The Cross-Sectional Intercept in Asset Pricing Tests

Alex Horenstein Paulo Maio*

August 2026

Abstract

Linear factor models for excess returns imply a zero cross-sectional regression (CSR) intercept (λ_0), yet one is routinely estimated. A free intercept absorbs the no-intercept CSR's average pricing error, mechanically amplified, especially when betas vary little relative to their size. It credits pricing failures as explained risk premia, inflates fit, distorts risk prices (often reversing sign), and weakens risk price identification. We derive restricted-unrestricted CSR mappings and penalized performance measures treating the intercept as pricing error. In our results, penalized fit is essentially always negative, and unrestricted market risk prices are never significantly positive. We recommend imposing $\lambda_0 = 0$ when evaluating models.

Keywords: asset pricing, cross-sectional regressions, two-step regression approach, linear factor models, pricing errors

*Horenstein: a.horenstein@miami.edu, Department of Economics, University of Miami; Maio: paulo.maio@hanken.fi, Department of Finance and Economics, Hanken School of Economics. We thank seminar participants at Miami Herbert Business School and the Econometric Society North American Summer Meeting (Atlanta) for helpful comments. We are grateful to Kenneth French, Bryan Kelly, Tyler Muir, and Lu Zhang for providing data on their web pages.

1 Introduction

The two-step regression approach of Black, Jensen, and Scholes (1972) and Cochrane (2005) has been the norm when estimating linear asset pricing models in which at least one of the factors is not traded (i.e., it does not represent an excess return).[1] Typically, such a method is applied to a cross-section of stock returns in excess of the risk-free rate (i.e., risk premia), which implies (theoretically) that the intercept should be excluded from the second-step cross-sectional regression (CSR). However, the vast majority of the studies in the literature include the cross-sectional intercept and typically do not report the results for the restricted CSR that excludes the intercept; when reported, these results are usually presented only briefly as ancillary robustness evidence rather than as a primary object of analysis. A common justification for including the intercept is that the observed risk-free rate (typically defined as the T-bill return) is not a good proxy for the "true" zero-beta rate due to frictions in the money market such as a liquidity premium or the discrepancy between the lending and borrowing interest rates (e.g., Brennan (1971); Black (1972); Burnside (2011); Frazzini and Pedersen (2014)). Yet, quite often the magnitudes of the intercept estimates are too large to be consistent with such an explanation. For example, Lewellen, Nagel, and Shanken (2010) argue that "(riskless) borrowing and lending rates are not sufficiently different, perhaps 1-2% annually, to justify the extremely high zero-beta estimates in many papers." Despite this clear warning, most studies in recent years continue to ignore this basic economic restriction on the cross-sectional intercept.[2] These include papers that propose better empirical methods to estimate and evaluate asset pricing models.[3] Frequently, researchers downplay the large magnitude of the intercept when it is not statistically significant, a practice that ignores the importance of addressing its economic significance. Importantly, virtually all studies in the literature include the intercept in the fitted risk premia explained by the model rather than as part of the pricing errors.[4] The intercept therefore plays a dual role: it relaxes a theoretical restriction, but it also becomes

[1] A related method is the procedure proposed by Fama and MacBeth (1973). The two methods are equivalent (i.e., produce the same risk price estimates) when the factor loadings are estimated from the full sample.

[2] Among the exceptions are Brennan, Wang, and Xia (2004), Campbell and Vuolteenaho (2004), Cochrane (2005), Burnside (2011), Maio and Santa-Clara (2017), Campbell et al. (2018), and Cooper, Ma, and Maio (2022). In these studies, the zero-intercept restriction is imposed ex ante when applying the two-step (or the Fama–McBeth) approach.

[3] Giglio and Xiu (2021) and Giglio, Xiu, and Zhang (2025) represent two notable exceptions, as the intercept is omitted.

[4] Maio (2024) and Maio and Min (2026) represent two recent exceptions.

part of what the model is credited with explaining.

Motivated by this dual role, we undertake a systematic theoretical and empirical analysis of the consequences of leaving the cross-sectional intercept unrestricted. We show that the intercept-versus-no-intercept choice materially affects model evaluation, risk price estimates, and identification. To evaluate this choice, the paper provides a structural decomposition of the intercept, an exact mapping between restricted and unrestricted estimates, and a collinearity index that signals when the unrestricted CSR becomes unreliable. Within a general linear asset pricing framework, the intercept in an excess-return CSR represents a common pricing offset relative to the tested factor span. One credible source of such an offset is omitted priced components. In this case the intercept reflects the risk-price-weighted average of the omitted loading component that is not spanned by the included factor loadings. This result provides a structural foundation for treating the intercept as a pricing error component rather than as risk premia explained by the tested model. This in turn implies new composite performance metrics. We further show that the intercept equals the average pricing error from the restricted CSR divided by $1 - s$, where s is a collinearity index linked to the first two cross-sectional moments of the included factor loadings. A high value of s means that the constant vector is nearly spanned by the included factor loadings, or equivalently that some linear combination of betas has a large cross-sectional mean relative to its dispersion. Thus, even modest average pricing errors can generate large intercepts when s is close to one. We refer to this mechanism as the *collinearity amplification channel*.

We evaluate these theoretical implications using a battery of empirical cross-sectional estimations based on both quarterly and monthly data. At the quarterly frequency, the baseline CAPM and 12 macro-factor models are estimated on decile portfolios sorted by classical cross-sectional anomalies: investment, profitability, momentum, and book-to-market. Including both single-anomaly and joint 40- and 17-portfolio (which contains the extreme four deciles within each group plus the market factor) cross-sections yields 78 quarterly estimations. At the monthly frequency, eight macro models, in addition to the CAPM and the 6-factor model of Fama and French (2018) (FF6), are estimated over the two cross-sections associated with all portfolio groups, yielding a total of 20 estimations.

The empirical results illustrate the high quantitative importance of the intercept mechanism. When we treat the intercept as part of the total pricing error, the fit is negative in 76 of the 78

quarterly estimations, which means a very poor performance: the average squared pricing error (which includes the intercept) exceeds the cross-sectional variation in the raw risk premia that we seek to explain in the first place. These results contrast sharply with conventional assessments: R^2_{ols}, which is constructed exclusively from the CSR residuals, is significantly positive (based on a bootstrap inference) for 43 of these 78 estimations. This large discrepancy arises because estimated intercepts are economically large, often exceeding 2% per quarter, partially because many factor loadings (or their linear combination) exhibit precisely the pattern that amplifies non-zero average pricing errors (from the restricted CSR)—high cross-sectional means (in magnitude) and low cross-sectional dispersion. A complementary pricing-error metric (denoted by $RRMSE_{\text{pen}}$) yields a similar qualitative conclusion: when the intercept is included in the definition of pricing errors, the total composite pricing error exceeds the raw composite risk premium that we seek to explain (i.e., $RRMSE_{\text{pen}}$ above 100%) in 54 of the 78 quarterly estimations. In contrast, using the conventional pricing errors (i.e., exclusively the CSR residuals), there are only three cases among the 78 estimations (i.e., 3.8%), in which the relative composite pricing error ($RRMSE$) is above 30%. At the monthly frequency, the results are perhaps even clearer. Notably, R^2_{pen} is largely negative in all 20 estimations, compared to positive (and statistically significant) R^2_{ols} in nearly all cases. On the other hand, $RRMSE_{\text{pen}}$ is above 100% in all 20 monthly estimations (and actually above 200% in 15 of these estimations), compared to $RRMSE$ below 30% for 18 of the 20 cases. Again, this extreme decoupling between the two metrics stems from the very large intercept estimates, which are above 1% per month (or more than 12% per year) in 16 of the 20 estimations.

Importantly, the unrestricted CSR, when appropriately treating the intercept as the average pricing error, often leads to similar qualitative empirical evidence as the corresponding restricted CSR, especially in terms of rejecting weak models. However, it is far from a perfect substitute for the restricted counterpart when it comes to correctly assessing linear factor models. FF6 provides a useful example in this respect. The fit in the restricted CSR is as large as 90%, with significance at the 1% level, in the estimation with the 17 portfolios. However, the corresponding R^2_{pen} is very negative, while $RRMSE_{\text{pen}}$ is close to 300%, compared to a very small $RRMSE$ in the restricted CSR, 14%.[5] The explanation for this very large discrepancy hinges on the extreme value of s,

[5]Since the FF6 factors are traded and the 17 portfolios are constructed to load strongly on these factors, the time-series approach provides the natural benchmark for assessing the model; the CSR evidence discussed here should therefore be interpreted as illustrating the s-mechanism and not as evidence against the model.

0.9996, which leads to a large intercept, around 1.73% per month, even though the average pricing error in the restricted CSR is very close to zero, 0.0007% per month. Thus, a very large s can generate "overshooting": even good models may be rejected when the unrestricted CSR intercept is appropriately treated as the average pricing error.

The distortion in fit measures (driven by the intercept) has a direct counterpart in risk price estimation. Thus, as another theoretical contribution, we derive the exact mapping between restricted and unrestricted risk price estimates. The key insight is that near-collinearity affects risk prices through two linked components. First, as s approaches one, near-collinearity amplifies the intercept relative to the average pricing error from the restricted CSR. Second, the resulting common pricing offset is transmitted into the risk price estimates through a vector $\boldsymbol{u}$, built from the same first and second cross-sectional beta moments that also determine s. The two objects are therefore distinct but linked components of the same mapping: s governs intercept amplification, while u governs how the inflated intercept is absorbed by the estimated risk prices. This mechanism is the asset-pricing manifestation of the familiar ill-conditioning of OLS under near-collinearity. The theoretical prediction is especially transparent in the single-factor case, such as the CAPM: holding the other components fixed, the magnitude of the risk price gap is larger when the restricted CSR has a larger average pricing error in magnitude, when the factor betas have smaller cross-sectional dispersion, and when the magnitude of their cross-sectional average is larger. In the single-factor case, the combined effect of s and $\boldsymbol{u}$ is especially simple: small beta dispersion magnifies the risk price gap, while a larger magnitude of the average beta scales the resulting distortion. Consistent with the theoretical mapping, including an intercept frequently flips the sign of estimated risk prices or substantially reduces their magnitudes. The market factor illustrates the mechanism sharply: across the nine quarterly macro-factor models that include a market factor (and considering the two more relevant cross-sections associated with all four portfolio groups), every market risk price estimate turns negative in unrestricted CSRs, whereas restricted CSRs yield uniformly positive and statistically significant estimates. A similar pattern holds for the eight monthly models that include the market factor (and for both cross-sections, 16 estimations total).

We also formalize (theoretically) the identification cost of including an intercept. Adding a constant strips out the mean component of the factor loadings, so risk prices are identified only from cross-sectional dispersion in betas around their means. This weakens the rank condition

whenever some linear combination of betas is close to being constant across assets. In settings where betas display little cross-sectional dispersion, unrestricted-intercept estimates are therefore especially prone to weak identification. Applying the identification-robust rank tests of Kleibergen and Zhan (2020), we confirm that identification is weaker when the intercept is included, indicating that unrestricted risk price estimates are less precisely identified.

Furthermore, the economic content of the intercept can also be assessed through the zero-beta rate and equity premium it implies under a literal benchmark-wedge interpretation. The mapping is mechanical: the implied zero-beta rate equals the observed risk-free rate (used to form excess returns) plus the intercept, and the implied equity premium equals the sample average market excess return minus the intercept. In our data, most intercept estimates therefore imply economically large benchmark-rate wedges and negative implied aggregate equity premia. Specifically, all 20 monthly intercept estimates imply a negative market equity premium, something that reveals a fundamental economic violation of any asset pricing model: taken at face value, these estimates suggest that investors pay to bear systematic risk rather than demand compensation for it.

Monte Carlo simulations reinforce our interpretation of a free intercept as a sign of misspecification: a misspecified model generates an average intercept whose magnitude is much larger than under correct specification, consistent with our theoretical result that large cross-sectional intercepts arise from omitted priced risk factors.

Finally, we extend the analysis to a GLS second pass. The same logic carries over once the equal cross-sectional weighting of the OLS second pass is replaced by inverse first-pass residual-covariance weighting. A free GLS intercept absorbs the GLS-weighted mean pricing error from the restricted CSR, amplified by $(1 - s_{\text{gls}})^{-1}$, and the penalized-fit decomposition carries over, with the GLS collinearity index s_{gls} scaling the restricted-versus-penalized fit gap. Empirically, GLS does not overturn the main conclusions from OLS: intercept estimates remain economically large and statistically significant in all cases, whereas the market risk price again reverses sign (relative to the restricted regression). Furthermore, the penalized GLS fit is negative in nearly every reported specification, even when the conventional GLS fit metric is significantly positive.

In sum, if the intercept is included, fit measures should not credit it as explained variation; instead, one should report penalized pricing-error metrics that treat the intercept as part of total pricing error. The two measures address distinct questions. The restricted CSR and the related

metric $R^2_{\text{ols}*}$ assess total pricing error under the zero-intercept restriction, whereas R^2_{pen} (which imposes the restriction that the intercept is the average pricing error) does so for the unrestricted CSR, in both cases scaled by cross-sectional dispersion. Conflating these objects can lead to materially different conclusions about model performance. Even this correction is only a partial remedy: although one can compute pricing errors correctly under the unrestricted CSR, the associated risk price estimates remain highly sensitive to collinearity amplification (often at odds with the underlying theoretical model) and, as the rank tests show, are often weakly identified. Therefore, the main message from this paper is clear: when estimating and evaluating linear macro factor models over cross-sectional risk premia (which represents the standard practice in the literature), researchers should rely on the restricted CSR (and associated performance metrics) as the benchmark empirical setting. The results from the unrestricted CSR (where the intercept is adequately treated as the average pricing error) should be only complementary and still interpreted with considerable caution.[6]

Our analysis contributes to three strands of the empirical asset pricing literature. First, it builds on work on two-pass model evaluation and cross-sectional fit. Lewellen, Nagel, and Shanken (2010) caution that conventional asset-pricing tests can deliver misleadingly strong evidence in favor of proposed models and recommend, among other prescriptions, taking the magnitudes of cross-sectional intercept estimates seriously when theory provides guidance. We implement that prescription by undertaking a comprehensive theoretical and empirical analysis of how the unrestricted intercept affects pricing errors, model evaluation, risk price estimates, and identification. This analysis constitutes the paper's key contribution. Kleibergen and Zhan (2015) derive the large-sample distribution of the OLS R^2 of the second-pass CSR and show that it can be large when the observed factors are only weakly correlated with the factors generating the return covariance structure, in which case the fit reflects estimation error in the first-pass loadings. Their OLS measure is the R^2_{ols} we study, computed from the same unrestricted CSR in which the zero-beta return is estimated. We complement this work by identifying the unrestricted intercept as a mechanism that reallocates pricing error into fitted values and by proposing a penalized fit measure that

[6] In particular, this requires reporting the corresponding restricted average pricing error, the collinearity index s, and the wedges that drive the decoupling in fit and risk price estimates with respect to the counterparts from the restricted CSR. Reporting the results of the unrestricted CSR is also informative when the intercept estimate is close to zero: consistent with the identities in Section 2, the two types of CSR then deliver similar risk price estimates and fit measures.

treats the intercept as pricing error rather than explained variation. Second, the paper relates to work on omitted, useless, and weak factors, all of which concern the factor content of tested models. Kan and Zhang (1999) show that useless factors can appear priced in two-pass tests; Giglio and Xiu (2021) propose a three-pass procedure for estimating risk premia in the presence of omitted priced factors; Giglio, Xiu, and Zhang (2025) connect weak factors to the choice of test assets; and Dello-Preite et al. (2025) allow expected excess returns to include compensation for unsystematic risk. Our focus is different: even holding the tested factor span fixed, freeing the intercept changes fit, risk prices, and identification. Third, our identification results relate to work on multicollinear, near-invariant, and rank-deficient beta matrices, including Ahn, Perez, and Gadarowski (2013), Bryzgalova (2015), Ahn, Horenstein, and Wang (2018), and Kleibergen and Zhan (2020). We show that the intercept creates a distinct identification cost by removing the mean component of factor loadings and relying only on their cross-sectional dispersion.

The rest of the paper proceeds as follows. Section 2 develops the theoretical framework. Section 3 describes the data. Section 4 examines the impact of the intercept on pricing errors and model fit. Section 5 analyzes effects on risk price estimates and identification. Section 6 connects the intercept to the zero-beta rate and equity premium. Section 7 presents simulations. Section 8 extends the analysis to GLS. Section 9 concludes.

2 Theory

This section provides the theoretical background. To save space and keep the focus, the proofs for all propositions, corollaries, and lemmas below are presented in the Internet Appendix (IA).

2.1 The Cross-Sectional Intercept and Missing Factors

In empirical cross-sectional (CS henceforth) asset pricing, a nonzero intercept in the second-stage CSR often signals model misspecification, concretely omitted risk factors. This subsection formalizes how missing factors map into the intercept and into biased risk price estimates in the model being tested. We show that the intercept equals the omitted factors' risk prices times the unspanned component of their average loadings. This provides a clear economic rationale for treating the intercept as a common pricing offset (a common level pricing error) rather than as a free parameter that contributes to explained fitted risk premia.

Throughout the entire section, all vectors are column vectors. Let $\mathbb{E}(R_{i,t})$ be the expected

excess return of asset $i = 1, \ldots, n$ (relative to the risk-free rate), with sample analogue given by $\mu_i = (1/T)\sum_{t=1}^{T} R_{i,t} > 0$.[7] Let $\overline{\mu} = \mathbb{E}_n(\mu_i) > 0$ denote the CS average risk premium among the test assets. The associated $n \times 1$ vector of sample risk premia on all assets is $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$. Let $\widehat{\boldsymbol{\beta}}_i = (\widehat{\beta}_{i,1}, \ldots, \widehat{\beta}_{i,K})'$ denote the first-stage betas (for asset i) estimated from time-series multivariate regressions, and let $\widehat{\boldsymbol{\beta}}$ be the corresponding $n \times K$ matrix of factor loadings. The $K \times 1$ vector of factor risk prices is $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)'$. We work with CS operators, $\mathbb{E}_n(\cdot)$, $\mathrm{Var}_n(\cdot)$, and $\mathrm{Cov}_n(\cdot, \cdot)$, taken over $i = 1, \ldots, n$. To simplify the notation, write $\boldsymbol{m} = \mathbb{E}_n\left(\widehat{\boldsymbol{\beta}}_i\right)$ as the $K \times 1$ vector of (cross-sectionally) average betas and $\boldsymbol{M} = \mathbb{E}_n\left(\widehat{\boldsymbol{\beta}}_i\widehat{\boldsymbol{\beta}}_i'\right)$ as the $K \times K$ matrix of betas' second moments. It follows that the covariance matrix of the factor loadings is given by $\boldsymbol{\Sigma} = \mathrm{Var}_n\left(\widehat{\boldsymbol{\beta}}_i\right) = \boldsymbol{M} - \boldsymbol{m}\boldsymbol{m}'$. A "hat" denotes a sample-based estimate or statistic, typically obtained by OLS (either in the time-series or in the cross-section).[8]

Consider the second-stage OLS CSR with an intercept,

$$\mu_i \;=\; \widehat{\lambda}_0 \;+\; \sum_{j=1}^{K} \widehat{\lambda}_j \widehat{\beta}_{i,j} \;+\; \widehat{\omega}_i \;=\; \widehat{\lambda}_0 \;+\; \widehat{\boldsymbol{\lambda}}'\widehat{\boldsymbol{\beta}}_i \;+\; \widehat{\omega}_i, \qquad i = 1, \ldots, n, \tag{1}$$

which is often used to test a K-factor model:

$$\mathbb{E}(R_{i,t}) \;=\; \sum_{j=1}^{K} \lambda_j \beta_{i,j} + \alpha_i.$$

We allow for model misspecification, i.e., $\alpha_i \neq 0$ for at least one test asset. In matrix form, the baseline CSR is defined as

$$\boldsymbol{\mu} = \widehat{\lambda}_0 \mathbf{1} + \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\lambda}} + \widehat{\boldsymbol{\omega}}, \tag{2}$$

where $\widehat{\boldsymbol{\omega}} = (\widehat{\omega}_1, \ldots, \widehat{\omega}_n)'$ is the $n \times 1$ vector of CSR residuals (pricing errors) and $\mathbf{1}$ denotes the $n \times 1$ vector of ones.

[7] The assumption of positive risk premia facilitates the interpretation of the theoretical results presented in this section, while being consistent with most empirical setups in CS asset pricing tests. Yet, such an assumption is not necessary for the derivations in this section.

[8] The CSRs are run on sample mean excess returns and estimated factor loadings. Accordingly, $\mu_i = (1/T)\sum_{t=1}^{T} R_{i,t}$ denotes the second-pass dependent variable, while $\mathbb{E}(R_{i,t})$ denotes its population counterpart. When we provide population interpretations, they refer to the analogues obtained by, for example, replacing $(\mu_i, \widehat{\boldsymbol{\beta}}_i, \widehat{\boldsymbol{\rho}}_i)$ with $(\mathbb{E}(R_{i,t}), \boldsymbol{\beta}_i, \boldsymbol{\rho}_i)$.

Assumption 1. Let $\boldsymbol{\rho} \in \mathbb{R}^{n\times L}$ collect additional loadings that, together with the K included loadings in $\boldsymbol{\beta}$, span expected excess returns:

$$\mathbb{E}(R_{i,t}) \;=\; \sum_{j=1}^{K} \theta_j \beta_{i,j} \;+\; \sum_{l=1}^{L} \gamma_l \rho_{i,l}.$$

The corresponding (augmented) second-pass CSR is

$$\mu_i \;=\; \widehat{\theta}_0 \;+\; \sum_{j=1}^{K} \widehat{\theta}_j \widehat{\beta}_{i,j} \;+\; \sum_{l=1}^{L} \widehat{\gamma}_l \widehat{\rho}_{i,l} \;+\; \widehat{\varpi}_i \;=\; \widehat{\theta}_0 + \widehat{\boldsymbol{\theta}}' \widehat{\boldsymbol{\beta}}_i + \widehat{\boldsymbol{\gamma}}' \widehat{\boldsymbol{\rho}}_i + \widehat{\varpi}_i, \tag{3}$$

where $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_K)'$ are the estimated risk prices on the included K factors. Although the loadings $\boldsymbol{\rho}$ are omitted from the baseline second-pass regression (2), it is convenient to define the augmented two-pass regression that would obtain if these additional loadings were also included. Accordingly, $\widehat{\boldsymbol{\rho}}$ and $\widehat{\boldsymbol{\gamma}}$ denote the additional OLS objects from this augmented (counterfactual) procedure; they are introduced solely as an analytical benchmark for the omitted-factor algebra and are not estimated in the baseline implementation. Hence, $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_L)'$ are the OLS coefficients on the additional factors in the augmented CSR, and $\widehat{\boldsymbol{\rho}}_i$ denotes the corresponding first-pass loading estimates for asset i (the transpose of the ith row of $\widehat{\boldsymbol{\rho}}$) in that augmented specification.

