMP_OLS	0.078	0.281	0.622	0.388	0.342
FMP_Stein	0.078	0.281	0.622	0.388	0.342
FMP_LM	0.066	0.288	0.628	0.412	0.348
FMP_SB	0.019	0.499	0.641	0.687	0.462
FMP_TS	0.002	0.007	0.016	0.009	0.009

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

	$corr(\tilde{f}, f)$ =30%				
	CG	CPI	IP	UE	Average
FMP_FP	0.002	0.002	0.008	0.004	0.004
FMP_OLS	0.203	0.135	0.168	0.105	0.153
FMP_Stein	0.203	0.135	0.168	0.105	0.153
FMP_LM	0.190	0.134	0.158	0.144	0.156
FMP_SB	0.012	0.123	0.168	0.229	0.133
FMP_TS	0.001	0.004	0.010	0.013	0.007
	$corr(\tilde{f}, f)$ =10%				
	CG	CPI	IP	UE	Average
FMP_FP	0.002	0.001	0.007	0.003	0.004
FMP_OLS	0.038	0.071	0.121	0.158	0.097
FMP_Stein	0.038	0.071	0.121	0.158	0.097
FMP_LM	0.027	0.074	0.135	0.163	0.100
FMP_SB	0.010	0.091	0.127	0.187	0.104
FMP_TS	0.001	0.002	0.005	0.005	0.003

Table D2. Impact of measurement error on the inference of the risk premium under misspecification.

This table shows the statistical inference of the risk premiums of FMPs estimated from our proposed four-procedure method under the case of misspecification. We follow a similar procedure as Table 3. The only difference is that we capture misspecification. We intentionally remove the FMPs of unemployment rate and estimate the risk premiums of the FMPs of the other three factors. Panel A reports the mean value of the estimated risk premiums of the FMPs. Panel B reports the critical values of the *t*-statistics at different percentiles. We report three cases in which the correlation between the true and observed factor, $corr(\tilde{f}, f)$, is 30%, 20%, and 10%, respectively. The values in each table are the mean values across 1,000 simulations. The last column of Panel B is the simple average of the values of the three factors. The sample period is 630 months. To be included in the simulations, individual stocks must have at least 60 continuous months of returns on CRSP. The macro factors include unexpected consumption growth (CG), unexpected changes in the CPI (CPI), and unexpected changes in industrial production (IP).

Panel A: Bias in the risk premium estimation

$corr(\tilde{f}, f)$	CG	CPI	IP
30%	-0.001	0.004	0.000
20%	0.001	-0.002	0.001
10%	0.001	-0.002	-0.002

Panel B: Critical value of the risk premium estimation

		$corr(\tilde{f}, f)$ = 30%		
Percentile	CG	CPI	IP	Average
2.5%	-1.7994	-1.9843	-1.9526	-1.912
5%	-1.524	-1.7066	-1.5906	-1.607
10%	-1.2302	-1.2454	-1.2451	-1.240
50%	0.0846	0.0009	0.0392	0.042
90%	1.3703	1.257	1.3339	1.320
95%	1.6757	1.6375	1.6845	1.666
97.5%	2.0567	1.9992	2.0646	2.040
		$corr(\tilde{f}, f)$ = 20%		
Percentile	CG	CPI	IP	Average
2.5%	-2.045	-1.920	-2.020	-1.995
5%	-1.662	-1.688	-1.739	-1.696
10%	-1.355	-1.302	-1.414	-1.357
50%	0.002	0.025	-0.070	-0.014
90%	1.231	1.277	1.194	1.234
95%	1.629	1.612	1.476	1.572
97.5%	1.922	1.868	1.858	1.883

	$corr(\tilde{f}, f) = 10\%$			
Percentile	CG	CPI	IP	Average
2.5%	-1.531	-1.459	-1.773	-1.587
5%	-1.007	-0.999	-1.280	-1.095
10%	-0.615	-0.662	-0.884	-0.720
50%	0.001	0.001	-0.016	-0.005
90%	0.521	0.678	0.902	0.700
95%	0.854	1.014	1.206	1.024
97.5%	1.377	1.695	1.961	1.678

Appendix E. FMPs constructed by the 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 (FMP_TS) and following Lamont (2001) in our main text, we examine three more methods. Giglio and Xiu (2021) create FMPs using the time-series approach with principal components (PCs) of 202 portfolio returns as the basis assets. We use the first 10 PCs and construct time-series FMPs using the fitted value from the regression. FMP_GX10PC denotes these FMPs. Next, we use all 202 portfolios from Giglio and Xiu (2021) (these portfolios capture most of the cross section anomalies) as the basis assets to construct the FMPs (denoted by FMP_GX202). However, the time-series regressions suffer from the curse of the dimensionality and the overfitting problem when the number of regressors is large. We mitigate these problems by applying the time-series approach with a variable selection method to the large portfolio set. Specifically, we adopt the Lasso method to solve the overfitting problem and select a small number of basis assets from the 202 portfolios, which is the third method, denoted by "FMP_GX202_LASSO."