If the population analogue of the augmented pricing relation is correctly specified (consistent with a linear SDF applied to excess returns), then $\theta_0 = 0$ in population and there are no structural pricing errors; in finite samples, nonzero $(\widehat{\theta}_0, \widehat{\boldsymbol{\varpi}})$ can still arise due to sampling error and first-pass estimation error. In matrix form, the augmented CSR is given by

$$\boldsymbol{\mu} = \widehat{\theta}_0 \mathbf{1} + \widehat{\boldsymbol{\beta}} \widehat{\boldsymbol{\theta}} + \widehat{\boldsymbol{\rho}} \widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\varpi}}, \tag{4}$$

where $\widehat{\boldsymbol{\varpi}} = (\widehat{\varpi}_1, \ldots, \widehat{\varpi}_n)'$ is the $n \times 1$ vector of pricing errors. For interpretive convenience we focus on the empirically relevant case in which (i) the omitted factors are priced, $\boldsymbol{\gamma} \neq \mathbf{0}$, and (ii) $\gamma_l \mathbb{E}_n(\rho_{i,l}) > 0$ for $l = 1, \ldots, L$. These restrictions are not required for the algebraic results below.[9]

[9] These two interpretive restrictions on the omitted factors make economic sense since omitted factors should be priced to be relevant in the first place. Moreover, given positive risk premia for each asset, $\mu_i > 0$, it is natural to focus on the case in which the omitted factors generate positive risk premia on average; otherwise, the omitted-factor component would do worse than a constant when pricing the test assets.

Assumption 2. Let $\boldsymbol{W} = \left[\mathbf{1} \;:\; \widehat{\boldsymbol{\beta}}\right] \in \mathbb{R}^{n\times(K+1)}$, with $\mathbf{1}$ denoting the $n \times 1$ vector of ones. We assume $\boldsymbol{W}$ has full column rank.

Assumption 3. The first-stage betas on the included factors, $\widehat{\boldsymbol{\beta}}$, and (when they appear) the first-stage loadings on the omitted factors, $\widehat{\boldsymbol{\rho}}$, are estimated once from the time-series and treated as fixed regressors in the second-pass CSRs. All results in Section 2 are therefore finite-sample algebra conditional on these first-stage estimates.

Assumption 3 makes explicit that our theoretical derivations in this section are exact finite-sample OLS identities conditional on the realized first-pass estimates $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\rho}})$. When the analysis is embedded in standard two-pass asymptotics with $T \to \infty$ (and fixed n), the first-pass estimates satisfy $\widehat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i + O_p(T^{-1/2})$ under regularity conditions (e.g., Shanken (1992); Kan, Robotti, and Shanken (2013)). Hence, the same identities characterize probability limits of the corresponding statistics by continuous mapping.

Consider the auxiliary CS projection of omitted loadings onto the included betas,

$$\widehat{\boldsymbol{\rho}} \;=\; \mathbf{1}\widehat{\boldsymbol{\delta}}_0' \;+\; \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\Delta}} \;+\; \widehat{\boldsymbol{E}},$$

where $\widehat{\boldsymbol{\delta}}_0 \in \mathbb{R}^L$ is the intercept coefficient vector, $\widehat{\boldsymbol{\Delta}} \in \mathbb{R}^{K\times L}$ is the slope coefficient matrix, and $\widehat{\boldsymbol{E}}$ is an $n \times L$ matrix of residuals orthogonal (in sample) to both $\mathbf{1}$ and $\widehat{\boldsymbol{\beta}}$. For an arbitrary omitted loading, $l = 1, \ldots, L$, we have:

$$\widehat{\rho}_{i,l} = \widehat{\delta}_{0,l} + \sum_{j=1}^{K} \widehat{\delta}_{j,l}\widehat{\beta}_{i,j} + \widehat{e}_{i,l}, \qquad i = 1, \ldots, n.$$

The OLS estimates are:

$$\widehat{\boldsymbol{\Delta}} \;=\; \boldsymbol{\Sigma}^{-1}\,\mathrm{Cov}_n\left(\widehat{\boldsymbol{\beta}}_i, \widehat{\boldsymbol{\rho}}_i'\right), \qquad \widehat{\boldsymbol{\delta}}_0 \;=\; \mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i) \;-\; \widehat{\boldsymbol{\Delta}}'\boldsymbol{m}. \tag{5}$$

The following proposition states the usual omitted-variable-bias mapping in our specific setting.

Proposition 1. *Let $(\widehat{\theta}_0, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\gamma}})$ be the OLS coefficients from the augmented CSR (4), where $[\mathbf{1} : \widehat{\boldsymbol{\beta}} : \widehat{\boldsymbol{\rho}}]$ is assumed to have full column rank.*[10] *Conditional on the realized first-pass estimates $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\rho}})$ and*

[10] The rank condition makes the augmented coefficients unique and the identities in (6) hold for any least-squares solution of the augmented CSR.

under Assumption 2, the baseline second-pass OLS coefficients from (2) satisfy

$$\widehat{\lambda}_0 \;=\; \widehat{\theta}_0 \;+\; \widehat{\boldsymbol{\delta}}_0'\widehat{\boldsymbol{\gamma}}, \qquad \widehat{\boldsymbol{\lambda}} \;=\; \widehat{\boldsymbol{\theta}} \;+\; \widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\gamma}}, \tag{6}$$

where

$$\begin{bmatrix} \widehat{\boldsymbol{\delta}}_0' \\ \widehat{\boldsymbol{\Delta}} \end{bmatrix} = (\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\widehat{\boldsymbol{\rho}}$$

is the OLS coefficient matrix from the CS projection of $\widehat{\boldsymbol{\rho}}$ on $\boldsymbol{W} = [\mathbf{1} : \widehat{\boldsymbol{\beta}}]$. If, in population, the augmented model is correctly specified with $\theta_0 = 0$ and no structural pricing errors, then the population analogue of (6) implies $\lambda_0 = \boldsymbol{\delta}_0'\boldsymbol{\gamma}$, that is, the intercept is driven by the price-weighted unspanned mean of the omitted loadings.

Relative to the risk prices in the augmented CSR, the baseline prices on the included factors are tilted by $\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\gamma}}$, while the baseline intercept equals the intercept in the augmented CSR plus the price-weighted average omitted loading that is not explained by the included betas.[11] Two economically distinct special cases of Proposition 1 are worth highlighting. From (5), we have $\widehat{\boldsymbol{\delta}}_0 = \mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i) - \widehat{\boldsymbol{\Delta}}'\boldsymbol{m}$. Hence, whenever $\widehat{\boldsymbol{\Delta}}'\boldsymbol{m} = \mathbf{0}$,

$$\widehat{\lambda}_0 = \widehat{\theta}_0 + \widehat{\boldsymbol{\gamma}}'\mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i) \approx \widehat{\boldsymbol{\gamma}}'\mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i)\,,$$

where the approximation treats $\widehat{\theta}_0$ as a sampling term under $\theta_0 = 0$ in population. Two natural configurations deliver this simplification. First, if $\boldsymbol{m} = \mathbf{0}$, then $\widehat{\boldsymbol{\delta}}_0 = \mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i)$. Equivalently, the fitted component $\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\Delta}}$ has zero CS mean, so the projection of the omitted loadings on the included betas cannot absorb any part of the mean omitted loading. The entire risk-price-weighted average omitted exposure therefore appears in the intercept. The included risk prices, however, can still be tilted, since $\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\gamma}}$ need not vanish. Second, if $\mathrm{Cov}_n\left(\widehat{\boldsymbol{\beta}}_i, \widehat{\boldsymbol{\rho}}_i'\right) = \mathbf{0}$ then, by (5), $\widehat{\boldsymbol{\Delta}} = \mathbf{0}$. The included and omitted loadings have no CS comovement after demeaning across assets, so the projection has no slope part and OLS does not tilt the included risk prices: $\widehat{\boldsymbol{\lambda}} = \widehat{\boldsymbol{\theta}}$. The same simplified intercept

[11] In recent closely related work, Giglio and Xiu (2021) identify an observable factor's risk premium under omitted priced factors via a three-pass PCA-based spanning approach. Our result instead provides an exact finite-sample omitted-variable-bias decomposition for the misspecified CSR: the intercept $\widehat{\boldsymbol{\delta}}_0$ and slope $\widehat{\boldsymbol{\Delta}}$ from the auxiliary projection of $\widehat{\boldsymbol{\rho}}$ on $[\mathbf{1}, \widehat{\boldsymbol{\beta}}]$, weighted by the omitted risk prices $\widehat{\boldsymbol{\gamma}}$, are respectively the unrestricted intercept and the tilt in the estimated included-factor risk prices.

formula follows, but now because the omitted-variable-bias channel in the slopes is absent.

The following proposition provides a useful interpretation of the CS intercept.

Proposition 2. *Combining* (5) *with* (6) *yields*

$$\widehat{\lambda}_0 \;=\; \widehat{\theta}_0 \;+\; \widehat{\boldsymbol{\gamma}}'\mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i) \;-\; \boldsymbol{m}'\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\gamma}} \;=\; \widehat{\theta}_0 \;+\; \widehat{\boldsymbol{\gamma}}'\mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i) \;-\; \boldsymbol{m}'\left(\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\theta}}\right).$$

Thus, abstracting from any structural intercept (e.g., under a linear SDF with $\theta_0 = 0$), the baseline intercept has two components: (i) The price-weighted average exposure to omitted factors, $\widehat{\boldsymbol{\gamma}}'\mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i)$; and (ii) the part of that average risk premia that the baseline model reallocates to the included factors by tilting their prices from $\widehat{\boldsymbol{\theta}}$ to $\widehat{\boldsymbol{\lambda}}$, namely $\boldsymbol{m}'(\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\theta}})$. This term enters with a minus sign in the identity above; its quantitative effect can be positive or negative depending on the direction of the tilt on the risk prices. The relative magnitude of the two components is not determined a priori. Concretely, the second component equals $\boldsymbol{m}'\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\gamma}}$ and is small whenever the tilt $\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\gamma}}$ is small or nearly orthogonal to $\boldsymbol{m}$.

Importantly, Proposition 2 provides guidance on the actual intercept estimates obtained in CS tests (under the assumption that $\widehat{\theta}_0 \approx 0$). First, it follows that large (positive) intercept estimates, which arise in many empirical applications (as discussed in the following sections), are more likely to occur in two scenarios: (i) when $-\boldsymbol{m}'(\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\theta}}) > 0$, that is, when the tilt in the included-factor risk prices has a negative projection on the mean-loading direction $\boldsymbol{m}$, which implies $\widehat{\lambda}_0 \approx \widehat{\boldsymbol{\gamma}}'\mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i) - \boldsymbol{m}'(\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\theta}}) > \widehat{\boldsymbol{\gamma}}'\mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i)$;[12] and (ii) when $\boldsymbol{m}'(\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\theta}}) \approx 0$, which implies $\widehat{\lambda}_0 \approx \widehat{\boldsymbol{\gamma}}'\mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i) > 0$, and can arise from either small average betas for the included factors, $\boldsymbol{m} \approx \mathbf{0}$, or an approximate zero bias in risk price estimates, $\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\theta}} \approx \mathbf{0}$, with this second case being discussed further below. Second, given the assumption that $\widehat{\boldsymbol{\gamma}}'\mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i) > 0$, the rarer case of a negative intercept estimate ($\widehat{\lambda}_0 < 0$) requires $-\boldsymbol{m}'(\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\theta}}) < 0$, equivalently a positive projection of the tilt $\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\theta}}$ on the mean-loading direction $\boldsymbol{m}$.[13]

Another important implication from Proposition 2 is that $\widehat{\lambda}_0$ can be close to zero even if there are omitted factors ($\widehat{\boldsymbol{\gamma}} \neq \mathbf{0}$) whenever $\widehat{\boldsymbol{\gamma}}'\mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i) \approx \boldsymbol{m}'\left(\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\theta}}\right) > 0$ (assuming that $\widehat{\theta}_0$ is negligible).

[12] In the single-factor case ($K = 1$), this condition means that the average beta (m) and the tilt in the risk price estimate, $\widehat{\lambda} - \widehat{\theta}$, have opposite signs.

[13] In the single-factor case ($K = 1$), this condition means that the average beta (m) and the tilt in the risk price estimate, $\widehat{\lambda} - \widehat{\theta}$, have the same sign.

Hence, a zero intercept in the cross-section is necessary for correct model specification, but not sufficient. In other words, a zero intercept does not preclude missing risk factors, whereas, within this framework, a non-zero intercept implies omitted priced risks. Therefore, a zero CS intercept can occur in empirical scenarios in which the tilt in risk prices for the included factors is large (in magnitude), which indicates a large correlation between the included and omitted loadings (i.e., $\widehat{\boldsymbol{\Delta}}$ is "large").

A further relevant question is under what conditions do the baseline risk prices equal the counterparts in the augmented CSR so that the intercept reduces (up to $\widehat{\theta}_0$) to $\widehat{\boldsymbol{\gamma}}'\mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i)$.[14] From Proposition 1, we have, $\widehat{\boldsymbol{\lambda}} = \widehat{\boldsymbol{\theta}} \iff \widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\gamma}} = \mathbf{0}$, which holds if $\widehat{\boldsymbol{\Delta}} = \mathbf{0}$, i.e., if the omitted exposures have zero CS covariance with the included betas after demeaning across assets. Therefore, in that configuration the baseline risk prices are close to their augmented counterparts and the intercept absorbs approximately the full price-weighted average omitted exposure. The results above are purely algebraic and require no time-series orthogonality restrictions. The main conclusion is that large CS intercepts are consistent with economically significant omitted risks with high average exposure.

Finally, one may ask whether empirical specifications with missing factors are compatible with the underlying theory of a given factor model. For the (unconditional) Consumption–CAPM of Lucas (1978) and Breeden (1979), and many tightly parameterized extensions (e.g., Epstein and Zin (1991); Yogo (2006)), the SDF has a structural form in consumption growth (and related macro factors) and a small set of preference parameters. Allowing additional terms in the SDF, which are not validated by theory, severs the mapping from the risk price estimates to the underlying structural parameters and affects the economic interpretation of the model; put differently, a nonzero intercept generated by omitted factors reflects misspecification relative to the theoretical consumption-based model that one wants to test in the first place, something that represents a violation of the underlying theory.[15] This contrasts with more generic frameworks such as Merton's ICAPM (Merton (1973)), where the SDF depends on innovations in state variables that govern the investment opportunity set, but the theory does not identify those variables. Empirical

[14] A similar expression for the intercept is obtained in Maio and Min (2026) by relying on a simplified framework: $K = L = 1$ combined with the assumption that the betas for the observed factor have a zero mean.

[15] Additional priced components can be rationalized by measurement error in consumption; in that case the "extra" terms are part of the extended theoretical structure, not violations.

implementations may legitimately omit some hedging (non-market) risks without contradicting the model's theoretical content. Proposition 1 then characterizes the algebraic consequences for the baseline with-intercept CSR; see the first identity in (6) for the corresponding intercept mapping. The same logic applies to conditional tests of the Consumption–CAPM (e.g., Lettau and Ludvigson (2001); Cochrane (2005)): the lagged instruments that generate the scaled factors are not singled out by theory, so an empirical design may omit relevant interactions while remaining consistent with the underlying conditional Euler equation. In such cases, the intercept in the corresponding unconditional approximation is expected to be nonzero and equals the price-weighted unspanned mean of the omitted scaled loadings.

In the IA, we show that classical measurement error in the factors rescales risk prices but cannot generate a nonzero CSR intercept unless the measurement-error component carries systematic exposures across assets, a case observationally equivalent to an omitted factor.

2.2 The CS Intercept and Pricing Errors

Building on the omitted-factor characterization in Subsection 2.1, we now examine how the CS intercept affects pricing errors and common measures of model pricing performance. The specification being tested corresponds to the baseline CSR (with an intercept) in (1). Let $s = \boldsymbol{m}'\boldsymbol{M}^{-1}\boldsymbol{m} = \boldsymbol{m}'\boldsymbol{u}$ denote the R^2 from the auxiliary (no-intercept) regression $1 = \widehat{\boldsymbol{\beta}}_i'\boldsymbol{u} + \varepsilon_i$, with $\boldsymbol{u} = \boldsymbol{M}^{-1}\boldsymbol{m}$ (share of the constant vector $\mathbf{1}$ explained by the span of $\widehat{\boldsymbol{\beta}}$).[16] Under Assumption 2, $s \in [0, 1)$. In addition to **Assumptions 1 − 3**, the following assumption holds:

Assumption 4. The mean excess returns, μ_i, exhibit CS variation, that is, $\mathrm{Var}_n(\mu_i) > 0$.

When returns are expressed in excess of the relevant benchmark rate (usually the risk-free return), theory implies a zero second-pass intercept. This makes the no-intercept CSR the natural benchmark for interpreting the unrestricted specification. It measures total pricing error under the maintained model. It also isolates the mean pricing error that the unrestricted intercept absorbs once the constant is allowed to enter the cross-section. We therefore introduce the OLS restricted CSR,

$$\mu_i \;=\; \widetilde{\boldsymbol{\lambda}}'\widehat{\boldsymbol{\beta}}_i \;+\; \widehat{\nu}_i, \tag{7}$$

[16] When $K = 1$, s coincides with the invariance-of-beta (IB) collinearity diagnostic of Ahn, Perez, and Gadarowski (2013) for near-constant market betas; for $K > 1$, it extends that diagnostic to the span of all factor loadings.

where $\widetilde{\boldsymbol{\lambda}}$ are the OLS risk prices and $\widehat{\nu}_i$ are the corresponding OLS CS residuals.

The next lemma states a normal-equation identity connecting the restricted and unrestricted CSRs. We use it repeatedly below and in Subsections 2.3 and 2.4.

Lemma 1. *Under Assumption 2, letting* $\boldsymbol{u} = \boldsymbol{M}^{-1}\boldsymbol{m}$ *and* $s = \boldsymbol{m}'\boldsymbol{u}$,

$$\widehat{\boldsymbol{\lambda}} - \widetilde{\boldsymbol{\lambda}} = -\widehat{\lambda}_0\, \boldsymbol{u}.$$

The lemma shows that introducing an intercept changes the risk prices in a structured way: the unrestricted risk prices differ from the restricted counterparts only through a shift in the direction $-\boldsymbol{u}$. The next proposition gives the corresponding expression for the intercept in terms of the average pricing error from the restricted CSR.

Proposition 3. *Given Assumption 2, the intercept in the unrestricted CSR satisfies*

$$\widehat{\lambda}_0 = \frac{\mathbb{E}_n(\widehat{\nu}_i)}{1-s} = \frac{\overline{\mu} - \boldsymbol{m}'\widetilde{\boldsymbol{\lambda}}}{1-s}. \tag{8}$$

A large intercept can therefore arise from a large CS mean pricing error in the restricted CSR, $\mathbb{E}_n(\widehat{\nu}_i)$, from a large value of s, or from both. Since s is large when the factor loadings have large CS means relative to their CS dispersion, the unrestricted CSR magnifies the restricted mean pricing error precisely when the constant is nearly in the span of the betas. In light of Propositions 1 and 2, a nonzero $\mathbb{E}_n(\widehat{\nu}_i)$ is consistent with omitted priced factors. Equation (8) therefore links a large intercept directly to model misspecification and to weak CS dispersion in the included betas.

Given the discussion in the previous subsection, we interpret the intercept as a common pricing offset and define the pricing-error vector shifted by the intercept,

$$\widehat{\alpha}_i \;=\; \mu_i - \widehat{\boldsymbol{\lambda}}'\widehat{\boldsymbol{\beta}}_i \;=\; \widehat{\omega}_i \;+\; \widehat{\lambda}_0, \tag{9}$$

where $\widehat{\omega}_i$ are the OLS residuals, with $\mathbb{E}_n(\widehat{\omega}_i) = 0$. Hence, the average pricing error becomes the intercept estimate, $\mathbb{E}_n(\widehat{\alpha}_i) = \widehat{\lambda}_0$. In other words, in comparison to the CS residuals, $\widehat{\omega}_i$ (which average to zero), $\widehat{\alpha}_i$ shifts all pricing errors by a constant, which is the common intercept ($\widehat{\lambda}_0$).

Given the baseline CS OLS R^2, $R^2_{\text{ols}} = 1 - \mathbb{E}_n\left(\widehat{\omega}_i^2\right) / \operatorname{Var}_n(\mu_i)$, we define a similar metric that

is based on the "total" pricing errors (rather than the regression residuals):

$$R^2_{\text{pen}} = 1 - \frac{\mathbb{E}_n\left(\widehat{\alpha}_i^2\right)}{\text{Var}_n\left(\mu_i\right)}. \tag{10}$$

R^2_{ols} credits the intercept as part of the fitted risk premia from the CSR, while R^2_{pen} treats it as part of total pricing error, consistent with the omitted-factor interpretation of the intercept in Subsection 2.1.[17] We therefore interpret R^2_{pen} as a dispersion-scaled total-pricing-error diagnostic for the unrestricted CSR rather than as a pure dispersion-fit statistic. Keeping $\text{Var}_n\left(\mu_i\right)$ in the denominator preserves comparability with the conventional CS R^2 and yields the exact wedge decomposition below. The corresponding restricted-CSR diagnostic, $R^2_{\text{ols}*}$, is presented below; it is likewise a total-error measure, since the restricted residuals $\widehat{\nu}_i$ need not average to zero. The (negative) intercept's contribution to the model's fit (the *Wedge*) is given by $\text{Wedge} = \widehat{\lambda}_0^2/\text{Var}_n(\mu_i)$, as shown in the following proposition.

Proposition 4. *Assuming* $\text{Var}_n(\mu_i) > 0$ *(Assumption 4) so that the* R^2 *ratios are well-defined, with* $\widehat{\alpha}_i = \widehat{\lambda}_0 + \widehat{\omega}_i$ *and* $\mathbb{E}_n(\widehat{\omega}_i) = 0$*,*

$$R^2_{\text{pen}} \;=\; R^2_{\text{ols}} - \frac{\widehat{\lambda}_0^2}{\text{Var}_n(\mu_i)} \;=\; R^2_{\text{ols}} - \text{Wedge}, \qquad \textit{equivalently} \qquad \mathbb{E}_n\left(\widehat{\alpha}_i^2\right) \;=\; \mathbb{E}_n\left(\widehat{\omega}_i^2\right) + \widehat{\lambda}_0^2. \tag{11}$$

Moreover, $R^2_{\text{pen}} = R^2_{\text{ols}}$ *iff* $\widehat{\lambda}_0 = 0$*.*

This proposition makes clear that the intercept has a negative effect on the model's fit, and that this effect increases with the magnitude of the intercept. Thus, given two models, ceteris paribus, the model that produces a larger (in magnitude) intercept estimate will have a lower CS fit. The following corollary provides further intuition on the new CS R^2.

Corollary 1. *Let* $\text{Wedge} = \widehat{\lambda}_0^2/\text{Var}_n(\mu_i)$*. Then* $R^2_{\text{pen}} = R^2_{\text{ols}} - \text{Wedge}$ *and, since* $R^2_{\text{ols}} \in [0,1]$ *for the OLS regression with an intercept,*

$$R^2_{\text{pen}} \in [-\text{Wedge}, 1-\text{Wedge}], \qquad R^2_{\text{pen}} < 0 \iff \widehat{\lambda}_0^2 > \text{Var}_n(\mu_i)\, R^2_{\text{ols}} \iff \widehat{\lambda}_0^2 > \text{Var}_n(\mu_i) - \mathbb{E}_n\left(\widehat{\omega}_i^2\right).$$

[17] Another frequently used R^2 relies on the CS variance (rather than the CS second moment) of the pricing errors, $\text{Var}_n\left(\widehat{\alpha}_i\right)$. However, this alternative measure is not impacted by the new definition of the pricing errors (in the CSR with an intercept), $R^2 = 1 - \text{Var}_n\left(\widehat{\alpha}_i\right)/\text{Var}_n\left(\mu_i\right) = 1 - \text{Var}_n\left(\widehat{\omega}_i\right)/\text{Var}_n\left(\mu_i\right) = 1 - \mathbb{E}_n\left(\widehat{\omega}_i^2\right)/\text{Var}_n\left(\mu_i\right)$ given that $\widehat{\lambda}_0$ is common across test assets.

Corollary 1 implies that a large estimated common pricing offset shifts the feasible range of R^2_{pen} downward. In particular, if $\widehat{\lambda}_0^2 > \text{Var}_n(\mu_i)$, then even the upper bound, $1 - \widehat{\lambda}_0^2/\text{Var}_n(\mu_i)$, is negative. The lower bound, $-\widehat{\lambda}_0^2/\text{Var}_n(\mu_i)$, is attained when $R^2_{\text{ols}} = 0$. More generally, negative values of R^2_{pen} are more likely when the estimated common pricing offset is large, when the unrestricted CS residual variance $\mathbb{E}_n\left(\widehat{\omega}_i^2\right)$ is large, or when the CS dispersion in risk premia, $\text{Var}_n(\mu_i)$, is small. In this sense, $R^2_{\text{pen}} < 0$ indicates that the penalty from the common pricing offset outweighs the CS variation explained by the unrestricted CSR.

It is also useful to compare the new fit measure from the unrestricted CSR with the standard OLS fit from the restricted CSR. Each statistic is the appropriate explanatory ratio for its own specification. The next corollary gives the exact relation between them.

Corollary 2. *Given*

$$R^2_{\text{ols}*} = 1 - \frac{\mathbb{E}_n\left(\widehat{\nu}_i^2\right)}{\text{Var}_n(\mu_i)},$$

we have:

$$R^2_{\text{pen}} \;\leq\; R^2_{\text{ols}*}, \qquad R^2_{\text{ols}*} - R^2_{\text{pen}} \;=\; s\frac{\widehat{\lambda}_0^2}{\text{Var}_n(\mu_i)} = s\text{Wedge}.$$

Hence $R^2_{\text{pen}} = R^2_{\text{ols}*}$ *iff either* $\widehat{\lambda}_0 = 0$ *or* $s = 0$ *(equivalently* $\boldsymbol{m} = \mathbf{0}$*).*

Thus, once the intercept is treated as pricing error, the unrestricted CSR cannot outperform the restricted CSR in terms of fit. The gap $R^2_{\text{ols}*} - R^2_{\text{pen}}$ widens with s directly and also indirectly because Proposition 3 implies $\widehat{\lambda}_0 = \mathbb{E}_n(\widehat{\nu}_i)/(1-s)$. Hence, a given mean pricing error in the restricted CSR generates a larger fit gap when the constant is nearly in the span of the betas. If $\mathbb{E}_n(\widehat{\nu}_i) = 0$, then $\widehat{\lambda}_0 = 0$ and the gap collapses to zero. Since $\boldsymbol{M} \succ 0$, $s = \boldsymbol{m}'\boldsymbol{M}^{-1}\boldsymbol{m} = 0$ if and only if $\boldsymbol{m} = \mathbf{0}$. In that case, $\boldsymbol{u} = \mathbf{0}$ and the two regressions produce the same risk price estimates.[18] The next corollary characterizes when this gap exceeds the restricted fit itself, so that the penalized fit is negative while the restricted fit is positive.[19]

[18] In this case, $\widehat{\lambda}_0 = \overline{\mu}$ and $\widehat{\nu}_i - \widehat{\omega}_i = \widehat{\lambda}_0$, so $\widehat{\nu}_i = \widehat{\alpha}_i$ and therefore $R^2_{\text{pen}} = R^2_{\text{ols}*}$.

[19] Kan, Robotti, and Shanken (2013) rely on $R^2_{\text{ols}**} = 1 - \mathbb{E}_n\left(\widehat{\nu}_i^2\right)/\mathbb{E}_n\left(\mu_i^2\right)$ in their analysis of the restricted CSR. The denominator is the uncentered CS second moment of expected returns, rather than the CS variance. In line with most of the literature, we do not adopt this metric because $R^2_{\text{ols}**}$ is mechanically bounded below by zero and can therefore be uninformative about whether a model prices the cross-section better than a constant. Empirically, a model can produce a large $R^2_{\text{ols}**}$, suggesting good pricing performance, while generating a negative $R^2_{\text{ols}*}$. Maio and Santa-Clara (2017) document this pattern for the CAPM.

Corollary 3. *Overshooting, in the sense that*

$$R^2_{\text{pen}} < 0 < R^2_{\text{ols}*},$$

occurs if and only if

$$0 < R^2_{\text{ols}*} < s\frac{\widehat{\lambda}_0^2}{\text{Var}_n(\mu_i)} = \frac{s}{(1-s)^2}\frac{\mathbb{E}_n(\widehat{\nu}_i)^2}{\text{Var}_n(\mu_i)}.$$

Equivalently,

$$\mathbb{E}_n(\widehat{\nu}_i^2) < \text{Var}_n(\mu_i) < \mathbb{E}_n(\widehat{\nu}_i^2) + s\widehat{\lambda}_0^2.$$

Corollary 3 shows that overshooting arises when the restricted CSR delivers a positive $R^2_{\text{ols}*}$ but leaves a nonzero mean pricing error that is strongly amplified by a large value of s. The condition above is therefore useful for distinguishing cases in which a negative R^2_{pen} reflects amplification of a common pricing offset from cases in which fit is weak under both specifications.

Another widely used performance metric is the root mean squared pricing error relative to the root mean squared risk premium (e.g., Lettau, Ludvigson, and Ma (2019)), $RRMSE = \sqrt{\mathbb{E}_n\left(\widehat{\omega}_i^2\right)/\mathbb{E}_n\left(\mu_i^2\right)}$.[20] Similarly to the R^2 metrics, the pricing errors should contain the intercept estimate, and failure to do so might (substantially) understate the relative pricing error. This is summarized in the following proposition.

Proposition 5. *The relative root mean squared total pricing error in the unrestricted CSR is*

$$RRMSE_{\text{pen}} = \sqrt{\frac{\widehat{\lambda}_0^2}{\mathbb{E}_n\left(\mu_i^2\right)} + \frac{\mathbb{E}_n\left(\widehat{\omega}_i^2\right)}{\mathbb{E}_n\left(\mu_i^2\right)}} \geq \sqrt{\frac{\mathbb{E}_n\left(\widehat{\omega}_i^2\right)}{\mathbb{E}_n\left(\mu_i^2\right)}}, \tag{12}$$

which means that $\widehat{\lambda}_0 \neq 0 \Rightarrow RRMSE_{\text{pen}} > RRMSE$.

Similarly to the case of the CS R^2, the intercept introduces an additive wedge in the squared relative pricing error,

$$RRMSE^2_{\text{pen}} - RRMSE^2 = \frac{\widehat{\lambda}_0^2}{\mathbb{E}_n\left(\mu_i^2\right)} = \text{Wedge} \times \frac{\text{Var}_n(\mu_i)}{\mathbb{E}_n\left(\mu_i^2\right)}.$$

This term is typically ignored in the standard definition, which is based only on the CS residuals,

[20] A closely related version is based on the mean absolute pricing error (e.g., Maio (2024)), $RMAE = \sum_{i=1}^n |\widehat{\omega}_i| / \sum_{i=1}^n |\mu_i|$.

$\widehat{\omega}_i$. Thus, models with large intercept estimates are more likely to exhibit a large $RRMSE_{\text{pen}}$. Trivially, when the intercept is zero it follows that the two metrics coincide, $RRMSE_{\text{pen}} = RRMSE$.

The magnitude of $\widehat{\lambda}_0$ matters for pricing-error measures even when it is statistically insignificant. Just as all CS residuals enter the computation of the average pricing error and the CS R^2 regardless of their individual significance, the intercept must also enter these metrics whenever it is included in the CSR.

The key implication is that there is no free lunch when adopting the two-pass procedure to test linear asset pricing models. If one includes the intercept in the CSR, something inconsistent with theory, then the intercept must be treated as a common pricing error, and performance metrics must be adjusted accordingly. Failure to do so yields a misleading picture of model fit.

2.3 The CS Intercept and Risk Price Estimates

Having examined how the intercept affects pricing errors and fit metrics, we now study how including an intercept changes the estimated factor risk prices. We consider the unrestricted OLS CSR in (1) and its restricted counterpart (without intercept) in (7). The following proposition complements Lemma 1 by expressing the unrestricted risk prices in closed form, in terms of the restricted counterparts and the CS moments of the loadings.

Proposition 6. *Given Assumption 2 (so that $\boldsymbol{M} \succ 0$ and $\boldsymbol{\Sigma} \succ 0$, equivalently $s < 1$), both the restricted and unrestricted CSRs are identified. Then*

$$\widehat{\boldsymbol{\lambda}} \;=\; \boldsymbol{\Sigma}^{-1}\big(\boldsymbol{M}\widetilde{\boldsymbol{\lambda}} - \boldsymbol{m}\,\overline{\mu}\big) \;=\; \widetilde{\boldsymbol{\lambda}} \;+\; \frac{\boldsymbol{m}'\widetilde{\boldsymbol{\lambda}} - \overline{\mu}}{1-s}\,\boldsymbol{u}, \tag{13}$$

or equivalently, recovering the identity of Lemma 1,

$$\widehat{\boldsymbol{\lambda}} - \widetilde{\boldsymbol{\lambda}} = -\widehat{\lambda}_0 \boldsymbol{u}. \tag{14}$$

Proposition 3 shows that $\widehat{\lambda}_0 = \mathbb{E}_n(\widehat{\nu}_i)/(1-s)$, so a large slope gap ultimately reflects three primitives: a large mean pricing error in the restricted CSR, large average loadings, and weak CS variation in those loadings through $\boldsymbol{M}^{-1}$. Conditional on the intercept, the shift in risk prices is larger when $\boldsymbol{M}^{-1}\boldsymbol{m}$ is large. Recall from Subsection 2.2 that s is the R^2 from the auxiliary (no-intercept) regression $1 = \widehat{\boldsymbol{\beta}}_i'\boldsymbol{u} + \varepsilon_i$ with $\boldsymbol{u} = \boldsymbol{M}^{-1}\boldsymbol{m}$. Thus $1 - s = \mathbb{E}_n(\varepsilon_i^2)$ is the mean squared

residual from projecting the constant vector onto the span of the factor loadings and therefore controls how close the design matrix $\boldsymbol{W} = [\mathbf{1} : \widehat{\boldsymbol{\beta}}]$ is to being rank deficient; Section 2.4 formalizes this link.[21]

In this regime, (8) and (14) imply that even a modest nonzero mean pricing error in the restricted CSR, $\mathbb{E}_n(\widehat{\nu}_i)$, can translate into a large change in the unrestricted estimates because the intercept scales with $(1-s)^{-1}$, and the slope shift inherits this amplification through $\widehat{\lambda}_0$. As $s \to 1$, the intercept diverges for any fixed nonzero mean pricing error, and the slope shift can also become arbitrarily large.[22] We refer to this asset-pricing manifestation of familiar OLS ill-conditioning as the *collinearity amplification channel*. It is most acute when the constant vector is nearly in the span of the factor loadings, for example, when a factor's betas are nearly constant across assets at a nonzero level. A salient empirical example is the market factor, whose betas are often close to constant across assets (e.g., Ahn, Perez, and Gadarowski (2013); Ahn and Horenstein (2020)).

Equation (14) also implies that when $\mathbb{E}_n(\widehat{\nu}_i) = \overline{\mu} - \boldsymbol{m}'\widetilde{\boldsymbol{\lambda}} > 0$ (equivalently, $\widehat{\lambda}_0 > 0$), the sign of each individual risk price shift is determined by u_j: $\widehat{\lambda}_j - \widetilde{\lambda}_j < 0$ when $u_j > 0$, and $\widehat{\lambda}_j - \widetilde{\lambda}_j > 0$ when $u_j < 0$. More generally, the componentwise gap satisfies $|\widehat{\lambda}_j - \widetilde{\lambda}_j| = |\widehat{\lambda}_0|\,|u_j|$, so larger intercepts and larger $|u_j|$ mechanically widen the difference between unrestricted and restricted risk price estimates. Corollaries 5 and 6 below characterize the resulting attenuation and sign-flip conditions.

The following corollary states the conditions under which the gap in risk price estimates is zero:

Corollary 4. *Under* $\boldsymbol{M} \succ 0$ *and* $\boldsymbol{\Sigma} \succ 0$,

$$\widehat{\boldsymbol{\lambda}} - \widetilde{\boldsymbol{\lambda}} = \mathbf{0} \quad \Longleftrightarrow \quad \Big(\widehat{\lambda}_0 = 0\Big) \text{ or } \Big(\boldsymbol{m} = \mathbf{0}\Big).$$

By Proposition 3, $\widehat{\lambda}_0 = 0$ *is equivalent to* $\mathbb{E}_n(\widehat{\nu}_i) = 0$.

Thus, a zero intercept, or equivalently, a zero mean residual in the restricted CSR, is the most direct case in which the two sets of risk prices coincide. The same conclusion also holds when the average loadings for all factors are zero. Empirically, models with either a small intercept or small average factor loadings are more likely to deliver similar risk price estimates under the unrestricted

[21] Geometrically, near-collinearity arises when the mean loading vector $\boldsymbol{m}$ is large relative to the CS dispersion of the loadings along its direction, so that $\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{m}\|^2$ is large and s is close to one. This can reflect large mean loadings, small dispersion, or both.

[22] For a related omitted-variable sensitivity result linking near multicollinearity to coefficient-sign fragility, see Masten and Poirier (2026).

and restricted CSRs.

The next corollary provides the relations for each individual risk price (under the generic case).

Corollary 5. *For $j = 1, \ldots, K$,*

$$\widehat{\lambda}_j \;=\; \widetilde{\lambda}_j \;-\; \widehat{\lambda}_0\, u_j,$$

where u_j is the jth element of $\boldsymbol{u} = \boldsymbol{M}^{-1}\boldsymbol{m}$ and $\widehat{\lambda}_0 = (\overline{\mu} - \boldsymbol{m}'\widetilde{\boldsymbol{\lambda}})/(1-s)$ is the intercept estimate in the unrestricted CSR. It follows:

(i) Attenuation without sign change. *If $\widetilde{\lambda}_j > 0$, attenuation ($0 < \widehat{\lambda}_j < \widetilde{\lambda}_j$) occurs iff $0 < \widehat{\lambda}_0 u_j < \widetilde{\lambda}_j$; if $\widetilde{\lambda}_j < 0$, attenuation ($\widetilde{\lambda}_j < \widehat{\lambda}_j < 0$) occurs iff $\widetilde{\lambda}_j < \widehat{\lambda}_0 u_j < 0$.*

(ii) Sign flip. $\operatorname{sign}(\widehat{\lambda}_j) \neq \operatorname{sign}(\widetilde{\lambda}_j)$ *iff*

$$\left(\widetilde{\lambda}_j - \widehat{\lambda}_0 u_j\right)\widetilde{\lambda}_j < 0, \quad \textit{equivalently} \quad \begin{cases} \widehat{\lambda}_0 u_j > \widetilde{\lambda}_j, & \textit{if } \widetilde{\lambda}_j > 0, \\ \widehat{\lambda}_0 u_j < \widetilde{\lambda}_j, & \textit{if } \widetilde{\lambda}_j < 0. \end{cases}$$

Therefore, for fixed u_j and mean pricing error, a rise in s increases the componentwise gap $|\widehat{\lambda}_j - \widetilde{\lambda}_j| = |\mathbb{E}_n(\widehat{\nu}_i)|\,|u_j|/(1-s)$. To obtain a better intuition on the decoupling between the two sets of risk price estimates, we assume the single-factor model. The following corollary holds as a special case of Corollary 5 under $K = 1$.

Corollary 6. *Suppose $K = 1$. Let $m = \mathbb{E}_n\left(\widehat{\beta}_i\right)$, $M = \mathbb{E}_n\left(\widehat{\beta}_i^2\right)$, $\sigma^2 = \operatorname{Var}_n\left(\widehat{\beta}_i\right) = M - m^2 > 0$, and $\overline{\mu} = \mathbb{E}_n(\mu_i)$. Then*

$$\widehat{\lambda} \;=\; \frac{M\,\widetilde{\lambda} - \overline{\mu}\,m}{\sigma^2}, \qquad \widehat{\lambda} - \widetilde{\lambda} \;=\; -\widehat{\lambda}_0\,\frac{m}{M} \;=\; \frac{m(\widetilde{\lambda}m - \overline{\mu})}{\sigma^2}.$$

Thus, in the single-factor case the gap in risk prices is governed by three primitive objects: the mean pricing error from the restricted CSR, through $\widehat{\lambda}_0 = \mathbb{E}_n(\widehat{\nu}_i)/(1-s)$; the average beta m; and the inverse beta second moment M^{-1}.

Assume $\widetilde{\lambda} > 0$ and $m > 0$. Then:

(i) Sign flip. $\widehat{\lambda} < 0$ *occurs iff $\widetilde{\lambda} < \overline{\mu}m/M$.*

(ii) Attenuation without sign change. $0 < \widehat{\lambda} < \widetilde{\lambda}$ *occurs iff $\overline{\mu}m/M < \widetilde{\lambda} < \overline{\mu}/m$. Analogous conditions obtain for $\widetilde{\lambda} < 0$ and $m < 0$ by symmetry.*

Therefore, in the single-factor case with $\widetilde{\lambda} > 0$ and $m > 0$, the discrepancy between unrestricted and restricted risk price estimates is best described by the exact thresholds

$$\widehat{\lambda} < 0 \iff \widetilde{\lambda} < \frac{\overline{\mu} m}{M} = \frac{\overline{\mu}}{m} s, \qquad 0 < \widehat{\lambda} < \widetilde{\lambda} \iff \frac{\overline{\mu}}{m} s < \widetilde{\lambda} < \frac{\overline{\mu}}{m},$$

where $s = m^2/M$ and $\sigma^2 = M - m^2$. A higher s shrinks the attenuation region and, for fixed $(\widetilde{\lambda}, \overline{\mu}, m)$, raises the threshold for a sign flip. Separately, a small σ^2 amplifies the magnitude of the unrestricted estimate through $\widehat{\lambda} = (M\widetilde{\lambda} - \overline{\mu} m)/\sigma^2$. Equivalently, since $v = M\widetilde{\lambda}$, a sign flip occurs iff $v < \overline{\mu} m$.

2.4 Identification and Collinearity

The previous subsection related risk price estimates from restricted and unrestricted CSRs. We now discuss identification and collinearity issues associated with including an intercept.

Let $\boldsymbol{v} = \mathbb{E}_n\left(\widehat{\boldsymbol{\beta}}_i \mu_i\right)$ and $\boldsymbol{c} = \mathrm{Cov}_n\left(\widehat{\boldsymbol{\beta}}_i, \mu_i\right) = \boldsymbol{v} - \boldsymbol{m}\,\overline{\mu}$. Identification of the unrestricted CSR requires $\boldsymbol{\Sigma} = \boldsymbol{M} - \boldsymbol{m}\boldsymbol{m}' \succ \boldsymbol{0}$, since $\widehat{\boldsymbol{\lambda}} = \boldsymbol{\Sigma}^{-1}\boldsymbol{c}$. By contrast, the restricted CSR requires $\boldsymbol{M} \succ 0$ because $\widetilde{\boldsymbol{\lambda}} = \boldsymbol{M}^{-1}\boldsymbol{v}$ (Assumption 2). Both conditions are primitive properties of the regressors in the CSRs, $\{\widehat{\boldsymbol{\beta}}_i\}_{i=1}^n$. We note that

$$s = \boldsymbol{m}'\boldsymbol{M}^{-1}\boldsymbol{m} = \frac{\boldsymbol{m}'\boldsymbol{\Sigma}^{-1}\boldsymbol{m}}{1 + \boldsymbol{m}'\boldsymbol{\Sigma}^{-1}\boldsymbol{m}} \in [0, 1), \qquad 1 - s = \frac{1}{1 + \boldsymbol{m}'\boldsymbol{\Sigma}^{-1}\boldsymbol{m}},$$

so that $s = \|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{m}\|^2/(1 + \|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{m}\|^2)$ and $(1 - s)^{-1} = 1 + \|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{m}\|^2$. The scalar s thus measures how close the constant vector is to the column span of the factor loadings; the following corollary makes this precise.

Corollary 7. *Let $\boldsymbol{W} = [\mathbf{1} : \widehat{\boldsymbol{\beta}}] \in \mathbb{R}^{n \times (K+1)}$ and let $\boldsymbol{M}_1 = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'$. Then the unrestricted CSR is identified (i.e., the OLS coefficients on $[\mathbf{1}, \widehat{\boldsymbol{\beta}}]$ are unique) if and only if*

$$\mathit{rank}(\boldsymbol{M}_1\widehat{\boldsymbol{\beta}}) = K \qquad (\textit{equivalently, } \widehat{\boldsymbol{\beta}}'\boldsymbol{M}_1\widehat{\boldsymbol{\beta}} \textit{ is nonsingular}).$$

Since $\frac{1}{n}\widehat{\boldsymbol{\beta}}'\boldsymbol{M}_1\widehat{\boldsymbol{\beta}} = \mathbb{E}_n\big[(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{m})(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{m})'\big] = \boldsymbol{\Sigma}$, the condition above is equivalent to $\boldsymbol{\Sigma} \succ 0$. Because identification in the unrestricted CSR requires $\boldsymbol{\Sigma} \succ 0$ rather than $\boldsymbol{M} \succ 0$, adding an intercept strictly tightens the rank condition. Empirically, market betas are often close to constant

across assets, pushing s toward one (as noted in Subsection 2.3), a pattern frequently associated with negative market risk price estimates in the unrestricted CSR. More broadly, Ahn, Horenstein, and Wang (2018) document that beta matrices from many macro-factor models fail to achieve full column rank, implying that risk prices in these specifications are underidentified even before adding an intercept. A free intercept then raises the identification hurdle further.[23]

Because $s = \boldsymbol{m}'\boldsymbol{M}^{-1}\boldsymbol{m}$, the same configuration that produces a large gap between unrestricted and restricted risk price estimates also weakens identification in the unrestricted CSR. In this configuration, the constant becomes nearly spanned by the factor loadings, $\boldsymbol{\Sigma}$ approaches singularity, and $(\widehat{\lambda}_0, \widehat{\boldsymbol{\lambda}})$ become both more weakly identified and more sensitive to small perturbations in the data.[24]

2.5 Simple Returns versus Excess Returns

A few studies in the literature (e.g., Kan, Robotti, and Shanken (2013) and Gospodinov, Kan, and Robotti (2019)) employ simple rather than excess returns, in which case the intercept is interpreted as the zero-beta return (Black (1972)). This subsection translates the results of this section to that convention. Let $\overline{R}_i$ denote the sample mean simple return of asset i and $\overline{R}_f$ the sample mean risk-free return used to construct excess returns, so that $\mu_i = \overline{R}_i - \overline{R}_f$. The unrestricted simple-return CSR is

$$\overline{R}_i \;=\; \widehat{r}^{zb} \;+\; \widehat{\boldsymbol{\lambda}}'\widehat{\boldsymbol{\beta}}_i \;+\; \widehat{\omega}_i, \qquad i = 1, \ldots, n, \tag{15}$$

where $\widehat{r}^{zb}$ denotes the estimated zero-beta return. Because $\overline{R}_f\mathbf{1}$ is proportional to the intercept column and Assumption 2 makes the coefficients unique, subtracting $\overline{R}_f$ from the dependent variable alters only the intercept. Regressions (15) and (1) share the same slope estimates $\widehat{\boldsymbol{\lambda}}$, residuals $\widehat{\omega}_i$, and R^2_{ols}, while the intercepts satisfy $\widehat{\lambda}_0 = \widehat{r}^{zb} - \overline{R}_f$. All statements in this subsection are exact

[23] A common normalization is to center the regressors across assets by replacing $\widehat{\boldsymbol{\beta}}_i$ with $\widehat{\boldsymbol{\beta}}_i - \boldsymbol{m}$. Centering orthogonalizes the intercept and the regressors in sample and can improve numerical stability and exposition. However, centering does not change $\boldsymbol{\Sigma} = \text{Var}_n(\widehat{\boldsymbol{\beta}}_i)$ and therefore does not resolve weak identification when $\boldsymbol{\Sigma}$ is ill-conditioned; it only reparameterizes the regression. In particular, when factor loadings exhibit little CS dispersion (so that $\boldsymbol{\Sigma}$ has small eigenvalues), the unrestricted CSR remains ill-conditioned regardless of centering.