Table E1 summarizes the empirical results of the FMPs constructed by the three methods described above. Panel B of Table E1 shows the risk premiums of these FMPs. Of the three methods, FMP_GX202_LASSO dominates the other two in terms of the risk premium's magnitude and significance. Consumption growth and industrial production have significant risk premiums and are robust when controlling for the FMPs of the three Fama-French factors. It is noteworthy that the consumption growth constructed by FMP_GX202_LASSO does not pass the first criteria. Moreover, even though IP is significant, it has the opposite sign of what is predicted intuitively. The FMP_GX202 method leads to an insignificant risk premium. 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 E1. 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 construct the FMPs. FMP_GX10PC includes the FMPs that are constructed by the time-series approach using 10 principal components [from 202 Giglio and Xiu (2021) portfolios] as predictors. FMP_GX202 consists of the FMPs constructed by the time-series approach with the 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 the 202 portfolios. Panel A shows the correlations between the FMPs and the systematic risk factors extracted from the individual stock returns (the second criteria). *t*-statistics breaching the 5% (1%) critical level are in boldface. The factors that pass the necessary conditions are highlighted in gray. Panel B presents the estimated risk premiums (the third criteria).

Panel A. 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 B. 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 F. Proofs of the propositions

Proof of Proposition 1

The objective function can be written as

$$\left(E(\boldsymbol{R}) - \boldsymbol{\beta} E(\boldsymbol{w}'\boldsymbol{R})\right)' \boldsymbol{\Sigma} \left(E(\boldsymbol{R}) - \boldsymbol{\beta} E(\boldsymbol{w}'\boldsymbol{R})\right).$$

Taking the first-order condition with respect to $\boldsymbol{w}$, we obtain

$$0 = -2\boldsymbol{\beta}'\boldsymbol{\Sigma} E(\boldsymbol{R})E(\boldsymbol{R}) + 2\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta} E(\boldsymbol{R})E(\boldsymbol{R})'\boldsymbol{w}.$$

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

Proof of Proposition 2

From Eq. (10) (the first-pass regression) and using the assumption in the proposition, the estimated coefficient is

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

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

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

Therefore, the estimated coefficient in Eq. (18) becomes

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

Proof of Proposition 3

1. In the first-pass regression, when T converges to infinity, we have

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

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,

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

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

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

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

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

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

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

(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

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

. 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 $\boldsymbol{R^i} = \alpha^i + \beta_1^{i*}\boldsymbol{f_1^*} + \boldsymbol{\varepsilon}^i$, even if the true model follows regression $\boldsymbol{R_t}' = \alpha_t + \lambda_{1t}\widehat{\boldsymbol{\beta}}_1^{*\prime} + \lambda_{2t}\widehat{\boldsymbol{\beta}}_{\boldsymbol{2}}^{*\prime} + \boldsymbol{\nu_t}$. 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 (17), 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

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 $\boldsymbol{R^i} = \alpha^i + \beta_1^{i*}\boldsymbol{f_1^*} + \boldsymbol{\varepsilon}^i$,
for any asset i,

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

Recall that $\tilde{f}_{1t} = f_{1t} + \varepsilon_{f_{1,t}}$. The goal is to prove that any term related to the 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 the 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 the numerator and denominator, we obtain

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

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 ($O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, i\right)$) and is of order $O(\frac{1}{\sqrt{T}})$.

In the second-pass regression (17), 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 the denominator, we obtain

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

where

$$O(\frac{1}{\sqrt{T}},\varepsilon_{f_1},i)=\frac{\beta_1^{i*}\left(\frac{1}{T}\sum_{t=1}^{T}((f_{1t}-E(f_1))(\varepsilon_{f_1t}))\right)}{var(f_1)+var(\varepsilon_{f_1})}+\frac{\beta_1^{i*}var(f_1)}{(var(f_1)+var(\varepsilon_{f_1}))^2}(\left(\frac{1}{T}\sum_{t=1}^{T}\left((\varepsilon_{f_1t})^2\right)-var(\varepsilon_{f_1})\right)+2\frac{1}{T}\sum_{t=1}^{T}((f_{1t}-E(f_1))(\varepsilon_{f_1t}))).$$