[24] Kleibergen and Zhan (2020) show that once a zero-beta return is included, identification requires the augmented regressor matrix $[\mathbf{1} : \widehat{\boldsymbol{\beta}}]$ to have full column rank; without the zero-beta return, full rank of $[\widehat{\boldsymbol{\beta}}]$ alone is sufficient. In our notation, these conditions are $\boldsymbol{\Sigma} = \boldsymbol{M} - \boldsymbol{m}\boldsymbol{m}' \succ 0$ for the unrestricted CSR and $\boldsymbol{M} \succ 0$ for the restricted CSR. Conditional on $\boldsymbol{M} \succ 0$, the scalar $s = \boldsymbol{m}'\boldsymbol{M}^{-1}\boldsymbol{m}$ summarizes how close the unrestricted case is to failure since, by $\det \boldsymbol{\Sigma} = \det \boldsymbol{M}\,(1 - s)$, s approaching one is equivalent to $\boldsymbol{\Sigma}$ becoming nearly singular relative to $\boldsymbol{M}$ and $[\mathbf{1} : \widehat{\boldsymbol{\beta}}]$ becoming ill-conditioned, with the shift in estimated risk prices magnified by $(1 - s)^{-1}$.

when both specifications use the same first-pass loadings.[25]

The zero-intercept restriction of the excess-return model does not translate into a zero simple-return intercept. It states that the zero-beta rate equals the risk-free rate, and its sample counterpart fixes the simple-return intercept at $\overline{R}_f$ rather than estimating it,

$$\overline{R}_i \;=\; \overline{R}_f \;+\; \widetilde{\boldsymbol{\lambda}}'\widehat{\boldsymbol{\beta}}_i \;+\; \widehat{\nu}_i. \tag{16}$$

Subtracting $\overline{R}_f$ from both sides recovers the restricted CSR in (7), so $\widetilde{\boldsymbol{\lambda}}$ and $\widehat{\nu}_i$ are unchanged as well. Every quantity entering the results of Subsections 2.2–2.4 is therefore numerically identical under the two conventions, and those results carry over after replacing $\widehat{\lambda}_0$ with the zero-beta spread $\widehat{r}^{zb} - \overline{R}_f$. In particular, Lemma 1 and Proposition 3 become

$$\widehat{\boldsymbol{\lambda}} \;=\; \widetilde{\boldsymbol{\lambda}} \;-\; \big(\widehat{r}^{zb} - \overline{R}_f\big)\boldsymbol{u}, \qquad \widehat{r}^{zb} - \overline{R}_f \;=\; \frac{\mathbb{E}_n(\widehat{\nu}_i)}{1-s}.$$

Corollaries 2 and 3, and the risk-price analysis in Subsection 2.3, all of which involve $\widehat{\lambda}_0$, hold under the simple-return convention once $\widehat{\lambda}_0$ is read as the spread. The identification analysis in Subsection 2.4 involves only the regressors and therefore needs no translation. The regressor matrix $\boldsymbol{W} = [\mathbf{1} : \widehat{\boldsymbol{\beta}}]$ is the same under both conventions, and so are s, $\boldsymbol{u}$, and the rank condition in Corollary 7. Fit is equally invariant since $\mathrm{Var}_n(\overline{R}_i) = \mathrm{Var}_n(\mu_i)$ because $\overline{R}_f$ has no CS variation, and the simple-return total pricing error, $\overline{R}_i - \overline{R}_f - \widehat{\boldsymbol{\lambda}}'\widehat{\boldsymbol{\beta}}_i = (\widehat{r}^{zb} - \overline{R}_f) + \widehat{\omega}_i$, coincides with $\widehat{\alpha}_i$ in (9). Consequently, R^2_{ols}, $R^2_{\text{ols}*}$, R^2_{pen}, and the Wedge are identical under the two conventions, and so are $RRMSE$ and $RRMSE_{\text{pen}}$ provided their denominator remains $\mathbb{E}_n(\mu_i^2) = \mathbb{E}_n[(\overline{R}_i - \overline{R}_f)^2]$; replacing it with $\mathbb{E}_n(\overline{R}_i^2)$ would define a different statistic.

Two readings of (15) should be distinguished. If (15) merely rewrites the excess-return model in simple returns, the results of this section are invariant once the intercept is read as the spread $\widehat{r}^{zb} - \overline{R}_f$. Alternatively, the zero-beta rate can be treated as a structural benchmark different to the observed risk-free rate, as when borrowing and lending rates diverge (Brennan (1971)). In this case,

[25] Re-estimating the loadings on simple returns is a different procedure, not a reparameterization. With a common sample and an intercept in each first-pass regression, the simple-return loadings equal their excess-return counterparts plus a vector common to all assets, namely the slope coefficients from projecting the risk-free return on the factors. This common shift leaves the demeaned loadings, $\boldsymbol{\Sigma}$, the rank condition in Corollary 7, and the unrestricted slopes and residuals unchanged, while $\boldsymbol{m}$, s, $\boldsymbol{u}$, the restricted slopes, and the spread $\widehat{r}^{zb} - \overline{R}_f$ can change. The magnitude of those changes is an empirical matter.

the restriction is relaxed rather than reparameterized, and a nonzero spread should be interpreted as a benchmark-rate gap rather than a common pricing error. This reading carries an economic restriction of its own, namely that the estimated spread must be plausible as a benchmark-rate or borrowing–lending spread.[26] Section 6 conducts this evaluation for our estimations. Under either reading, and with the first-pass loadings held fixed, switching to simple returns mitigates neither the collinearity amplification nor the weak identification described in this section, because when factor loadings are nearly constant across assets, the constant vector is nearly spanned under both conventions. The leading case is the market factor, whose betas cluster near one (Ahn, Perez, and Gadarowski (2013); Ahn and Horenstein (2020)) whether estimated from simple or excess returns. Hence, these problems therefore arise in any specification that includes the market, regardless of the return convention.

3 Empirical Background

Having established the theoretical implications of the CS intercept, we now proceed to describe the variables employed in the empirical analysis that follows, including the specific asset pricing models and test portfolios.

3.1 Models

Our choice of models stems from two conditions. First, the easy access to the original factor data from the authors, or alternatively the ability to construct the factors based on publicly available data. Second, the original papers do not report the zero-intercept case or instead report it merely in a secondary and restricted analysis. Our objective is not to consider a comprehensive list of models; instead our aim is to employ a minimally rich family of models for the purpose of this study: to illustrate the quantitative impact of the CS intercept. Moreover, we do not aim at replicating the original studies. Thus, we employ a common sample period to enable a proper comparison among models and factors. When particular models perform notably well or poorly in our common empirical design, we interpret those outcomes primarily through the lens of the intercept mechanism, rather than as definitive judgments about the models in the empirical settings for which they were originally proposed.

[26] The intercept t-ratios in Kan, Robotti, and Shanken (2013) test the null that the zero-beta rate equals the average T-bill rate. They also consider an excess-return specification without a constant, on the grounds that the estimated zero-beta rates exceed the risk-free rate by more than plausible borrowing and lending spreads.

We start by describing the models that employ quarterly factors. First, we consider the 3-factor model proposed by Jagannathan and Wang (1996) (JW3 henceforth), which augments the baseline CAPM with a labor income growth factor (Δy) and the lagged default spread (def, which represents a conditioning variable),

$$\mathbb{E}\left(R_{i,t}\right) = \lambda_{RM}\beta_{i,RM} + \lambda_{Y}\beta_{i,Y} + \lambda_{def}\beta_{i,def},$$

where RM is the excess market return. In this model, the growth in per capita labor income is used as a proxy for the unobserved human capital return. We also consider a restricted version of JW3, where the lagged scaling variable (def_{t-1}) is omitted ($\lambda_{def} = 0$), which represents the human-capital CAPM (JW2 henceforth). The simple CAPM is obtained as a special case of JW2 by setting $\lambda_{Y} = 0$. The fourth model is the 4-factor model of Santos and Veronesi (2006) (SV4),

$$\mathbb{E}\left(R_{i,t}\right) = \lambda_{RM}\beta_{i,RM} + \lambda_{Y}\beta_{i,Y} + \lambda_{RMsw}\beta_{i,RMsw} + \lambda_{Ysw}\beta_{i,Ysw},$$

which augments JW2 with a scaled market factor ($RM_{t}sw_{t-1}$) and a scaled labor income factor ($\Delta y_{t}sw_{t-1}$), where the instrument is the lagged aggregate labor income-to-consumption ratio (sw_{t-1}). The first nested model of SV4 (denoted by SV2) ignores the two labor income factors ($\lambda_{Y} = \lambda_{Ysw} = 0$), whereas the second nested model (SV3) removes the scaled market factor ($\lambda_{RMsw} = 0$). Next, we consider the 2-factor model suggested by Lettau and Ludvigson (2001),

$$\mathbb{E}\left(R_{i,t}\right) = \lambda_{RM}\beta_{i,RM} + \lambda_{RMcay}\beta_{i,RMcay},$$

where the scaled market factor is based on the lagged log consumption-to-wealth ratio (cay_{t-1}). The eighth model is the 2-factor model of Adrian, Etula, and Muir (2014) (AEM2),

$$\mathbb{E}\left(R_{i,t}\right) = \lambda_{RM}\beta_{i,RM} + \lambda_{LEV}\beta_{i,LEV},$$

where LEV denotes the financial intermediaries leverage factor. The nested single-factor leverage model (AEM1) imposes $\lambda_{RM} = 0$. Next, we estimate the 2-factor model (HKM2) of He, Kelly, and

Manela (2017),

$$\mathbb{E}\left(R_{i,t}\right) = \lambda_{RM}\beta_{i,RM} + \lambda_{IC}\beta_{i,IC},$$

where the new factor is intermediary capital (IC). The nested single-factor model (HKM1) assumes $\lambda_{RM} = 0$. Finally, we consider two consumption-based models. Lustig and Van Nieuwerburgh (2005) propose a conditional version of the Consumption-CAPM (LVN3) in which the conditioning variable is the lagged housing collateral ratio (my_{t-1}),

$$\mathbb{E}\left(R_{i,t}\right) = \lambda_{C}\beta_{i,C} + \lambda_{my}\beta_{i,my} + \lambda_{Cmy}\beta_{i,Cmy},$$

and the key factor is scaled consumption growth, $\Delta c_t my_{t-1}$ (with Δc_t representing log aggregate consumption growth). The baseline CCAPM of Lucas (1978) and Breeden (1979) is obtained by setting $\lambda_{my} = \lambda_{Cmy} = 0$. Our sample period is 1970:I to 2017:III, where the starting (ending) quarter is constrained by the availability of data on IC (LEV).

Next, we describe the models estimated at the monthly frequency. The first model is the baseline CAPM. The second model is the 5-factor model of Hahn and Lee (2006) (HL5),

$$\mathbb{E}\left(R_{i,t}\right) = \lambda_{RM}\beta_{i,RM} + \lambda_{\Delta TS}\beta_{i,\Delta TS} + \lambda_{\Delta DS}\beta_{i,\Delta DS} + \lambda_{SMB}\beta_{i,SMB} + \lambda_{HML}\beta_{i,HML},$$

where the non-market factors are the innovations in the term spread (ΔTS) and default spread (ΔDS) in addition to the size (SMB) and value (HML) traded factors of Fama and French (1993). The nested 3-factor model (HL3) is obtained by setting $\lambda_{SMB} = \lambda_{HML} = 0$. The fourth monthly model is the 7-factor model of Petkova (2006) (P7),

$$\mathbb{E}\left(R_{i,t}\right) = \lambda_{RM}\beta_{i,RM}+\lambda_{\Delta TS}\beta_{i,\Delta TS}+\lambda_{\Delta DS}\beta_{i,\Delta DS}+\lambda_{\Delta DY}\beta_{i,\Delta DY}+\lambda_{\Delta TB}\beta_{i,\Delta TB}+\lambda_{SMB}\beta_{i,SMB}+\lambda_{HML}\beta_{i,HML},$$

which represents an extension of HL5 by incorporating the innovations in the aggregate dividend yield (ΔDY) and in the T-bill rate (ΔTB). Petkova (2006) also considers the nested 5-factor model (P5), which arises from ignoring the size and value factors, $\lambda_{SMB} = \lambda_{HML} = 0$. Next, we

estimate the 4-factor model of Liu and Zhang (2008) (LZ4),

$$\mathbb{E}\left(R_{i,t}\right) = \lambda_{RM}\beta_{i,RM} + \lambda_{SMB}\beta_{i,SMB} + \lambda_{HML}\beta_{i,HML} + \lambda_{MP}\beta_{i,MP},$$

which contains industrial production growth (MP) in addition to the three traded factors of Fama and French (1993). The nested single-factor model (LZ1) is obtained by setting $\lambda_{RM} = \lambda_{SMB} = \lambda_{HML} = 0$. The eighth model is the 6-factor model estimated in Chen, Roll, and Ross (1986),

$$\mathbb{E}\left(R_{i,t}\right) = \lambda_{RM}\beta_{i,RM} + \lambda_{MP}\beta_{i,MP} + \lambda_{TS}\beta_{i,TS} + \lambda_{DS}\beta_{i,DS} + \lambda_{UI}\beta_{i,UI} + \lambda_{EI}\beta_{i,EI},$$

which is designated by CRR6 and extends LZ1. In this model, TS (DS) represents the level (rather than the innovation) in the term (default) spread. On the other hand, UI denotes unexpected inflation, whereas EI stands for the expected change in inflation. The nested 5-factor model (CRR5) results from imposing $\lambda_{RM} = 0$. Finally, we estimate the 6-factor model of Fama and French (2018) (FF6),

$$\mathbb{E}\left(R_{i,t}\right) = \lambda_{RM}\beta_{i,RM}+\lambda_{SMB}\beta_{i,SMB}+\lambda_{HML}\beta_{i,HML}+\lambda_{RMW}\beta_{i,RMW}+\lambda_{CMA}\beta_{i,CMA}+\lambda_{UMD}\beta_{i,UMD},$$

where RMW, CMA, and UMD denote, respectively, the profitability, investment, and momentum factors. All factors in this model are traded (i.e., excess equity returns), which implies that the correct approach to estimate it is the standard time-series method (e.g., Fama and French (1993); Cochrane (2005)). Yet, we employ FF6 as a benchmark for the macro models described above, as the model was designed to explain the specific cross-section described below. The sample period in the monthly tests is 1970:01 to 2017:09, in order to preserve consistency with the quarterly estimations.

3.2 Test Assets

The testing assets employed in the asset pricing tests are associated with prominent CAPM anomalies, that is, patterns in CS risk premia left unexplained by the baseline CAPM. Concretely, we consider decile portfolios sorted on four stock characteristics: book-to-market ratio (BM10); investment (INV10); operating profitability (OP10); and price momentum (M10). We do not consider portfolios sorted on market capitalization, given the abundant evidence showing that

the size premium in the U.S. market has declined significantly since the late 1980s. These four portfolio groups are associated with the four non-market factors (ignoring the size factor) in the 6-factor model of Fama and French (2018), which represents one of the current "horse-race" equity factor models in the literature. Hence, these portfolios can be interpreted as a "minimum" cross-section to judge the pricing performance of the different factor models described above. These test assets have also been a popular choice in the literature in recent years and are sufficient for our purposes, illustrating the quantitative impact of the CS intercept on pricing errors and model fit. The data on all monthly portfolio returns are retrieved from Kenneth French's data library. Monthly excess portfolio returns are obtained by subtracting the monthly T-bill return, also available from French's library. Quarterly portfolio returns are obtained by compounding the corresponding monthly returns. Quarterly excess returns are then formed by subtracting a quarterly benchmark risk-free return constructed from the monthly 3-month T-bill rate from the St. Louis Fed (FRED). Specifically, TBQ_t denotes the annualized 3-month T-bill rate (in percent per annum) observed in the first month of quarter t, and we set the quarterly risk-free return as $TBQ_t/4$.[27]

The IA provides the CS summary statistics for both quarterly and monthly portfolio risk premia (i.e., average excess portfolio returns). The average risk premium across all 40 portfolios is 1.85% per quarter and 0.59% per month. These estimates are close to the sample means of the market factor (1.72% per quarter and 0.54% per month). The momentum deciles exhibit the largest CS dispersion among the four portfolio groups by a wide margin, which helps explain why the strongest restricted-CSR fit appears in that group, as documented in the next section.[28]

4 Empirical Analysis: Pricing Errors and Performance Metrics

In this section, we evaluate the impact of the CS intercept on the pricing performance of the various factor models described in the last section.

[27] Thus, $TBQ_t/4$ should be interpreted as a simple approximation to the one-quarter holding-period return on a 3-month bill purchased at the beginning of the quarter, rather than as a compounded sequence of monthly short rates.

[28] The minimum values of risk premia within INV10, OP10, and BM10 are all positive, thus showing that all portfolios in these groups have positive average excess returns. The exception holds for the momentum deciles, as the extreme loser portfolio (first momentum decile) has a negative premium (-0.33% per quarter).

4.1 Quarterly Models

To test the null of a zero fit (against the alternative of a positive fit), we follow Maio and Min (2026) in computing empirical p-values based on a bootstrap simulation, with the details being presented in the IA.[29] The bootstrap distribution is likely a better approximation to finite samples than an asymptotic distribution given the relatively small T at the quarterly frequency.[30] The empirical p-value for a given explanatory ratio (R^2_{pen}, R^2_{ols}, or $R^2_{\text{ols}*}$) represents the fraction of artificial samples where the (pseudo) R^2 is higher than the corresponding sample R^2.[31] The return generating process assumes that the model has no pricing power, that is, all factor risk prices are constrained to zero. The IA contains an analysis of the power and size of the bootstrap-based tests for the three measures of fit.

We present results for both the single-anomaly estimations as well as the augmented estimation (denoted by All) including all four portfolio groups (for a total of 40 portfolios). As discussed in Maio (2024), while the augmented estimation is the most challenging one, because it assesses the capacity of a given model to explain all portfolio groups on a joint basis, it can place substantial weight on inter-group dispersion.[32] Hence, it is also important to provide the asset pricing results for the individual portfolio groups or anomalies, something that is also in line with most of the literature. Furthermore, and following Cooper, Ma, and Maio (2022), in addition to the estimation with the 40 portfolios, we conduct an augmented estimation (designated by All*) that includes only the extreme first two and last two deciles within each portfolio group in addition to the market

[29] More precisely, the null does not represent zero fit in the case of $R^2_{\text{ols}*}$, but rather no explanatory power (i.e., zero risk prices).

[30] As a robustness check, the IA derives fixed-n, large-T inference for the null $R^2_{\text{pen}} = 0$. This null is an interior restriction when the population intercept is non-zero ($\lambda_0 \neq 0$). Under the maintained time-series CLT, differentiability, and nondegeneracy condition that the delta-method variance of $\sqrt{T}\widehat{R}^2_{\text{pen}}$ is strictly positive, the corresponding delta-method statistic (computed with a HAC estimate of the long-run covariance matrix) is asymptotically standard normal under the null. We nonetheless rely on the bootstrap for the paper's finite-sample inference. Beyond the small-sample concern, especially relevant at the quarterly frequency, the asymptotic approximation appears to have finite-sample power close to *zero* for R^2_{pen} in the designs considered. A likely reason is that when s is close to one, the near-collinearity between the constant and the beta span makes the intercept and risk price estimates highly sensitive to small perturbations in the estimated moments, which can substantially increase the delta-method variance of $\widehat{R}^2_{\text{pen}}$. Thus, the weak finite-sample performance of the asymptotic R^2_{pen}-test is another symptom of the collinearity-amplification channel.

[31] The empirical p-values associated with either R^2_{pen} or $R^2_{\text{ols}*}$ are not reported when these sample metrics assume negative values. The reason is that the one-sided positive-fit test is not relevant in those cases.

[32] A given model might then produce a large fit by matching inter-group dispersion in risk premia (e.g., the difference in risk premia between the first momentum decile and the last BM decile) rather than the economically relevant intra-group dispersion in risk premia (e.g., the difference in risk premia between the first and the last BM deciles).

factor (for a total of 17 portfolios). The choice of removing the intermediate deciles stems from the abundant evidence showing that most of the CS dispersion in risk premia is associated with the "extreme portfolios" within a given portfolio sort. Thus, such a choice can enable a more powerful asset pricing test.

The results for the performance evaluation metrics associated with the quarterly models are shown in Tables 1 and 2. We have a total of 78 estimations, which stems from 13 models and six different cross-sections. When the intercept is restricted to zero, there are several models that mimic the baseline CAPM in terms of producing a negative $R^2_{\text{ols}*}$, which implies that these models do worse than a constant when it comes to pricing the cross-section. Concretely, there are 40 such cases, which represents about half of the 78 estimations being conducted. This pattern is especially evident in the estimations with INV10 or OP10, as 10 models (out of the 13 within each of these two portfolio groups) produce a negative explanatory ratio. In the cases of the two potentially more challenging cross-sections (All and All*), there are five models with a negative $R^2_{\text{ols}*}$ in both settings (SV2, LL2, HKM1, HKM2, in addition to the baseline CAPM). If we switch attention to the cases in which the sample $R^2_{\text{ols}*}$ is positive, there are several models failing to achieve statistically significant positive fit, as the null of a zero $R^2_{\text{ols}*}$ is not rejected (at the 5% level). In fact, an insignificant explanatory ratio actually holds for models with a seemingly reasonable sample fit (explanatory ratios around or above 40%), such as JW3 (in the estimations with OP10 and All*), AEM1 (INV10), HKM2 (INV10), and SV3 and SV4 (both in the estimation with All*). These results reveal the importance of accounting for the sampling error associated with the CS R^2.

When the intercept is added to the second-pass CSR, at first sight one would be tempted to conclude for a largely positive pricing performance across the board, given that the estimates of R^2_{ols} assume quite high magnitudes in many cases. Yet, the null of a zero fit is not rejected (at the 5% level) in 35 of the 78 estimations (which means about 45% of the cases), including several specifications where the sample explanatory ratio is around or above 50%. This pattern is especially notable in the estimations with INV10 or OP10, where there are only three models (in each portfolio group) for which the null is rejected. Again, these results illustrate the importance of accounting for the estimation error in the sample explanatory ratio. On the other hand, the cases of CAPM and LL2 are noteworthy: the fit moves from being negative ($R^2_{\text{ols}*} < 0$) to values above 30% (with significance in both cases) by estimating the CSR with an intercept over the full cross-section.

This pattern becomes even more accentuated in the estimation with the 17 portfolios, as the R^2_{ols} associated with those two models register larger magnitudes (relative to the estimation with All), with both SV2 and HKM2 also exhibiting a similar relative performance (i.e., a large gap between R^2_{ols} and $R^2_{\text{ols}*}$). These results illustrate the strong quantitative effect of relaxing the zero-intercept restriction and using R^2_{ols} as a performance metric. In particular, the baseline CAPM produces a large and statistically significant R^2_{ols} for four of the six cross-sections (OP10, M10, All, and All*), which reflects the mechanical inflation documented in Proposition 4: once the intercept absorbs the common pricing offset, R^2_{ols} can rise materially even when the total pricing errors remain large. Because $R^2_{\text{ols}} = R^2_{\text{pen}} + \text{Wedge}$, a large wedge can leave R^2_{pen} negative, so standard goodness-of-fit metric is not a measure of total pricing performance.

When one uses the penalized metric (R^2_{pen}), the overall picture is substantially different, as nearly all models produce negative explanatory ratios (often with very large magnitudes). The only two exceptions (which represent less than 3% of the 78 estimations) refer to AEM1 (in the estimation with the momentum deciles) and AEM2 (BM10), where the sample R^2_{pen} is positive. Therefore, in the two more challenging estimations (All and All*), there is not a single model that delivers a positive penalized R^2, and the same holds in the estimations with the investment and profitability deciles. Such a pattern compares with a scenario in which the R^2_{ols} is significantly positive in the majority of the estimations associated with All and All*. Particularly striking are the estimations over All*, where nearly all models (with the exceptions of SV4 and HKM1) produce simultaneously an economically large R^2_{ols} (i.e., around or higher than 50%) jointly with a negative R^2_{pen}. A similar pattern applies to JW2, JW3, SV3, SV4, AEM1, AEM2, and LVN3 in the estimation with All. These results document the extent to which R^2_{ols} overstates pricing performance when intercept estimates are large.

The magnitude of the intercept's contribution to the pricing performance can also be inferred from the sample relative composite pricing errors. By using the incorrect definition of pricing errors (i.e., by ignoring the intercept and considering exclusively the CSR residuals), one obtains composite pricing errors that represent at most 36% of the raw composite risk premia, as indicated by the estimates of $RRMSE$. In fact, there are only three cases among the 78 estimations (i.e., 3.8%), in which $RRMSE$ is above 30% (CAPM and AEM1 both in the estimation with M10 and HKM1 in the estimation with All*). For several models, the $RRMSE$ estimates are clearly below

the counterparts associated with the restricted CSR ($RRMSE*$), thus suggesting a reasonably good pricing performance. However, when one relies on $RRMSE_{\text{pen}}$, which imposes the restriction that the intercept represents the average pricing error, the assessment changes materially: the composite pricing errors exceed the composite risk premia that one seeks to explain (i.e., $RRMSE_{\text{pen}}$ above one) in the majority of the cases (more precisely, 54 out of 78 or nearly 70% of the total estimations). To have an even better idea, it turns out that only in eight estimations (i.e., around 10% of the total) do we obtain a $RRMSE_{\text{pen}}$ below 50%: AEM1 (estimations with INV10 and M10), AEM2 (INV10 and BM10), HKM1 (BM10), LVN3 (BM10), and CCAPM (All and All*). Even in those cases, we observe that $RRMSE_{\text{pen}}$ remains clearly above the corresponding $RRMSE$, reflecting the corresponding large intercept estimates; the main exception being AEM2 with BM10, where the very small intercept (0.06%) leaves $RRMSE_{\text{pen}}$ essentially at $RRMSE$, consistent with Proposition 5. In sum, the picture provided by $RRMSE_{\text{pen}}$ is qualitatively similar to that provided by R^2_{pen}, that is, the intercept estimates contribute the bulk of the total pricing errors, so standard residual-based metrics substantially understate total pricing error when the intercept is economically relevant.

Consistent with this argument, and as shown in Table 4 below, the estimates of λ_0 for most models are above 2% per quarter in the estimation with All. These intercept magnitudes are economically large and are consistent with a sizeable common pricing offset in the unrestricted CSR. One legitimate interpretation, developed in Propositions 1 and 2, is that omitted risk factors contribute materially to that offset. Indeed, the three models producing intercept estimates below 2% (AEM1, CCAPM, and LVN3) are still associated with economically relevant average pricing errors (i.e., λ_0 varying between 0.46% and 1.20% per quarter). Therefore, Wedge (which equals minus the lower bound for R^2_{pen} based on Corollary 1) is above one in nearly all cases, often with very large magnitudes. The very few exceptions are AEM1 (estimation with M10 and All*), AEM2 (BM10), and CCAPM (All and All* portfolios). Yet, three of these five cases still point to a large Wedge (between 0.67 and 0.99). By Corollary 1, these estimates imply that the upper bound for R^2_{pen} is negative in 73 of the 78 specifications (i.e., around 94%), a finding that is especially notable. Additionally, those estimates drive a large decoupling between the penalized and conventional fit measures (i.e., R^2_{pen} falls well below the corresponding R^2_{ols}), consistent with Proposition 4 and Corollary 1. The very high Wedge also explains why $RRMSE_{\text{pen}}$ substantially exceeds $RRMSE$ in nearly all 78 estimations (consistent with Proposition 5). The above mentioned

case of AEM2 represents the main outlier to those patterns, as Wedge assumes a value around zero in the estimation with BM10 (0.02). This explains why R^2_{pen} is very similar to both R^2_{ols} and $R^2_{\text{ols}*}$ (81% versus 83%), something that is consistent with Proposition 4 and Corollary 2.

The scalar s also plays an important role in explaining the discrepancy between the (appropriate) fit among the two types of CSR ($R^2_{\text{ols}*} - R^2_{\text{pen}}$), as described in Corollary 2. The estimates are very close to one in nearly all cases, thus justifying why we obtain R^2_{pen} that are substantially below the $R^2_{\text{ols}*}$ counterparts (i.e., more negative or negative versus positive). Hence, the large s reinforce the large values of Wedge (arising from large intercepts, which in turn can emerge from large s) in producing a considerable discrepancy in fit between the two CSRs. In other words, the same near-collinearity (i.e., small dispersion in betas relative to the respective magnitudes) widens the fit gap through two reinforcing channels rather than one. In fact, we obtain a s below 0.90 only in four cases (i.e., around 5% of the estimations), all being associated with AEM1 (estimation with M10, BM10, All, and All*). This helps explaining why the R^2_{pen} for AEM1 in the first of these estimations is significantly positive and not very far apart the corresponding $R^2_{\text{ols}*}$ (16% versus 46%). The large values of Wedge in the remaining three cases, especially in the estimations with BM10 and All, more than offset the smaller contribution from s, thus producing a large gap between $R^2_{\text{ols}*}$ and R^2_{pen}. The case of AEM2 (over BM10) mentioned above represents an opposite scenario, as the very large s estimate (0.99) does not originate a decoupling in the fit measures due to the tiny value of Wedge.

A key takeaway from the results in Tables 1 and 2 is that performance metrics counting the intercept as pricing error often lead to qualitative conclusions similar to those from the corresponding restricted CSR. In particular, $R^2_{\text{ols}*}$ and R^2_{pen} frequently point in the same direction about which models receive weak support in this empirical design. However, occasionally, employing correct metrics under the unrestricted CSR (R^2_{pen} or $RRMSE_{\text{pen}}$) can produce a type of "overshooting", as shown in Corollary 3. Overshooting occurs when $R^2_{\text{pen}} < 0 < R^2_{\text{ols}*}$. The equivalent characterization from Corollary 3,

$$0 < R^2_{\text{ols}*} < \frac{s}{(1-s)^2} \frac{\mathbb{E}_n(\widehat{\nu}_i)^2}{\text{Var}_n(\mu_i)},$$

is useful in practice: it shows that a model can keep its total pricing error below the CS dispersion of risk premia under the excess return normalization (i.e., $R^2_{\text{ols}*} > 0$) and still receive a negative

R^2_{pen} when a nonzero mean pricing error in the restricted CSR is amplified by a large value of s. Accordingly, when the results from the unrestricted CSR are reported, the triplet $\left(R^2_{\text{ols}*}, s, \mathbb{E}_n(\widehat{\nu}_i)\right)$ indicates whether a negative R^2_{pen} mainly reflects poor fit (i.e., $R^2_{\text{pen}} < R^2_{\text{ols}*} < 0$) or a non-zero common pricing offset magnified by near-collinearity ($R^2_{\text{pen}} < 0 < R^2_{\text{ols}*}$). Fundamentally, these two scenarios are largely distinct. LVN3 estimated on the 17-portfolio cross-section represents an example of such a scenario. The restricted CSR yields $R^2_{\text{ols}*} = 0.72$, which is statistically significant at the 1% level. At the same time, despite the quite small $\mathbb{E}_n(\widehat{\nu}_i)$ of 0.0242% per quarter, s is very close to one (0.9829). Corollary 3 therefore predicts a large fit gap (i.e., overshooting), and the unrestricted CSR indeed delivers a very negative explanatory ratio ($R^2_{\text{pen}} = -2.10$). Relatedly, the penalized relative composite pricing error of 73% is well above the corresponding metric from the restricted CSR ($RRMSE*$ of 22%). Other examples of such a decoupling are AEM2 (over INV10), LL2 and HKM2 (both in the estimation with BM10), and several models in the estimation with the momentum deciles (JW2, JW3, SV3, SV4, and LVN3): in all of these cases, we detect both economic (estimates above 70%) and statistical significance for $R^2_{\text{ols}*}$, while the corresponding R^2_{pen} being negative.

4.2 Monthly Models

Next, we shift attention to the monthly models, whose results appear in Table 3. Given that some of these models contain as many as six or seven factors, we restrict the analysis to the augmented cross-sections (All and All*) to avoid overfitting. Overall, the results are not very different from the quarterly counterparts on a qualitative basis. Specifically, and excluding the special case of FF6 (which is discussed below), none of the models produces a significantly positive $R^2_{\text{ols}*}$ in the estimation with all portfolios. The picture is marginally improved in the cross-section with 17 portfolios, as the null (of a zero explanatory ratio) is rejected (at the 5% level) for both HL5 and P7.[33]

The pricing performance appears substantially more favorable when the intercept is unrestricted, as most of the 18 estimations (i.e., excluding the two estimations of FF6) are associated with a significantly positive R^2_{ols}. The few exceptions are associated with LZ1, where the baseline explanatory ratios (13% and 41% for All and All*, respectively) are not significant at the 5% level.

[33] Naturally, the good performance of both HL5 and P7 for All* stems from the presence of the non-market traded factors of Fama and French (1993) (SMB, and especially, HML).

However, such evidence is illusory and stems from the very large intercept estimates. In fact, the Wedge estimates are even more extreme than the quarterly counterparts in Table 2, ranging from 10 (LZ1) to 203 (LZ4) over All, and from 7 (LZ1) to 152 (P5) over All*. Consequently, both the sample R^2_{pen}, as well as the respective upper bounds, are extremely negative in all cases, in line with Proposition 4 and Corollary 1. Similar conclusions can be drawn when we rely on the relative composite pricing error. Indeed, the estimates of $RRMSE_{\text{pen}}$ are above 100% in all 18 monthly estimations and actually above 200% in 14 of these estimations. In comparison, the corresponding estimates of $RRMSE$ are below 30% for 16 of the 18 specifications.

The case of FF6 deserves a separate discussion, as it represents an even more extreme example of overshooting than the case of LVN3 discussed above. The 6-factor models produces a large $R^2_{\text{ols}*}$ of 80% over All (with significance at the 1% level), given that the factors in the model are mechanically correlated (i.e., by construction) with the test portfolios. At the same time, the model generates a very negative R^2_{pen} (-17), which suggests a clear empirical rejection. In a similar vein, $RRMSE_{\text{pen}}$ is above 100% (1.33), compared to only 14% for both $RRMSE$ and $RRMSE*$. By Corollaries 2 and 3 and Proposition 5, the explanation for such a substantial decoupling hinges on the very extreme value of s (0.9993), which leads to a large intercept (around 0.80% per month), and thus a very large Wedge (around 18), even though the average pricing error in the restricted CSR is very close to zero (0.0006% per month). Recall that such an effect comes in addition to the direct effect of s in terms of widening the gap $R^2_{\text{ols}*} - R^2_{\text{pen}}$. In the estimation with All*, the situation is even more pronounced: the fit in the restricted CSR is as large as 90% (again with significance at the 1% level), given that the middle deciles (which typically display little CS dispersion in risk premia) are removed from the cross-section. However, the penalized fit is twice as negative as in the estimation with All (-35 versus -17). Consistent with such a decoupling, it turns out that $RRMSE_{\text{pen}}$ is not far from 300% (2.71) compared to a $RRMSE*$ close to zero (14%). Again, despite the negligible $\mathbb{E}_n(\widehat{\nu}_i)$ of 0.0007% per month, the s in the close neighborhood of one (0.9996) produces an intercept as large as 1.73% per month, which in turns leads to a Wedge of 36. What makes this example informative is not the magnitude of Wedge (most monthly models produce even more extreme estimates of this quantity), but the role of s: very low CS dispersion in betas relative to the respective average magnitude for (at least some of) the factors in the model completely shifts the qualitative assessment, by transforming a very large fit into an extreme empirical rejection. This

pattern represents the fit-metric manifestation of the collinearity amplification channel.

P7 and HL5 (both over All*) represent other relevant cases of overshooting, as adding the intercept transforms $R^2_{\text{ols}*}$ of 84% and 68% (both significant at the 5% or better significance level) into extremely negative R^2_{pen} (-40 and -77, respectively). Almost equivalently, the intercept shifts $RRMSE*$ of respectively 18% and 26% into $RRMSE_{\text{pen}}$ of nearly 300% and 400%, respectively. Again, the key driving force is s, with estimates around one (0.9992 and 0.9976, respectively), which clearly outweigh the very small $\mathbb{E}_n(\widehat{\nu}_i)$ in both cases (0.0015% and 0.0061% per month, respectively).

In sum, while R^2_{pen} very often aligns with $R^2_{\text{ols}*}$, as discussed above and in the previous subsection, it can also be highly misleading in a few cases. The implication is that the unrestricted CSR, even if appropriately treating the intercept as the average pricing error, is not a perfect substitute to the restricted counterpart when it comes to assessing linear factor models. Consequently, we view the restricted CSR (in which the restriction that the zero-beta rate equals the risk-free rate is imposed ex ante) and $R^2_{\text{ols}*}$ (as well as $RRMSE*$) as the appropriate primary tools when evaluating macro factor models. When the unrestricted CSR is also reported, R^2_{pen} can be informative about the model's total pricing-error burden (often pointing toward rejection of weak models), but it should be read jointly with s and $\mathbb{E}_n(\widehat{\nu}_i)$ to avoid the kind of trap dissected above.

Finally, LZ1 represents a peculiar case: the estimates of s (0.22 and 0.50 in the estimations with All and All*, respectively) are notable outliers, as all other 18 estimates are around or above 0.96. Following Proposition 3, this implies lower intercept estimates, and thus lower values of Wedge, as described above. Consequently, the positive gaps in fit, $R^2_{\text{ols}} - R^2_{\text{pen}}$ and $R^2_{\text{ols}*} - R^2_{\text{pen}}$, are not as extreme as with the remaining monthly models, in line with Proposition 4 and Corollary 2. In other words, R^2_{pen} aligns with $R^2_{\text{ols}*}$ in establishing an empirical rejection of that model.

5 Empirical Analysis: Risk Price Estimates and Identification

Beyond pricing errors and model fit, the CS intercept shapes risk price estimates and their identification. This section examines how the intercept alters the magnitude, sign, and statistical significance of estimated factor risk premia.

5.1 Quarterly Risk Price Estimates

The quarterly risk price estimates for both types of CSR are reported in Table 4. To keep the focus and save space, we concentrate the analysis on the more relevant estimation with all 40 portfolios. To assess the statistical significance of the risk prices, we employ the popular Shanken standard errors (which correct for the error-in-variables in the first-pass betas, as shown in Shanken (1992)).

Several patterns emerge from Table 4. First, we observe that most of the intercept estimates are statistically significant (at the 5% level). The few exceptions hold for AEM1, CCAPM, and LVN3, in which cases the estimates of λ_0 are insignificant even at 10% level. Second, a notable finding is that none of the nine market risk price estimates remains positive in the unrestricted CSRs, something that compares with significantly positive estimates in all the corresponding restricted CSRs. Indeed, some of the negative estimates of λ_M (CAPM, SV2, LL2, and HKM2) are statistically significant.[34] Third, there are other relevant cases of sign reversal. Concretely, the estimates of λ_{RMsw} (within SV2) and λ_{IC} (within both HKM1 and HKM2) move from being significantly positive in the restricted CSR to negative (albeit insignificant) estimates in the corresponding unrestricted CSR. Fourth, another pattern from Table 4 is that the magnitudes of the risk price estimates (that preserve sign) are reduced when the CSR includes the intercept, and this holds for both positive and negative estimates. In several cases, the decline in magnitude is substantial (by 50% or more). The few exceptions to such a pattern hold for λ_{Ysw} (within SV3 and SV4) and λ_{my}, in which cases the magnitude either does not change or increases very marginally by including the intercept. These results show that there is a robust pattern of attenuation in the risk prices under the unrestricted CSR.

Those patterns (sign flip and attenuation in magnitude) are not surprising and are consistent with the theoretical analysis in Section 2, confirming the quantitative importance of the intercept for risk price estimates. As shown in Proposition 6 and Corollary 5, the risk price shift satisfies $\widehat{\boldsymbol{\lambda}} - \widetilde{\boldsymbol{\lambda}} = -\widehat{\lambda}_0 \boldsymbol{u}$, or more transparently $\widehat{\lambda}_j - \widetilde{\lambda}_j = -\widehat{\lambda}_0 u_j$. Since Table 4 shows $\mathbb{E}_n(\widehat{\nu}_i) > 0$ for all 13 models, Proposition 3 implies $\widehat{\lambda}_0 > 0$ throughout. The empirical implication is therefore

[34] We obtain similar qualitative statistical inference for the market risk prices by relying on more robust standard errors that account for model misspecification (see Kan, Robotti, and Shanken (2013)). Results are available upon request.

not a uniformly negative componentwise gap (i.e., $\widehat{\lambda}_j - \widetilde{\lambda}_j < 0$), but rather a shift whose sign is determined by u_j and whose magnitude scales with $\widehat{\lambda}_0|u_j| = \mathbb{E}_n(\widehat{\nu}_i)\,|u_j|/(1-s)$. This often produces attenuation and, when the shift is large enough, sign reversals.[35]

Importantly, the gap in risk prices (across the two CSRs) is magnified when s is very high, something that applies to most models under consideration, as discussed in the previous section. This is especially relevant for the models including the market factor, as their s estimates are always very close to one (in a tight range of 0.98-0.99), implying large λ_0 and thus material changes in risk prices (often with sign flip). Recall that a high s stems from a high (cross-sectionally) average beta and/or a low CS dispersion in betas. This is exactly what typically holds for the market factor when pricing equity portfolios: the average market beta tends to be positive with a large magnitude, while there is little CS dispersion in the market betas. Consistent with this description, there is only one case of a sign flip (HKM1) among the four models that do not contain the market factor.

To illustrate the amplification mechanism in a familiar case, consider the baseline CAPM. The restricted CSR yields $\widetilde{\lambda}_{RM} = 1.76\%$ per quarter, while the average portfolio risk premium is $\overline{\mu} = 1.85\%$ (as reported in the IA). Because $m = \mathbb{E}_n(\widehat{\beta}_{i,RM})$ is basically one (1.0049), the restricted-CSR mean pricing error, $\mathbb{E}_n(\widehat{\nu}_i) = \overline{\mu} - m\widetilde{\lambda}_{RM}$, is small but positive (0.09%). With $s \approx 0.98$, Eq. (8) implies $\widehat{\lambda}_0 = 0.0860\%/(1 - 0.9785) \approx 3.99\%$. Since $u = m/M = 0.97$ is also close to one in the single-factor case, the induced risk price shift is $\widehat{\lambda}_{RM} - \widetilde{\lambda}_{RM} = -u\widehat{\lambda}_0 = -3.99\% \times 0.9737 \approx -3.89\%$, so the unrestricted estimate falls to approximately -2.13%. The sign reversal therefore reflects the amplification channel in this specific case: even a modestly positive mean pricing error, when combined with near-collinearity between the market-beta vector and the constant, generates a large intercept and a large downward shift in the risk price, in line with Propositions 3 and 6 and Corollary 6. Given that the market factor is pervasive in the asset pricing literature (most linear multifactor models, including several macro models, include it), the evidence underscores the sensitivity of unrestricted CSR estimates to the collinearity amplification channel. While the pricing-error distortion can be corrected by including the intercept in the definition of total pricing error, as discussed in the last section, the risk price distortion has no analogous remedy. Notably,

[35] By Corollary 6, in the single-factor case ($K = 1$) we have $u = m/M$. Hence, with $\widetilde{\lambda} > 0$, if $m > 0$ (equivalently, $u > 0$) and $\mathbb{E}_n(\widehat{\nu}_i) > 0$ (equivalently, $\widehat{\lambda}_0 > 0$), then $\widehat{\lambda} - \widetilde{\lambda} < 0$, implying attenuation and, if the shift is large enough, sign reversal. Corollary 6 also implies that a sign flip is more likely to occur when s is high.

a significantly negative (or even an insignificantly positive) market price of risk determines an economic and statistical violation of either the baseline CAPM or a multifactor model that contains the market factor. The practical implication is the same as in the last section: the unrestricted CSR is best viewed as a supplementary diagnostic, with the restricted CSR serving as the primary tool for CS asset pricing tests.

AEM1 also represents an interesting case. Its restricted-CSR mean pricing error is the largest in Table 4 ($\mathbb{E}_n(\widehat{\nu}_i) = 0.15\%$), yet the corresponding intercept estimate (0.95% per quarter) is smaller than in most other models (the sole exception being the CCAPM) because the collinearity amplification channel is less extreme here: with $s = 0.8394$, Eq. (8) implies $\widehat{\lambda}_0 = 0.1526\%/(1 - 0.8394) \approx 0.95\%$. Thus, s is not offsetting the positive mean pricing error; rather, relative to the other models (with s much closer to one), AEM1 is subject to considerably less collinearity amplification. Even so, the positive $\mathbb{E}_n(\widehat{\nu}_i)$ together with a large value of $u = m/M \approx 10.32$ in the single-factor model still generates a substantial decline in the LEV risk price, $11.07\% - 20.87\% \approx -9.80\% = -0.95\% \times 10.32$, in line with Proposition 6 and Corollaries 5 and 6.

The IA contains the risk price estimates when we estimate the various models over the cross-section of 17 portfolios. The evidence is qualitatively similar to that in Table 4. Concretely, we find that apart from the cases of AEM1 and CCAPM, the intercept estimates assume very large magnitudes, ranging from 1.41% (LVN3) to 6.16% per quarter (SV2). Moreover, all the market risk prices become negative (with statistical significance in several cases) in the unrestricted CSRs, while the magnitudes of the risk prices (that preserve sign) tend to be lower under the unrestricted CSRs. Importantly, models that produce larger average pricing errors (in the restricted CSRs) tend to be associated with larger gaps in risk prices (often with sign flip), namely the already mentioned cases of the CAPM, SV2, LL2, HKM1, and HKM2, consistent with Propositions 3 and 6. One interesting feature is that the intercept estimate is negative (-0.78%) in the case of CCAPM, which represents an outlier among all 26 estimates (from both augmented cross-sections). Naturally, this stems from the negative $\mathbb{E}_n(\widehat{\nu}_i)$ (-0.04%). On the other hand, AEM1 again stands out: despite the relatively large average pricing error from the restricted CSR (0.24% per quarter), the intercept is "only" 0.83% because s is quite low (0.71) on a relative basis. Even so, the large estimate of u (9.45) generates a substantial attenuation in the corresponding risk price estimate

$(13.18\% - 21.03\% \approx -7.85\% = -0.83\% \times 9.45)$, following Proposition 6 and Corollaries 5 and 6.[36]

5.2 Monthly Risk Price Estimates

The monthly risk price estimates are displayed in Table 5. The qualitative picture is consistent with that observed in Table 4. Concretely, the intercept estimates assume even more extreme magnitudes than the quarterly counterparts, varying between 0.63% per month (about 1.89% per quarter) for LZ1 and 2.78% per month (8.34% per quarter) for LZ4. Most of the monthly estimates are statistically significant, with FF6 representing the sole exception (albeit with significance at the 10% level). Indeed, half of the 10 models have an intercept above 2% per month (or roughly 6% per quarter): HL3, HL5, P5, P7, and LZ4.