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

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,

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

---

[2] 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 $O^1(\frac{1}{\sqrt{T}},\varepsilon_{f_1},i)=\frac{\left(\frac{1}{T}\sum_{t=1}^{T}((f_{1t}-E(f_1))(\varepsilon_{f_1t}))\right)}{var(f_1)+var(\varepsilon_{f_1})}+\frac{var(f_1)}{(var(f_1)+var(\varepsilon_{f_1}))^2}(\left(\frac{1}{T}\sum_{t=1}^{T}\left((\varepsilon_{f_1t})^2\right)-var(\varepsilon_{f_1})\right)+2\frac{1}{T}\sum_{t=1}^{T}((f_{1t}-E(f_1))(\varepsilon_{f_1t})))$. The proof will be the same since $O(\frac{1}{\sqrt{T}},\varepsilon_{f_1},i)$ is $O\left(\frac{1}{\sqrt{T}},\varepsilon_{f_1}\right)$.

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

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 $O\left(\frac{1}{\sqrt{T}}, \varepsilon_{f_1}, t\right)$ over time, so that

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

The first two terms are both of order $O(\frac{1}{\sqrt{T}})$ and depend on $\varepsilon_{f_1 t}$ and 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, 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 is

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

Given that the measurement errors, $\varepsilon_{f_1,t}$, are uncorrelated with the 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 come from the measurement error of order $O(\frac{1}{T})$, and $O\left(\frac{1}{T},Other\right)$ is 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, we can write the estimated error that comes from the measurement error, ε_{f_1}, as order $O(\frac{1}{T})$.

In the second-pass regression (17), 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 is

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

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

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

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 $\frac{1}{\frac{1}{c^2}\overline{var}(\beta_1^*)}(\frac{1}{N}\left(\sum_{i=1}^{N}\left(O\left(\frac{1}{T},\varepsilon_{f_1},i\right)-\frac{1}{N}\sum_{j=1}^{N}O\left(\frac{1}{T},\varepsilon_{f_1},j\right)\right)\left(R_t^i-\right.\right.$
$\left.\left.\frac{1}{N}\sum_{j=1}^{N}R_t^j\right)\right))+\frac{\frac{1}{c}\overline{var}(\beta_1^*)(f_{1t}-E(f_1)+\gamma_1)}{\left(\frac{1}{c^2}\overline{var}(\beta_1^*)\right)^2}\frac{1}{N}\left(\sum_{i=1}^{N}\left(\beta_1^{i*}-\frac{1}{N}\sum_{j=1}^{N}\beta_1^{j*}\right)\left(O\left(\frac{1}{T},\varepsilon_{f_1},i\right)-\right.\right.$
$\left.\left.\frac{1}{N}\sum_{j=1}^{N}O\left(\frac{1}{T},\varepsilon_{f_1},j\right)\right)\right)$. Take the average of $O\left(\frac{1}{T},\varepsilon_{f_1},t\right)$ over time, so that

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

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

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

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

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

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

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

Proof of Proposition 7

From regression (21) and the regularity assumptions, when *T* converges to infinity, we get

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

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

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

Appendix G. Robustness tests

G.1. Using Giglio and Xiu's (2021) three-pass method to resolve model misspecification

To resolve the omitted variable problem from model misspecification in the risk premium estimation, Giglio and Xiu (2021) provide a three-pass framework. Their method can be described as follows. First, extract principal components from the covariance matrix of 202 portfolio returns. Second, run a Fama-MacBeth regression of cross-sectional individual stock returns on these principle components to estimate their risk premiums (denoted as $\boldsymbol{\gamma}_p$). Third, regress each factor on these principle components to get the estimated coefficients (denoted by $\boldsymbol{\delta}_p$). The risk premium of the factor is computed as $\boldsymbol{\gamma}_p{}'\boldsymbol{\delta}_p$.

We apply their methodology to the four-procedure FMPs of the four macro factors. The result shows that the monthly premium of the FMP_FP of CG is 0.09 percentage points and is highly significant as the *R*-squared is high, indicating that consumption growth survives from the omitted variable problems.

Table G1. Risk premium estimation using Giglio and Xiu's (2021) method.

This table shows the risk premium estimation using Giglio and Xiu's (2021) three-pass method. 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). The first row shows the estimated risk premium for the four-step FMP of each underlying factor. The second row shows the corresponding *R*-squared in the third pass, which captures explanatory power of the four selected principal components on the target FMP.

	CG	CPI	IP	UE
Risk premium	0.090**	-0.016	-0.013	0.010
t-ratio	(1.96)	(1.14)	(0.929)	(-1.667)
R^2	0.621	0.207	0.0347	0.169