Turning the attention to the λs, it turns out that the market prices of risk shift from being significantly positive in the restricted CSR in all eight cases to either significantly negative (the majority of the cases) or insignificantly negative (FF6 and CRR6). Other relevant cases of sign shift are the risk prices associated with ΔDS (within HL3), ΔDY (in both P5 and P7), ΔTB (within P5), SMB (in LZ4) and UI (within CRR6). On the other hand, there are several cases of attenuation in the magnitude (with preserved sign) that leads to insignificant risk prices in the unrestricted CSR: ΔTS (within HL5 and P7), HML (for HL5, P7, and LZ4), MP (in LZ4), and UI (within CRR5).

The case of LZ1 deserves some discussion: the restricted average pricing error is as large as 0.49% per month, representing a major outlier among the 10 models. Yet, the corresponding intercept estimate of 0.63% is the smallest among the monthly models. The reason hinges on the small amplification caused by the unusual low value of s (0.22), already noted in the last section. Nonetheless, the relatively large magnitude of $u = m/M \approx -4$ produces a relevant sign flip, with the risk price shifting from -1.83% (with no significance) to a significant positive estimate of 0.68% in the unrestricted CSR: $0.68\% - (-1.83\%) \approx 2.51\% = -0.63\% \times (-3.9985)$.

Another relevant finding from Table 5 is the strong attenuation effect of the intercept on the risk price estimates for HML. Thus, the estimates from the restricted CSR are much closer to the "correct" estimate (0.37%, which represents the sample mean of HML) than the unrestricted

[36] The analysis of the risk price estimates from both CSRs when the individual portfolio groups are the test assets is also included in the IA. The results are qualitatively consistent with those associated with both All and All*, albeit there are a few differences. For example, both AEM2 and HKM1 (in the estimation with BM10) have small gaps in risk prices due to small $\widehat{\lambda}_0$ (in magnitude).

counterparts. In fact, $\widehat{\lambda}_{HML}$ are insignificant in nearly all cases, the sole exception holding for FF6. This represents another clear example of the highly distorting effect provoked by the intercept.

The IA contains the monthly risk price estimates associated with All*. It turns out that most of the estimates of λ_0 are even higher than the counterparts associated with All (with P7 representing the sole exception), varying between 0.75% (LZ1) and 3.55% (P5). Furthermore, the pattern in the market prices of risk is very similar to that obtained in the estimation with all 40 portfolios.

5.3 Risk Price Identification

In this subsection, we study whether the factor risk prices are identified in the second-pass CSRs. Following Burnside (2011) and Kleibergen and Zhan (2020), rank failures in the factor-loading matrix make individual factor risk prices unidentified; near-rank failures make the corresponding estimates and inference unreliable. As established in Subsection 2.4, identification in the restricted CSR (where the intercept is set to zero) requires the $n \times K$ loading matrix used in the second pass, $\widehat{\boldsymbol{\beta}}$, to have full column rank, while in the unrestricted CSR the identified coefficient vector is $(\lambda_0, \boldsymbol{\lambda}')'$ and the relevant condition is full column rank of the augmented matrix $\widehat{\boldsymbol{W}} = [\mathbf{1} : \widehat{\boldsymbol{\beta}}]$. Equivalently, $\text{rank}(\boldsymbol{M}_1\widehat{\boldsymbol{\beta}}) = K$, with $\boldsymbol{M}_1 = \boldsymbol{I}_n - n^{-1}\mathbf{1}\mathbf{1}'$ (Corollary 7). We implement the rank tests of Kleibergen and Zhan (2020) for the quarterly factor models. The null hypothesis is reduced rank. More precisely, for the restricted CSR, reduced rank of $\widehat{\boldsymbol{\beta}}$; for the unrestricted CSR, reduced rank of $\widehat{\boldsymbol{W}}$. When the intercept is excluded, the statistic admits an $F(n+1-K, T-n-1)$ bounding distribution; when the intercept is included it admits an $F(n-K, T-n)$ bounding distribution.

Table 6 reports the resulting p-values. The p-value before the comma corresponds to the unrestricted CSR, and the p-value after the comma corresponds to the restricted CSR. Focusing first on the unrestricted CSR, the reduced-rank null is rejected at the 5% level in 27 of the 78 specifications. The rejections are concentrated in CAPM and HKM1 (all six cross-sections), HKM2 (all cross-sections except M10), SV2 (OP10, All, and All*), LL2 (OP10, M10, All, and All*), AEM1 (BM10 and All*), and AEM2 (All*). Hence, in most unrestricted specifications the rank test does not reject reduced rank.

We do not interpret a non-rejection as proof of exact rank deficiency, which is a knife-edge condition. Rather, a non-rejection is informative as a diagnostic of weak identification when it is accompanied by evidence that the second-pass design matrix is nearly collinear. The collinearity

index s in Tables 1–2 provides this additional diagnostic. Recall from Section 2 that s is the fitted R^2 from projecting the constant vector on the loading matrix, so values of s close to one indicate severe near-collinearity between the constant and the columns of $\widehat{\boldsymbol{\beta}}$. In most cases in which the unrestricted rank test fails to reject, s is close to one; the main exceptions are AEM1 in the M10 and All cross-sections, for which $s = 0.78$ and $s = 0.84$, respectively. Moreover, Eq. (14) shows that as s approaches one, the amplification channel makes risk price estimates highly sensitive to small perturbations in $(\boldsymbol{\mu}, \widehat{\boldsymbol{\beta}})$. Accordingly, when rank-test non-rejection coincides with a large value of s, we regard the unrestricted CSR estimates $(\widehat{\lambda}_0, \widehat{\boldsymbol{\lambda}})$ as weakly identified and therefore unreliable.

The restricted CSR gives a somewhat more favorable picture. Using the p-values after the comma, the reduced-rank null is rejected at the 5% level in 34 of the 78 specifications. Relative to the unrestricted CSR, the restricted test newly rejects reduced rank for SV2 (M10), AEM1 (All), AEM2 (BM10 and All), and CCAPM (INV10, OP10, All, and All*). In the opposite direction, we detect only one case, HKM2 with OP10, in which the unrestricted test rejects while the restricted test does not; even there, the restricted-CSR non-rejection is marginal, with a p-value of 5.7%.

These results indicate that weak identification of factor risk prices is more likely when the intercept is included, so the corresponding unrestricted risk price estimates carry less weight than those from the restricted CSR. The reason is that unrestricted identification requires more than factor loadings that are jointly different from zero: it requires K independent sources of CS loading variation after the intercept has been partialled out. Equivalently, conditional on the loading matrix itself being of full rank, the incremental identification problem created by the intercept arises when there exists a nonzero vector $\boldsymbol{a}$ such that $\widehat{\boldsymbol{\beta}}\boldsymbol{a}$ is nearly proportional to $\mathbf{1}$ (see Corollary 7). In a single-factor model, this reduces to the single beta vector being nearly proportional to $\mathbf{1}$. In multi-factor models, however, the precise statement is about a linear combination of beta columns, not necessarily about every beta column and not necessarily about the market beta alone. In the equity applications studied here, the market-beta column is nevertheless a natural and empirically important source of this near-collinearity, since market betas tend to have a large positive CS mean, typically close to one, and little CS dispersion across portfolios, as emphasized by Ahn, Perez, and Gadarowski (2013); Ahn and Horenstein (2020). Notably, this same intercept-loading near-collinearity is what drives the collinearity amplification channel in the risk price estimates discussed above.

6 Implied Zero-Beta Rates and Aggregate Equity Premia

A common justification for estimating an unrestricted intercept in an excess-return CSR is that the risk-free rate used to form excess returns may differ from the economically relevant zero-beta benchmark because of money-market frictions, liquidity premia, or borrowing–lending wedges. Under this interpretation, $\widehat{\lambda}_0$ is the sample wedge between the model-implied benchmark return and the observed risk-free rate. This section asks whether that interpretation is economically plausible given the intercept magnitudes documented in the previous section.

Let $\overline{R}_f$ denote the sample mean of the short rate used to construct excess returns. Under the zero-beta interpretation, the intercept implies the average benchmark return, $\widehat{r}^{zb} = \overline{R}_f + \widehat{\lambda}_0$. Thus, $\widehat{\lambda}_0$ is the implied excess of the zero-beta rate over the observed risk-free rate. If $\overline{R}_M$ denotes the sample mean market return and $\overline{RM} = \overline{R}_M - \overline{R}_f$ is the sample mean market excess return (i.e., the observed market premium), then the implied aggregate equity premium (relative to that benchmark return) is

$$\widehat{EP} = \overline{R}_M - \widehat{r}^{zb} = \overline{RM} - \widehat{\lambda}_0, \tag{17}$$

which means that $\widehat{EP} < 0$ implies $\widehat{\lambda}_0 > \overline{RM}$. In our sample, $\overline{R}_f = 1.19\%$ per quarter. For the 40-portfolio cross-section, even the smallest unrestricted intercept in Table 4, $\widehat{\lambda}_0 = 0.46\%$ for the CCAPM, implies $\widehat{r}^{zb} = 1.65\%$ per quarter (about 6.60% at an annual rate). This represents a gap of roughly 1.84 percentage points per year relative to the observed short-term rate. This is already at the upper end of the magnitudes typically invoked to motivate an unrestricted intercept (e.g., Lewellen, Nagel, and Shanken (2010)). However, most intercept estimates in Table 4 are considerably larger, thus leading to even more implausible implied zero-beta rates. The results are even more striking at the monthly frequency: with $\overline{R}_f = 0.39\%$, the smallest intercept estimate reported in Table 5 (0.63% for LZ1) yields $\widehat{r}^{zb} = 1.02\%$ per month (12.24% at an annual rate), which means a gap around 7.6 percentage points per year.

Naturally, these large implied zero-beta returns lead to (largely implausible) very small aggregate risk premia. At the quarterly frequency, we have $\overline{RM} = 1.72\%$ per quarter. Thus, for 10 of the 13 models in Table 4 (and similarly in the estimation with the 17 portfolios), $\widehat{\lambda}_0 > \overline{RM}$, so the implied aggregate equity premium is negative. At the monthly frequency, the sample market premium is 0.54% per month, which implies that the condition for a negative implied aggregate

equity premium is satisfied in all 20 estimations.[37] A negative (implied) equity premium represents a fundamental economic violation of any asset pricing model. Yet, this fact is typically ignored in CS asset pricing tests (with unrestricted CSRs), as the simple "back-of-the-envelope" computation above is not conducted.

The same point follows from the unrestricted CSR identity, $\widehat{\lambda}_0 = \overline{\mu} - \sum_{j=1}^{K} \widehat{\lambda}_j m_j$, with $m_j = \mathbb{E}_n(\widehat{\beta}_{i,j})$. Hence, a large intercept reflects a large gap between the CS average risk premium and the fitted beta component (i.e., the risk premium explained by the model). For the equity test assets used here, $\overline{\mu}$ is close to the sample average market factor.[38] A large positive $\widehat{\lambda}_0$ therefore means that the fitted factor component fails to match even the aggregate equity premium. This implication is especially transparent in the models that produce negative market prices of risk: when the market beta has a positive CS mean ($m_{RM} > 0$), which happens to be the case, a negative $\widehat{\lambda}_{RM}$ pushes $\sum_{j=1}^{K} \widehat{\lambda}_j m_j$ downward and mechanically raises $\widehat{\lambda}_0$ (often into the region where the implied aggregate equity premium becomes negative, as discussed above). In other words, a negative market risk price (as documented in Section 5) and a negative implied market premium are both manifestations of a very large intercept.

Overall, the zero-beta-rate interpretation provides a simple economic plausibility check on unrestricted CSRs. In our data, the implied benchmark-rate wedges and negative implied equity premia are too large and pervasive, which makes it difficult to interpret $\widehat{\lambda}_0$ as a money-market friction. They instead reinforce the paper's central interpretation of the intercept as a common pricing offset (for example, from omitted priced components) that should be counted as pricing error rather than treated as fitted compensation for risk.

7 Monte Carlo Simulations

We conduct two experiments to illustrate the relationship between model misspecification and the CS intercept, as predicted by Propositions 1 and 2. We employ the 40 portfolios as the test assets and either FF6 or the CAPM as the benchmark model. In the experiments described below, we simulate (jointly) the factors and the unexpected portfolio excess returns (with 10000 replications)

[37] The negative intercepts in our quarterly estimations (CCAPM on the 17 portfolios, HKM1 on BM10, and both AEM2 and CCAPM on M10, as reported in the IA) are likewise hard to reconcile with the money-market-wedge interpretation, since they imply a zero-beta rate below the observed short rate (see also Burnside (2011)).

[38] This is approximate rather than exact. Each anomaly group contains the full universe of stocks in a given period, but an equal-weighted average of value-weighted decile returns does not exactly coincide with the value-weighted market return. Nevertheless, the IA shows that the approximation is close in our data, as noted in Section 3.

by using the multivariate normal distribution and the first two sample moments of all variables. The quarterly non-market factors from FF6 are obtained by compounding the respective monthly factors.

First, we compute the demeaned return innovations from the auxiliary regression,

$$R_{i,t} = \mu_i + \varsigma_{i,t}, \tag{18}$$

which decomposes the total excess portfolio return into the expected (μ_i) and unexpected excess return ($\varsigma_{i,t}$). We use the same sample size as the original data ($T = 191$ or $T = 573$), which leads to the following pseudo series:

$$\left\{RM_t^m, SMB_t^m, HML_t^m, RMW_t^m, CMA_t^m, UMD_t^m, \varsigma_{i,t}^m\right\},$$
$$\left\{RM_t^m, \varsigma_{i,t}^{*m}\right\}, m = 1, ..., 10000.$$

Armed with the artificial data, we estimate the models (under the unrestricted CSR) and obtain a pseudo intercept estimate for each of the 10000 draws, $\left\{\widehat{\lambda}_0^m\right\}, m = 1, ..., 10000$. We then compute the average pseudo intercept across the 10000 replications, $\mathrm{E}^*(\widehat{\lambda}_0) = (1/10000)\sum_{m=1}^{10000}\widehat{\lambda}_0^m$.

The first experiment is based on the CAPM. The corresponding CSR,

$$\mu_i = \lambda_0 + \lambda_{RM}\widehat{\beta}_{i,RM} + \omega_i, \tag{19}$$

is estimated over the original samples. We assume that the CAPM is the true model in the population. Thus, the pseudo excess return for portfolio i (in each draw) equals

$$R_{i,t}^m = \widehat{\lambda}_{RM}\widehat{\beta}_{i,RM} + \varsigma_{i,t}^{*m}, \tag{20}$$

where the beta and risk price estimates are associated with (19). It follows that the average CAPM intercept estimates (across the 10000 replications) are very small at both the quarterly (-0.05%) and monthly (-0.02%) frequencies.

The second experiment is based on FF6, with the CS regression

$$\mu_i = \lambda_0 + \lambda_{RM}\widehat{\beta}_{i,RM} + \lambda_{SMB}\widehat{\beta}_{i,SMB} + \lambda_{HML}\widehat{\beta}_{i,HML} + \lambda_{RMW}\widehat{\beta}_{i,RMW} + \lambda_{CMA}\widehat{\beta}_{i,CMA} + \lambda_{UMD}\widehat{\beta}_{i,UMD} + \omega_i, \tag{21}$$

estimated over the original samples. We assume that FF6 holds in the population, with the return generating process given by

$$R^m_{i,t} = \widehat{\lambda}_{RM}\widehat{\beta}_{i,RM} + \widehat{\lambda}_{SMB}\widehat{\beta}_{i,SMB} + \widehat{\lambda}_{HML}\widehat{\beta}_{i,HML} + \widehat{\lambda}_{RMW}\widehat{\beta}_{i,RMW} + \widehat{\lambda}_{CMA}\widehat{\beta}_{i,CMA} + \widehat{\lambda}_{UMD}\widehat{\beta}_{i,UMD} + \varsigma^m_{i,t}, \tag{22}$$

where the factor loadings and risk prices are obtained from (21). The pseudo intercepts associated with both FF6 and the CAPM are estimated for each of the 10000 replications. We obtain average pseudo intercept estimates of -0.12% and 1.95% for FF6 and CAPM, respectively in the quarterly estimations. The monthly counterparts are -0.09% and 0.65%, respectively. Therefore, the misspecified model (CAPM) produces an average intercept that is about 16 times (or seven times) as large (in magnitude) the corresponding estimate for the "true" model. Overall, these experiments illustrate that model misspecification represents an important driver of large CS intercept estimates, consistent with Propositions 1 and 2.

8 Extension to GLS Estimation

The analysis above uses an OLS second pass, which gives equal CS weight to the original test portfolios. A natural alternative is GLS, motivated by efficiency under correct specification when the CS weighting matrix is the inverse of the residual or return covariance matrix; see Kandel and Stambaugh (1995), Lewellen, Nagel, and Shanken (2010), Kan, Robotti, and Shanken (2013), Kleibergen and Zhan (2015), and Gospodinov, Kan, and Robotti (2019). The omitted-factor economic interpretation of the intercept from Subsection 2.1 carries over from OLS to GLS since the GLS intercept is the $\boldsymbol{C}$-weighted analogue of the OLS one.

Let $\boldsymbol{C} = \widehat{\boldsymbol{\Sigma}}_\varepsilon^{-1} \succ 0$, where $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ is the first-pass residual covariance matrix; more generally, the identities below hold for any fixed positive-definite $\boldsymbol{C}$. Let $\boldsymbol{W} = [\mathbf{1} : \widehat{\boldsymbol{\beta}}]$. For any $n \times 1$ vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, define $\mathbb{E}_{n,\text{gls}}(x_i) = \mathbf{1}'\boldsymbol{C}\boldsymbol{x}/\mathbf{1}'\boldsymbol{C}\mathbf{1}$ and $\mathbb{E}_{n,\text{gls}}(x_i y_i) = \boldsymbol{x}'\boldsymbol{C}\boldsymbol{y}/\mathbf{1}'\boldsymbol{C}\mathbf{1}$, with $\text{Var}_{n,\text{gls}}(\cdot)$ and $\text{Cov}_{n,\text{gls}}(\cdot,\cdot)$ defined by the corresponding centered bilinear forms. Define $\overline{\mu}_{\text{gls}} = \mathbb{E}_{n,\text{gls}}(\mu_i)$, $\boldsymbol{m}_{\text{gls}} = \mathbb{E}_{n,\text{gls}}(\widehat{\boldsymbol{\beta}}_i)$,

$\boldsymbol{M}_{\text{gls}} = \mathbb{E}_{n,\text{gls}}(\widehat{\boldsymbol{\beta}}_i\widehat{\boldsymbol{\beta}}_i')$, and $\boldsymbol{\Sigma}_{\text{gls}} = \boldsymbol{M}_{\text{gls}} - \boldsymbol{m}_{\text{gls}}\boldsymbol{m}_{\text{gls}}'$. Further, define $\boldsymbol{v}_{\text{gls}} = \mathbb{E}_{n,\text{gls}}(\widehat{\boldsymbol{\beta}}_i\mu_i)$, $\boldsymbol{c}_{\text{gls}} = \boldsymbol{v}_{\text{gls}} - \boldsymbol{m}_{\text{gls}}\overline{\mu}_{\text{gls}}$, $\boldsymbol{u}_{\text{gls}} = \boldsymbol{M}_{\text{gls}}^{-1}\boldsymbol{m}_{\text{gls}}$, and $s_{\text{gls}} = \boldsymbol{m}_{\text{gls}}'\boldsymbol{M}_{\text{gls}}^{-1}\boldsymbol{m}_{\text{gls}}$. The scalar s_{gls} is the GLS analogue of s, the no-intercept auxiliary fit from projecting the constant on the beta span.[39] Throughout this section we assume $\boldsymbol{M}_{\text{gls}} \succ 0$ and $\boldsymbol{\Sigma}_{\text{gls}} \succ 0$, equivalently $s_{\text{gls}} \in [0, 1)$, so that the restricted and unrestricted GLS CSRs are identified; the IA establishes the corresponding identification conditions.

8.1 GLS Intercept and Risk Prices

The algebraic identities of Sections 2.2 and 2.3 carry over under GLS once ordinary CS moments are replaced by their $\boldsymbol{C}$-weighted counterparts. The restricted and unrestricted GLS CSRs are

$$\mu_i = \widetilde{\boldsymbol{\lambda}}_{\text{gls}}'\widehat{\boldsymbol{\beta}}_i + \widehat{\nu}_{i,\text{gls}}, \qquad \mu_i = \widehat{\lambda}_{0,\text{gls}} + \widehat{\boldsymbol{\lambda}}_{\text{gls}}'\widehat{\boldsymbol{\beta}}_i + \widehat{\omega}_{i,\text{gls}},$$

respectively, where $\widehat{\nu}_{i,\text{gls}}$ and $\widehat{\omega}_{i,\text{gls}}$ denote the restricted and unrestricted GLS CS residuals. We have $\widetilde{\boldsymbol{\lambda}}_{\text{gls}} = \boldsymbol{M}_{\text{gls}}^{-1}\boldsymbol{v}_{\text{gls}}$ and $\widehat{\boldsymbol{\lambda}}_{\text{gls}} = \boldsymbol{\Sigma}_{\text{gls}}^{-1}\boldsymbol{c}_{\text{gls}}$. All identities are exact finite-sample equalities conditional on $(\boldsymbol{\mu}, \widehat{\boldsymbol{\beta}}, \boldsymbol{C})$.

Lemma 2 (GLS analogue of Lemma 1)**.** *When the restricted and unrestricted GLS CSRs are well defined,*

$$\widehat{\boldsymbol{\lambda}}_{\text{gls}} - \widetilde{\boldsymbol{\lambda}}_{\text{gls}} = -\widehat{\lambda}_{0,\text{gls}}\boldsymbol{u}_{\text{gls}}.$$

Lemma 2 shows that a free GLS intercept changes the GLS risk price vector only in the direction $-\boldsymbol{u}_{\text{gls}}$. The direction depends on the $\boldsymbol{C}$-weighted first and second moments of betas, while the magnitude is governed by the intercept.

Proposition 7 (GLS analogue of Proposition 3)**.** *When the restricted and unrestricted GLS CSRs are well defined,*

$$\widehat{\lambda}_{0,\text{gls}} = \frac{\mathbb{E}_{n,\text{gls}}(\widehat{\nu}_{i,\text{gls}})}{1 - s_{\text{gls}}} = \frac{\overline{\mu}_{\text{gls}} - \boldsymbol{m}_{\text{gls}}'\widetilde{\boldsymbol{\lambda}}_{\text{gls}}}{1 - s_{\text{gls}}}.$$

Proposition 7 shows that the GLS intercept absorbs the GLS-weighted mean pricing error from the restricted CSR, amplified by $(1 - s_{\text{gls}})^{-1}$. Because $\boldsymbol{C}$ changes both $\boldsymbol{m}_{\text{gls}}$ and $\boldsymbol{M}_{\text{gls}}$, the OLS index s and the GLS index s_{gls} need not move one-for-one. The identification conditions for the GLS CSR are collected in the IA.

[39] Geometrically, s_{gls} is the squared projection of $\mathbf{1}$ onto the beta span in the $\boldsymbol{C}$-weighted inner product, normalized by the squared $\boldsymbol{C}$-length of $\mathbf{1}$.

The quarterly GLS estimates of λ_0, reported in the IA, are larger on average than the OLS counterparts, varying between 2.31% (AEM1) and 3.71% per quarter (HKM2), with significance at the 1% level in all 13 cases (i.e., t-ratios around or above 4.63). Consequently, we observe the sign reversal in all nine market risk prices. The key difference relative to the OLS estimations is that the unrestricted estimates of λ_{RM} within JW2, SV3, and SV4 become significant. The monthly intercept estimates vary between 0.68% (CRR5) and 1.50% per month (LZ4), also with statistical significance (5% level) in all cases. Relative to the OLS counterparts, the negative unrestricted estimate of λ_{RM} within the CAPM ceases to be significant. On the other hand, there is a sign flip in the risk price estimates associated with ΔDY, as in the OLS estimation.

8.2 GLS Pricing Errors and Fit

The unrestricted GLS pricing-error vector is

$$\widehat{\alpha}_{i,\text{gls}} = \mu_i - \widehat{\boldsymbol{\lambda}}'_{\text{gls}}\widehat{\boldsymbol{\beta}}_i = \widehat{\omega}_{i,\text{gls}} + \widehat{\lambda}_{0,\text{gls}},$$

where $\mathbb{E}_{n,\text{gls}}(\widehat{\omega}_{i,\text{gls}}) = 0$. This mirrors equation (9) in Section 2.2. Define

$$R^2_{\text{gls}} = 1 - \frac{\mathbb{E}_{n,\text{gls}}(\widehat{\omega}^2_{i,\text{gls}})}{\text{Var}_{n,\text{gls}}(\mu_i)}, \qquad R^2_{\text{pen,gls}} = 1 - \frac{\mathbb{E}_{n,\text{gls}}(\widehat{\alpha}^2_{i,\text{gls}})}{\text{Var}_{n,\text{gls}}(\mu_i)}, \qquad \text{Wedge}_{\text{gls}} = \frac{\widehat{\lambda}^2_{0,\text{gls}}}{\text{Var}_{n,\text{gls}}(\mu_i)}.$$

The statistic $R^2_{\text{pen,gls}}$ treats the intercept as part of total pricing error, while R^2_{gls} credits it as part of fitted risk premia.

Proposition 8 (GLS analogue of Proposition 4)**.** *If* $\text{Var}_{n,\text{gls}}(\mu_i) > 0$*, then*

$$R^2_{\text{pen,gls}} = R^2_{\text{gls}} - \frac{\widehat{\lambda}^2_{0,\text{gls}}}{\text{Var}_{n,\text{gls}}(\mu_i)} = R^2_{\text{gls}} - \text{Wedge}_{\text{gls}}, \qquad \mathbb{E}_{n,\text{gls}}(\widehat{\alpha}^2_{i,\text{gls}}) = \mathbb{E}_{n,\text{gls}}(\widehat{\omega}^2_{i,\text{gls}}) + \widehat{\lambda}^2_{0,\text{gls}}.$$

Moreover, $R^2_{\text{pen,gls}} = R^2_{\text{gls}}$ *if and only if* $\widehat{\lambda}_{0,\text{gls}} = 0$.

Corollary 8 (GLS analogue of Corollary 2)**.** *Let*

$$R^2_{\text{gls}*} = 1 - \frac{\mathbb{E}_{n,\text{gls}}(\widehat{\nu}^2_{i,\text{gls}})}{\text{Var}_{n,\text{gls}}(\mu_i)}.$$

Then

$$R^2_{\text{gls}*} - R^2_{\text{pen,gls}} = s_{\text{gls}} \frac{\widehat{\lambda}^2_{0,\text{gls}}}{\text{Var}_{n,\text{gls}}(\mu_i)} = s_{\text{gls}} \text{Wedge}_{\text{gls}}.$$

Proposition 8 establishes that the intercept enters GLS fit in the same way that it enters OLS fit. Corollary 8 establishes that the unrestricted GLS CSR cannot outperform the restricted GLS CSR once the intercept is treated as pricing error, with the gap widening in s_{gls}. [40]

The results (for both augmented estimations) tabulated in the IA indicate a similar qualitative picture to the OLS counterparts reported in Tables 1 to 3. Concretely, $R^2_{\text{pen,gls}}$ assumes negative values in all 20 monthly estimations and in 25 of the 26 quarterly estimations, compared to several significantly positive estimates of R^2_{gls}. As in the OLS estimations, we detect a few cases of overshooting (i.e., models producing simultaneously a significantly positive $R^2_{\text{gls}*}$ and a negative $R^2_{\text{pen,gls}}$), especially at the monthly frequency.[41]

9 Conclusion

This paper studies how the model-implied zero-intercept restriction affects model evaluation, risk price estimation, and identification in two-pass CSRs with excess-return test assets. Linear factor models restrict the CS intercept to zero, yet empirical work routinely estimates it freely. We show that a free intercept absorbs the average pricing error of the restricted (no-intercept) regression and mechanically reallocates pricing failures into fitted risk premia, inflating conventional measures of fit and distorting risk price estimates even though the tested model is unchanged. Structurally, when the model omits a priced factor, the intercept equals the risk-price-weighted average of the omitted loading component not spanned by the included factors, so it is informative about missing pricing content rather than about the explanatory power of the tested factors.

On the theoretical side, we derive an exact mapping between the restricted and unrestricted

[40] Our analysis generalizes the GLS results of Kandel and Stambaugh (1995). Their single-traded-index setting is the special case in which $K = 1$, the factor is the excess return on a portfolio p of the test assets, and $\boldsymbol{C} = \boldsymbol{V}^{-1}$; our identities also cover multiple factors, non-traded factors, and any fixed positive-definite CS weighting matrix. Our $R^2_{\text{gls}*}$ has no counterpart in their unrestricted analysis. The weighting choice is innocuous for the coefficient estimates too. When the first pass is an OLS time-series regression with an intercept and $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ is nonsingular, the restricted and unrestricted GLS coefficients coincide exactly under $\boldsymbol{C} = \widehat{\boldsymbol{V}}^{-1}$ and $\boldsymbol{C} = \widehat{\boldsymbol{\Sigma}}_\varepsilon^{-1}$ (see footnote 9 of Kandel and Stambaugh 1995). However, the weighted moments do depend on it, so the fit measures and s_{gls} reported in the IA are specific to the residual-covariance weighting.

[41] The main qualitative discrepancy with respect to the OLS estimation is that there are substantially fewer cases in which the standard metric (R^2_{gls}) postulates a significantly positive fit, a pattern that is particularly evident in the quarterly estimations. These results are in line with previous evidence showing that the GLS estimation represents a bigger hurdle for linear factor models (e.g., Lewellen, Nagel, and Shanken (2010) and Kan, Robotti, and Shanken (2013)).

second-pass estimates and characterize when the distortions are large. The key determinant is the near-collinearity between the factor loadings and a constant: when loadings have small CS dispersion relative to their average magnitude, even negligible average pricing errors generate large intercepts and correspondingly large changes in estimated premia, including sign reversals, through a familiar collinearity-amplification channel. Freeing the intercept tightens the CS identification condition, and near-collinearity brings it close to binding, weakening identification. These results motivate penalized fit measures that treat the intercept as pricing error and thereby quantify the distortion a free intercept introduces.

Empirically, these effects are first-order in standard applications. Across the CAPM and a broad set of macro-factor models estimated on widely used anomaly-sorted portfolios at both quarterly and monthly frequencies, intercept estimates are almost always economically large and the amplification conditions are typically present. Consequently, the penalized fit measure is negative in 76 of 78 quarterly and in all 20 monthly specifications, even where the conventional residual-based R^2 appears high; allowing an intercept frequently attenuates or reverses estimated factor premia, including the market price of risk; and rank tests indicate weaker identification in the unrestricted specification. The intercept magnitudes further imply zero-beta rates and aggregate equity premia that are difficult to reconcile with standard economic priors. We therefore recommend that excess-return CSRs be estimated without an intercept, reserving the unrestricted intercept as a diagnostic for a common pricing offset such as omitted priced factors.

This paper follows the related literature in focusing on CS tests for equity portfolio risk premia. Testing our theoretical propositions with CS tests on individual equity risk premia or risk premia from other asset classes (such as currencies, commodities, corporate bonds, treasury bonds, or options) would provide complementary evidence. This is beyond the scope of this paper and is left for future research.

References

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

Ahn, S. C., A. R. Horenstein, and N. Wang, 2018, Beta matrix and common factors in stock returns, *Journal of Financial and Quantitative Analysis* 53, 1417–1440.

Ahn, S. C. and A. Horenstein, 2020, Market betas in multi-factor models, SSRN Electronic Journal.

Ahn, S. C., M. F. Perez, and C. Gadarowski, 2013, Two-pass estimation of risk premiums with multicollinear and near-invariant betas, *Journal of Empirical Finance* 20, 1–17.

Black, F., 1972, Capital market equilibrium with restricted borrowing, *Journal of Business* 45, 444–455.

Black, F., M.C. Jensen, and M. Scholes, 1972, The capital asset pricing model: Some empirical tests, in M.C. Jensen, ed.: Studies in the Theory of Capital Markets (Praeger, New York).

Breeden, D.T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.

Brennan, M., 1971, Capital market equilibrium with divergent borrowing and lending rates, *Journal of Financial and Quantitative Analysis* 6, 1197–1205.

Brennan, M. J., A. W. Wang, and Y. Xia, 2004, Estimation and test of a simple model of intertemporal capital asset pricing, *Journal of Finance* 59, 1743–1775.

Bryzgalova, S., 2015, Spurious factors in linear asset pricing models, Working Paper.

Burnside, C., 2011, The cross section of foreign currency risk premia and consumption growth risk: Comment, *American Economic Review* 101, 3456–3476.

Campbell, J.Y., S. Giglio, C. Polk, and R. Turley, 2018, An intertemporal CAPM with stochastic volatility, *Journal of Financial Economics* 128, 207–33.

Campbell, J.Y., and T. Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review* 94, 1249–75.

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

Cochrane, J.H., 2005, *Asset Pricing* (Princeton University Press, Princeton, NJ).

Cooper, I., L. Ma, and P. Maio, 2022, What does the cross-section tell about itself? Explaining equity risk premia with stock return moments, *Journal of Money, Credit and Banking* 54, 73–118.

Epstein, L.G., and S.E. Zin, 1991, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis, *Journal of Political Economy* 99, 263–286.

Dello-Preite, M., R. Uppal, P. Zaffaroni, and I. Zviadadze, 2025, Cross-sectional asset pricing with unsystematic risk, SSRN Electronic Journal.

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

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

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

Frazzini, A., and L.H. Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Giglio, S., and D. Xiu, 2021, Asset pricing with omitted factors, *Journal of Political Economy* 129, 1947–1990.

Giglio, S., D. Xiu, and D. Zhang, 2025, Test assets and weak factors, *Journal of Finance* 80, 259–319.

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

Hahn, J., and H. Lee, 2006, Yield spreads as alternative risk factors for size and book-to-market, *Journal of Financial and Quantitative Analysis* 41, 245–269.

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

Jagannathan, R., and Z. Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Kan, R., C. Robotti, and J. Shanken, 2013, Pricing model performance and the two-pass cross-sectional regression methodology, *Journal of Finance* 68, 2617–2649.

Kan, R. and C. Zhang, 1999, Two-pass tests of asset pricing models with useless factors, *Journal of Finance* 54, 203–235.

Kandel, S. and R. F. Stambaugh, 1995, Portfolio inefficiency and the cross-section of expected returns, *Journal of Finance* 50, 157–184.

Kleibergen, F., and Z. Zhan, 2015, Unexplained factors and their effects on second pass R-squared's, *Journal of Econometrics* 189, 101–116.

Kleibergen, F., and Z. Zhan, 2020, Robust inference for consumption-based asset pricing, *Journal of Finance* 75, 507–550.

Lettau, M., and S.C. Ludvigson, 2001, Resurrecting the (C)CAPM: A cross sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Lettau, M., S.C. Ludvigson, and S. Ma, 2019, Capital share risk in U.S. asset pricing, *Journal of Finance* 74, 1753–1792.

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

Liu, L.X., and L. Zhang, 2008, Momentum profits, factor pricing, and macroeconomic risk, *Review of Financial Studies* 21, 2417–2448.

Lucas, R.E., 1978, Asset prices in an exchange economy, *Econometrica* 46, 1429–1445.

Lustig, H.N., and S.G. Van Nieuwerburgh, 2005, Housing collateral, consumption insurance, and risk premia: An empirical perspective, *Journal of Finance* 60, 1167–1219.

Maio, P., 2024, Is long-run risk really priced? Revisiting Liu and Matthies (2022), *Journal of Finance* 79, 2885–2900.

Maio, P., and B.-K. Min, 2026, Revisiting Lettau and Ludvigson (2001): Does *cay* really matter for cross-sectional risk premia? *Critical Finance Review* (forthcoming).

Maio, P., and P. Santa-Clara, 2017, Short-term interest rates and stock market anomalies, *Journal of Financial and Quantitative Analysis* 52, 927–961.

Masten, M. A., and A. Poirier, 2026, The effect of omitted variables on the sign of regression coefficients, *American Economic Review*, forthcoming.

Merton, R.C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Petkova, R., 2006, Do the Fama-French factors proxy for innovations in predictive variables? *Journal of Finance* 61, 581–612.

Santos, T., and P. Veronesi, 2006, Labor income and predictable stock returns, *Review of Financial Studies* 19, 1–44.

Shanken, J., 1992, On the estimation of beta pricing models, *Review of Financial Studies* 5, 1–34.

Yogo, M., 2006, A consumption-based explanation of expected stock returns, *Journal of Finance* 61, 539–580.

Table 1: Pricing Performance in Quarterly Asset Pricing Tests I

This table presents the evaluation results for the asset pricing models at the quarterly frequency. The method is the standard two-step regression approach. The testing portfolios are deciles sorted on investment (INV10), operating profitability (OP10), and momentum (M10). For a description of the different models and factors, see Section 3. Wedge equals $\lambda_0^2/\operatorname{Var}_n(\mu_i)$, where λ_0 denotes the intercept estimate and $\operatorname{Var}_n(\mu_i)$ is the cross-sectional variance of portfolio risk premia. s is the cross-sectional projection of one on the factor loadings. $RRMSE_{pen}$ is the relative root mean squared pricing error, while $RRMSE$ denotes a similar metric, except that it is based on pricing errors that exclude the intercept. R^2_{pen} denotes the cross-sectional R^2, while R^2_{ols} denotes a similar metric, except that it is based on pricing errors that exclude the intercept. $RRMSE*$ and R^2_{ols*} denote the performance metrics associated with the CSR without intercept. The values in parentheses represent bootstrap p-values to test the null of a zero fit ($R^2_{ols} = 0$ or $R^2_{pen} = 0$) or, in the case of R^2_{ols*}, the null of no explanatory power (zero risk prices). The sample is 1970:I to 2017:III.

			Intercept						**No intercept**		
	Wedge	s	$RRMSE$	$RRMSE_{pen}$	R^2_{ols}	p	R^2_{pen}	p	$RRMSE*$	R^2_{ols*}	p
Panel A: INV10											
CAPM	63.08	0.98	0.18	1.67	0.24	(0.20)	−62.85		0.30	−1.13	
JW2	146.41	0.99	0.11	2.53	0.75	(0.04)	−145.66		0.30	−1.08	
JW3	147.22	0.99	0.10	2.53	0.77	(0.06)	−146.45		0.30	−1.07	
SV2	62.93	0.98	0.18	1.67	0.28	(0.40)	−62.66		0.30	−1.09	
SV3	160.95	0.99	0.10	2.65	0.75	(0.07)	−160.20		0.24	−0.34	
SV4	168.87	0.99	0.10	2.71	0.77	(0.11)	−168.10		0.24	−0.34	
LL2	61.70	0.98	0.17	1.65	0.35	(0.27)	−61.35		0.29	−0.98	
AEM1	3.93	0.91	0.10	0.43	0.76	(0.00)	−3.16		0.16	0.39	(0.05)
AEM2	3.98	0.99	0.10	0.43	0.76	(0.03)	−3.22		0.11	0.73	(0.01)
HKM1	54.26	0.99	0.20	1.55	0.10	(0.39)	−54.16		0.27	−0.64	
HKM2	15.27	0.99	0.14	0.83	0.53	(0.09)	−14.74		0.16	0.38	(0.10)
CCAPM	24.03	0.98	0.21	1.04	0.00	(0.95)	−24.03		0.25	−0.47	
LVN3	49.68	0.99	0.14	1.48	0.53	(0.22)	−49.15		0.22	−0.14	
Panel B: OP10											
CAPM	149.97	0.99	0.10	2.21	0.69	(0.04)	−149.28		0.29	−1.52	
JW2	122.21	0.99	0.10	2.00	0.71	(0.06)	−121.50		0.18	−0.02	
JW3	88.68	1.00	0.09	1.70	0.74	(0.10)	−87.95		0.14	0.39	(0.27)
SV2	143.09	0.99	0.10	2.16	0.69	(0.07)	−142.40		0.22	−0.44	
SV3	107.69	0.99	0.08	1.87	0.80	(0.06)	−106.89		0.16	0.18	(0.64)
SV4	121.00	0.99	0.08	1.99	0.83	(0.10)	−120.17		0.16	0.18	(0.75)
LL2	185.59	0.99	0.10	2.46	0.72	(0.06)	−184.87		0.23	−0.68	
AEM1	57.84	0.97	0.17	1.38	0.14	(0.43)	−57.70		0.29	−1.56	
AEM2	146.11	0.99	0.10	2.18	0.71	(0.07)	−145.40		0.28	−1.40	
HKM1	134.59	0.98	0.09	2.09	0.75	(0.03)	−133.83		0.31	−1.88	
HKM2	124.34	0.99	0.09	2.01	0.76	(0.05)	−123.58		0.23	−0.70	
CCAPM	26.04	0.98	0.18	0.94	0.00	(0.91)	−26.04		0.23	−0.62	
LVN3	38.03	0.98	0.17	1.12	0.14	(0.88)	−37.89		0.22	−0.51	
Panel C: M10											
CAPM	29.06	0.96	0.34	2.43	0.43	(0.03)	−28.63		0.57	−0.63	
JW2	5.87	0.98	0.18	1.09	0.84	(0.01)	−5.03		0.24	0.72	(0.01)
JW3	3.33	0.98	0.13	0.82	0.91	(0.01)	−2.42		0.17	0.85	(0.01)
SV2	50.97	0.97	0.23	3.19	0.74	(0.03)	−50.23		0.57	−0.62	
SV3	2.67	0.98	0.13	0.74	0.92	(0.01)	−1.76		0.16	0.87	(0.00)
SV4	3.67	1.00	0.13	0.86	0.92	(0.02)	−2.75		0.14	0.90	(0.01)
LL2	17.84	0.97	0.15	1.89	0.88	(0.01)	−16.96		0.37	0.30	(0.23)
AEM1	0.39	0.78	0.30	0.41	0.55	(0.05)	0.16	(0.04)	0.33	0.46	(0.02)
AEM2	26.56	1.00	0.28	2.32	0.59	(0.20)	−25.97		0.30	0.56	(0.04)
HKM1	32.63	0.96	0.28	2.56	0.61	(0.02)	−32.02		0.59	−0.78	
HKM2	24.91	0.96	0.23	2.24	0.73	(0.03)	−24.18		0.48	−0.16	
CCAPM	1.56	0.93	0.22	0.60	0.76	(0.01)	−0.80		0.26	0.66	(0.01)
LVN3	6.84	0.99	0.14	1.17	0.90	(0.01)	−5.94		0.17	0.85	(0.01)

Table 2: Pricing Performance in Quarterly Asset Pricing Tests II

This table presents the evaluation results for the asset pricing models at the quarterly frequency. The method is the standard two-step regression approach. In Panel A, the testing portfolios are deciles sorted on book-to-market (BM10). Panel B contains the results for the augmented estimation including all 40 portfolios (INV10+OP10+M10+BM10), while Panel C contains the results for a restricted cross-section containing the first two and last two deciles in each group plus the market factor (for a total of 17 portfolios). In everything else, the table is identical to Table 1.

			Intercept						**No intercept**		
	Wedge	s	$RRMSE$	$RRMSE_{pen}$	R^2_{ols}	p	R^2_{pen}	p	$RRMSE*$	R^2_{ols*}	p
Panel A: BM10											
CAPM	13.80	0.99	0.20	0.77	0.01	(0.79)	−13.79		0.21	−0.09	
JW2	43.84	1.00	0.19	1.34	0.14	(0.79)	−43.70		0.21	−0.07	
JW3	103.56	1.00	0.15	2.04	0.42	(0.54)	−103.15		0.20	−0.00	
SV2	13.77	0.99	0.20	0.77	0.02	(0.96)	−13.75		0.21	−0.08	
SV3	36.37	1.00	0.18	1.22	0.14	(0.95)	−36.22		0.19	0.06	(0.94)
SV4	33.01	1.00	0.18	1.16	0.23	(0.95)	−32.78		0.18	0.16	(0.94)
LL2	11.41	0.99	0.08	0.68	0.83	(0.02)	−10.58		0.10	0.75	(0.01)
AEM1	8.76	0.83	0.09	0.60	0.77	(0.01)	−7.99		0.26	−0.72	
AEM2	0.02	0.99	0.08	0.09	0.83	(0.02)	0.81	(0.00)	0.08	0.83	(0.01)
HKM1	1.29	0.98	0.13	0.26	0.60	(0.04)	−0.69		0.13	0.58	(0.00)
HKM2	7.35	0.99	0.09	0.55	0.79	(0.04)	−6.56		0.10	0.74	(0.03)
CCAPM	8.74	0.97	0.19	0.62	0.11	(0.57)	−8.63		0.21	−0.13	
LVN3	4.93	0.98	0.16	0.47	0.40	(0.55)	−4.53		0.17	0.28	(0.52)
Panel B: All											
CAPM	50.18	0.98	0.24	2.08	0.32	(0.03)	−49.86		0.39	−0.76	
JW2	28.69	0.99	0.22	1.58	0.45	(0.02)	−28.25		0.29	0.03	(0.56)
JW3	25.28	0.99	0.22	1.48	0.46	(0.04)	−24.83		0.27	0.15	(0.44)
SV2	54.28	0.98	0.24	2.16	0.33	(0.07)	−53.95		0.38	−0.69	
SV3	29.12	0.99	0.21	1.59	0.49	(0.03)	−28.63		0.28	0.07	(0.65)
SV4	32.06	0.99	0.21	1.66	0.49	(0.05)	−31.57		0.28	0.09	(0.70)
LL2	48.45	0.98	0.23	2.04	0.38	(0.03)	−48.07		0.38	−0.66	
AEM1	2.84	0.84	0.21	0.53	0.49	(0.00)	−2.35		0.29	0.03	(0.34)
AEM2	15.89	0.99	0.19	1.18	0.55	(0.01)	−15.34		0.23	0.37	(0.03)
HKM1	37.98	0.98	0.26	1.81	0.21	(0.07)	−37.77		0.38	−0.71	
HKM2	49.00	0.98	0.24	2.05	0.33	(0.07)	−48.67		0.38	−0.71	
CCAPM	0.67	0.96	0.25	0.35	0.25	(0.05)	−0.41		0.26	0.23	(0.01)
LVN3	4.53	0.98	0.21	0.66	0.47	(0.02)	−4.06		0.23	0.35	(0.03)
Panel C: All*											
CAPM	49.59	0.98	0.29	2.95	0.51	(0.00)	−49.08		0.52	−0.54	
JW2	22.75	0.99	0.26	2.00	0.60	(0.01)	−22.15		0.33	0.38	(0.03)
JW3	22.12	0.99	0.26	1.98	0.60	(0.03)	−21.52		0.32	0.40	(0.12)
SV2	54.41	0.98	0.29	3.08	0.52	(0.02)	−53.89		0.51	−0.49	
SV3	23.19	0.99	0.26	2.02	0.62	(0.03)	−22.57		0.32	0.40	(0.12)
SV4	27.64	0.99	0.25	2.20	0.63	(0.05)	−27.00		0.32	0.40	(0.21)
LL2	45.47	0.98	0.29	2.82	0.53	(0.02)	−44.94		0.48	−0.36	
AEM1	0.99	0.71	0.27	0.50	0.57	(0.01)	−0.42		0.35	0.29	(0.07)
AEM2	17.95	0.99	0.23	1.78	0.70	(0.00)	−17.25		0.30	0.49	(0.01)
HKM1	26.93	0.97	0.36	2.19	0.27	(0.03)	−26.66		0.51	−0.51	
HKM2	50.14	0.98	0.28	2.96	0.54	(0.01)	−49.60		0.51	−0.51	
CCAPM	0.88	0.95	0.29	0.48	0.52	(0.01)	−0.35		0.30	0.48	(0.00)
LVN3	2.86	0.98	0.20	0.73	0.76	(0.00)	−2.10		0.22	0.72	(0.00)

Table 3: Pricing Performance in Monthly Asset Pricing Tests

This table presents the evaluation results for the asset pricing models at the monthly frequency. The method is the standard two-step regression approach. The testing portfolios are deciles sorted on investment (INV10), operating profitability (OP10), momentum (M10), and book-to-market (BM10). For a description of the different models and factors, see Section 3. Panel A contains the results for the augmented estimation including all 40 portfolios (INV10+OP10+M10+BM10), while Panel B contains the results for a restricted cross-section containing the first two and last two deciles in each group plus the market factor (for a total of 17 portfolios). Wedge equals $\lambda_0^2/\operatorname{Var}_n(\mu_i)$, where λ_0 denotes the intercept estimate and $\operatorname{Var}_n(\mu_i)$ is the cross-sectional variance of portfolio risk premia. s is the cross-sectional projection of one on the factor loadings. $RRMSE_{pen}$ is the relative root mean squared pricing error, while $RRMSE$ denotes a similar metric, except that it is based on pricing errors that exclude the intercept. R^2_{pen} denotes the cross-sectional R^2, while R^2_{ols} denotes a similar metric, except that it is based on pricing errors that exclude the intercept. $RRMSE*$ and R^2_{ols*} denote the performance metrics associated with the CSR without intercept. The values in parentheses represent bootstrap p-values to test the null of a zero fit ($R^2_{ols} = 0$) or, in the case of R^2_{ols*}, the null of no explanatory power (zero risk prices). The sample is 1970:01 to 2017:09.

			Intercept					**No intercept**		
	Wedge	s	$RRMSE$	$RRMSE_{pen}$	R^2_{ols}	p	R^2_{pen}	$RRMSE*$	R^2_{ols*}	p
Panel A: All										
CAPM	57.56	0.98	0.26	2.40	0.33	(0.01)	−57.22	0.39	−0.57	
HL3	120.56	0.99	0.22	3.45	0.49	(0.01)	−120.07	0.37	−0.35	
HL5	164.22	1.00	0.19	4.03	0.63	(0.00)	−163.59	0.26	0.33	(0.15)
P5	168.45	1.00	0.20	4.08	0.59	(0.01)	−167.87	0.28	0.22	(0.41)
P7	171.90	1.00	0.19	4.12	0.63	(0.01)	−171.27	0.23	0.44	(0.11)
LZ1	10.34	0.22	0.29	1.05	0.13	(0.22)	−10.21	0.94	−7.95	
LZ4	203.08	1.00	0.19	4.48	0.62	(0.00)	−202.46	0.32	−0.06	
CRR5	22.41	0.97	0.21	1.50	0.55	(0.01)	−21.86	0.32	−0.06	
CRR6	43.80	0.99	0.20	2.09	0.58	(0.01)	−43.22	0.26	0.31	(0.31)
FF6	17.84	1.00	0.14	1.33	0.81	(0.00)	−17.03	0.14	0.80	(0.00)
Panel B: All*										
CAPM	54.96	0.98	0.33	3.36	0.48	(0.00)	−54.49	0.53	−0.40	
HL3	102.67	0.99	0.26	4.58	0.66	(0.01)	−102.01	0.51	−0.30	
HL5	77.66	1.00	0.17	3.98	0.86	(0.00)	−76.79	0.26	0.68	(0.01)
P5	151.61	1.00	0.21	5.56	0.78	(0.01)	−150.82	0.32	0.51	(0.10)
P7	40.63	1.00	0.16	2.88	0.87	(0.01)	−39.76	0.18	0.84	(0.00)
LZ1	6.83	0.50	0.35	1.23	0.41	(0.10)	−6.42	0.91	−3.03	
LZ4	128.67	0.99	0.14	5.12	0.90	(0.00)	−127.77	0.39	0.24	(0.41)
CRR5	14.20	0.96	0.22	1.71	0.76	(0.01)	−13.44	0.40	0.21	(0.74)
CRR6	42.51	0.99	0.19	2.95	0.82	(0.01)	−41.69	0.31	0.54	(0.24)
FF6	35.85	1.00	0.13	2.70	0.92	(0.00)	−34.93	0.14	0.90	(0.00)

Table 4: Quarterly Risk Price Estimates: All Portfolios

This table presents the quarterly risk price estimates. The method is the standard two-step regression approach. The testing portfolios are deciles sorted on investment (INV10), operating profitability (OP10), momentum (M10), and book-to-market (BM10). λ_0 denotes the intercept estimate (in %). λ denote the risk price estimates (in %), while the numbers in parentheses represent the respective Shanken t-ratios (with * denoting statistical significance at the 5% level). The risk price estimates and t-ratios displayed after the comma are associated with the corresponding CSR with no intercept. $\mathbb{E}_n(\widehat{\nu}_i)$ denotes the average pricing errors (in %) of the restricted CSR. For a description of the different models and factors, see Section 3. The sample is 1970:I to 2017:III.

	λ_0	λ_{RM}	λ_Y	λ_{def}	λ_{RMsw}	λ_{Ysw}	λ_{RMcay}	$\mathbb{E}_n(\widehat{\nu}_i)$
CAPM	3.99	$-2.13, 1.76$						0.09
	(5.11*)	($-2.12^*, 2.76^*$)						
JW2	3.02	$-1.09, 1.95$	$0.68, 1.41$					0.04
	(3.24*)	($-0.98, 3.03^*$)	($2.22^*, 2.92^*$)					
JW3	2.84	$-0.89, 1.98$	$0.59, 0.93$	$-0.13, -0.35$				0.03
	(2.63*)	($-0.73, 3.11^*$)	($3.02^*, 3.30^*$)	($-0.66, -1.70$)				
SV2	4.15	$-2.27, 1.71$			$-0.06, 0.11$			0.08
	(5.06*)	($-2.19^*, 2.69^*$)			($-1.05, 1.98^*$)			
SV3	3.04	$-1.14, 1.93$	$0.68, 1.42$			$0.02, 0.02$		0.04
	(3.03*)	($-0.98, 2.97^*$)	($2.06^*, 2.82^*$)			(1.19, 0.83)		
SV4	3.19	$-1.27, 1.89$	$0.65, 1.40$		$-0.00, 0.12$	$0.02, 0.02$		0.04
	(3.27*)	($-1.11, 2.93^*$)	($2.15^*, 2.78^*$)		($-0.02, 1.40$)	(1.20, 0.83)		
LL2	3.93	$-1.99, 1.85$					$0.05, 0.07$	0.08
	(4.88*)	($-1.97^*, 2.92^*$)					(1.15, 1.62)	
	λ_0	λ_{RM}	λ_{LEV}	λ_{IC}	λ_C	λ_{my}	λ_{Cmy}	$\mathbb{E}_n(\widehat{\nu}_i)$
AEM1	0.95		$11.07, 20.87$					0.15
	(0.93)		($2.80^*, 1.70$)					
AEM2	2.25	$-0.47, 1.69$	$8.63, 14.21$					0.03
	(2.00*)	($-0.36, 2.64^*$)	($2.39^*, 4.02^*$)					
HKM1	3.48			$-0.03, 0.03$				0.08
	(4.41*)			($-1.66, 2.74^*$)				
HKM2	3.95	$-2.12, 1.69$		$-0.02, 0.03$				0.08
	(4.93*)	($-2.09^*, 2.67^*$)		($-0.87, 2.10^*$)				
CCAPM	0.46				$0.79, 1.04$			0.02
	(0.36)				($2.29^*, 1.80$)			
LVN3	1.20				$0.38, 0.89$	$-14.98, -14.11$	$-0.03, -0.04$	0.03
	(1.11)				($1.07, 2.25^*$)	($-2.15^*, -1.75$)	($-0.41, -0.48$)	

Table 5: Monthly Risk Price Estimates: All Portfolios

This table presents the monthly risk price estimates. The method is the standard two-step regression approach. The testing portfolios are deciles sorted on investment (INV10), operating profitability (OP10), momentum (M10), and book-to-market (BM10). λ_0 denotes the intercept estimate (in %). λ denote the risk price estimates (in %), while the numbers in parentheses represent the respective Shanken t-ratios (with * denoting statistical significance at the 5% level). The risk price estimates and t-ratios displayed after the comma are associated with the corresponding CSR with no intercept. $\mathbb{E}_n(\widehat{\nu}_i)$ denotes the average pricing errors (in %) of the restricted CSR. For a description of the different models and factors, see Section 3. The sample is 1970:01 to 2017:09.

	λ_0	λ_{RM}	$\lambda_{\Delta TS}$	$\lambda_{\Delta DS}$	$\lambda_{\Delta DY}$	$\lambda_{\Delta TB}$	λ_{SMB}	λ_{HML}	$\mathbb{E}_n(\widehat{\nu}_i)$
CAPM	1.48	$-0.88, 0.56$							0.02
	(5.26*)	$(-2.62^*, 2.95^*)$							
HL3	2.14	$-1.54, 0.58$	$-0.01, -0.09$	$-0.12, 0.07$					0.02
	(4.31*)	$(-2.94^*, 3.06^*)$	$(-0.08, -0.81)$	$(-2.35^*, 2.07^*)$					
HL5	2.50	$-1.92, 0.51$	$-0.15, -0.60$	$-0.04, 0.01$			$0.64, -0.06$	$0.21, 0.30$	0.00
	(4.09*)	$(-3.00^*, 2.63^*)$	$(-1.25, -2.42^*)$	$(-1.26, 0.37)$			$(2.10^*, -0.22)$	$(1.52, 2.08^*)$	
P5	2.53	$-1.92, 0.61$	$0.02, 0.18$	$-0.06, -0.04$	$2.05, -0.58$	$0.01, -0.04$			0.01
	(4.41*)	$(-3.22^*, 3.17^*)$	$(0.18, 1.02)$	$(-2.15^*, -0.84)$	$(3.11^*, -2.48^*)$	$(0.66, -2.39^*)$			
P7	2.56	$-1.98, 0.55$	$-0.16, -0.27$	$-0.04, -0.01$	$2.05, -0.42$	$0.02, 0.00$	$0.65, -0.16$	$0.21, 0.36$	0.00
	(4.66*)	$(-3.44^*, 2.92^*)$	$(-1.36, -1.97^*)$	$(-1.23, -0.45)$	$(3.27^*, -1.84)$	$(1.60, 0.09)$	$(2.61^*, -0.70)$	$(1.51, 2.59^*)$	
	λ_0	λ_{RM}	λ_{SMB}	λ_{HML}	λ_{MP}	λ_{RMW}	λ_{CMA}	λ_{UMD}	$\mathbb{E}_n(\widehat{\nu}_i)$
LZ1	0.63				$0.68, -1.83$				0.49
	(2.42*)				$(2.50^*, -0.83)$				
LZ4	2.78	$-2.18, 0.56$	$0.73, -0.55$	$0.15, 0.29$	$0.35, 0.57$				0.01
	(3.80*)	$(-2.92^*, 2.96^*)$	$(2.23^*, -2.33^*)$	$(1.02, 2.05^*)$	$(1.69, 2.27^*)$				
FF6	0.82	$-0.28, 0.53$	$0.51, 0.19$	$0.28, 0.32$		$0.30, 0.30$	$0.23, 0.24$	$0.58, 0.58$	0.00
	(1.86)	$(-0.59, 2.82^*)$	$(2.28^*, 0.98)$	$(2.11^*, 2.45^*)$		$(2.75^*, 2.77^*)$	$(2.49^*, 2.55^*)$	$(3.08^*, 3.05^*)$	
	λ_0	λ_{RM}	λ_{MP}	λ_{TS}	λ_{DS}	λ_{UI}	λ_{EI}		$\mathbb{E}_n(\widehat{\nu}_i)$
CRR5	0.92		$0.06, 0.10$	$-0.78, 1.99$	$-0.25, -0.41$	$-0.01, -0.38$	$0.12, 0.10$		0.03
	(3.63*)		$(0.36, 0.28)$	$(-1.34, 1.09)$	$(-1.45, -1.19)$	$(-0.12, -2.23^*)$	$(1.69, 0.71)$		
CRR6	1.29	$-0.69, 0.58$	$-0.25, -0.81$	$-0.21, -0.38$	$0.08, 0.05$	$0.02, -0.18$	$0.08, 0.18$		0.01
	(3.30*)	$(-1.63, 3.03^*)$	$(-0.66, -1.39)$	$(-1.27, -1.65)$	$(0.50, 0.20)$	$(0.41, -1.97^*)$	$(1.63, 1.84)$		

Table 6: Rank Tests on Factor Loadings

This table presents the p-values associated with the F-test on the quarterly factor loading estimates. For a description of the different models and factors, see Section 3. The p-values before the comma are associated with the null hypothesis that the matrix containing a vector of ones and the factor loadings has a reduced rank. The p-values after the comma are associated with the null hypothesis that the matrix containing the factor loadings has a reduced rank. The testing portfolios are deciles sorted on investment (INV10), operating profitability (OP10), momentum (M10), and book-to-market (BM10). "All" refers to the estimation including all 40 portfolios. "All*" refers to the estimation containing the first two and last two deciles in each group plus the market factor (for a total of 17 portfolios). The sample is 1970:I to 2017:III.

	INV10	OP10	M10	BM10	All	All*
CAPM	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000
JW2	0.446, 0.277	0.510, 0.336	0.120, 0.112	0.208, 0.225	0.168, 0.155	0.084, 0.097
JW3	0.507, 0.606	0.630, 0.546	0.390, 0.457	0.950, 0.944	0.493, 0.335	0.707, 0.564
SV2	0.132, 0.183	0.000, 0.000	0.075, 0.036	0.575, 0.671	0.000, 0.000	0.000, 0.000
SV3	0.800, 0.492	0.613, 0.656	0.363, 0.420	0.883, 0.812	0.590, 0.634	0.759, 0.776
SV4	0.850, 0.638	0.589, 0.649	0.435, 0.375	0.817, 0.758	0.560, 0.605	0.839, 0.795
LL2	0.191, 0.265	0.001, 0.001	0.018, 0.031	0.680, 0.769	0.000, 0.000	0.001, 0.002
AEM1	0.203, 0.067	0.754, 0.653	0.746, 0.070	0.047, 0.015	0.174, 0.042	0.012, 0.000
AEM2	0.164, 0.081	0.693, 0.772	0.956, 0.091	0.059, 0.019	0.220, 0.049	0.033, 0.003
HKM1	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000
HKM2	0.012, 0.004	0.042, 0.057	0.224, 0.196	0.000, 0.000	0.000, 0.000	0.000, 0.000
CCAPM	0.147, 0.019	0.139, 0.019	0.561, 0.130	0.242, 0.124	0.107, 0.049	0.334, 0.035
LVN3	0.380, 0.427	0.895, 0.941	0.991, 0.949	0.616, 0.671	0.904, 0.922	0.816, 0.843

A Internet Appendix: Proofs

A.1 Proof of Proposition 1

Proof. The $K+1$ OLS normal equations for the baseline CSR,

$$\boldsymbol{\mu} = \widehat{\lambda}_0 \mathbf{1} + \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\lambda}} + \widehat{\boldsymbol{\omega}} \Leftrightarrow \boldsymbol{\mu} = \boldsymbol{W} \begin{bmatrix} \widehat{\lambda}_0 \\ \widehat{\boldsymbol{\lambda}} \end{bmatrix} + \widehat{\boldsymbol{\omega}},$$

are

$$\boldsymbol{W}'\boldsymbol{\mu} \;=\; \boldsymbol{W}'\boldsymbol{W} \begin{bmatrix} \widehat{\lambda}_0 \\ \widehat{\boldsymbol{\lambda}} \end{bmatrix}.$$

For the augmented CSR,

$$\boldsymbol{\mu} = \widehat{\theta}_0 \mathbf{1} + \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\theta}} + \widehat{\boldsymbol{\rho}}\widehat{\gamma} + \widehat{\boldsymbol{\varpi}} \;\Leftrightarrow\; \boldsymbol{\mu} \;=\; \boldsymbol{W} \begin{bmatrix} \widehat{\theta}_0 \\ \widehat{\boldsymbol{\theta}} \end{bmatrix} \;+\; \widehat{\boldsymbol{\rho}}\widehat{\gamma} \;+\; \widehat{\boldsymbol{\varpi}},$$

the first $K+1$ OLS normal equations are

$$\boldsymbol{W}'\boldsymbol{\mu} \;=\; \boldsymbol{W}'\boldsymbol{W} \begin{bmatrix} \widehat{\theta}_0 \\ \widehat{\boldsymbol{\theta}} \end{bmatrix} \;+\; \boldsymbol{W}'\widehat{\boldsymbol{\rho}}\widehat{\gamma}.$$

Equate the two expressions for $\boldsymbol{W}'\boldsymbol{\mu}$ and premultiply by $(\boldsymbol{W}'\boldsymbol{W})^{-1}$ (Assumption 2) to obtain

$$\begin{bmatrix} \widehat{\lambda}_0 \\ \widehat{\boldsymbol{\lambda}} \end{bmatrix} \;=\; \begin{bmatrix} \widehat{\theta}_0 \\ \widehat{\boldsymbol{\theta}} \end{bmatrix} \;+\; (\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\widehat{\boldsymbol{\rho}}\widehat{\gamma}. \tag{A.1}$$

By the OLS projection of $\widehat{\boldsymbol{\rho}}$ on $\boldsymbol{W}$, one has:

$$(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\widehat{\boldsymbol{\rho}} \;=\; \begin{bmatrix} \widehat{\boldsymbol{\delta}}_0' \\ \widehat{\boldsymbol{\Delta}} \end{bmatrix}, \quad \widehat{\boldsymbol{\delta}}_0 \in \mathbb{R}^L,\; \widehat{\boldsymbol{\Delta}} \in \mathbb{R}^{K\times L}.$$

Substituting into (A.1) yields (6). □

A.2 Proof of Proposition 2

Proof. The first identity in (6) gives $\widehat{\lambda}_0 = \widehat{\theta}_0 + \widehat{\boldsymbol{\delta}}_0'\widehat{\boldsymbol{\gamma}}$. Substitute the expression for $\widehat{\boldsymbol{\delta}}_0$ from (5) to obtain

$$\widehat{\lambda}_0 \;=\; \widehat{\theta}_0 \;+\; \widehat{\boldsymbol{\gamma}}'\mathbb{E}_n(\widehat{\boldsymbol{\rho}}_i) \;-\; \boldsymbol{m}'\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\gamma}}.$$

Finally, use the second identity in (6), i.e., $\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\gamma}} = \widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\theta}}$, which yields the claimed representation. □

A.3 Proof of Lemma 1

Proof. Let

$$\boldsymbol{v} = \mathbb{E}_n(\widehat{\boldsymbol{\beta}}_i\mu_i).$$

From the normal equations of the restricted and unrestricted CSRs we have

$$\boldsymbol{M}\widetilde{\boldsymbol{\lambda}} = \boldsymbol{v}, \qquad \boldsymbol{M}\widehat{\boldsymbol{\lambda}} + \boldsymbol{m}\,\widehat{\lambda}_0 = \boldsymbol{v}.$$

Subtracting yields

$$\boldsymbol{M}(\widehat{\boldsymbol{\lambda}} - \widetilde{\boldsymbol{\lambda}}) = -\boldsymbol{m}\,\widehat{\lambda}_0.$$

Since $\boldsymbol{M} \succ 0$ under Assumption 2, premultiplying by $\boldsymbol{M}^{-1}$ gives

$$\widehat{\boldsymbol{\lambda}} - \widetilde{\boldsymbol{\lambda}} = -\widehat{\lambda}_0\,\boldsymbol{M}^{-1}\boldsymbol{m} = -\widehat{\lambda}_0\,\boldsymbol{u}.$$

□

A.4 Proof of Proposition 3

Proof. The unrestricted intercept normal equation is

$$\overline{\mu} = \widehat{\lambda}_0 + \boldsymbol{m}'\widehat{\boldsymbol{\lambda}}.$$

By Lemma 1,

$$\widehat{\boldsymbol{\lambda}} - \widetilde{\boldsymbol{\lambda}} = -\widehat{\lambda}_0\,\boldsymbol{u},$$

and combining the two equations gives

$$\overline{\mu} = \widehat{\lambda}_0 + \boldsymbol{m}'(\widetilde{\boldsymbol{\lambda}} - \widehat{\lambda}_0 \boldsymbol{u}) = \boldsymbol{m}'\widetilde{\boldsymbol{\lambda}} + (1-s)\widehat{\lambda}_0.$$

Averaging the restricted CSR (7) across i gives

$$\overline{\mu} = \boldsymbol{m}'\widetilde{\boldsymbol{\lambda}} + \mathbb{E}_n(\widehat{\nu}_i).$$

Finally, subtracting the two identities for $\overline{\mu}$ yields,

$$\widehat{\lambda}_0 = \frac{\mathbb{E}_n(\widehat{\nu}_i)}{1-s} = \frac{\overline{\mu} - \boldsymbol{m}'\widetilde{\boldsymbol{\lambda}}}{1-s},$$

where the last equality follows from the second identity for $\overline{\mu}$.

□

A.5 Proof of Proposition 4

Proof. From $\widehat{\alpha}_i = \widehat{\omega}_i + \widehat{\lambda}_0$,

$$\mathbb{E}_n\left(\widehat{\alpha}_i^2\right) = \mathbb{E}_n\left(\widehat{\omega}_i^2\right) + 2\widehat{\lambda}_0\,\mathbb{E}_n\left(\widehat{\omega}_i\right) + \widehat{\lambda}_0^2 = \mathbb{E}_n\left(\widehat{\omega}_i^2\right) + \widehat{\lambda}_0^2,$$

since $\mathbb{E}_n(\widehat{\omega}_i) = 0$ by the OLS normal equations for the unrestricted CSR. Divide by $\mathrm{Var}_n(\mu_i)$ to obtain (11). □

A.6 Proof of Corollary 1

Proof. From Proposition 4, $R^2_{\text{pen}} = R^2_{\text{ols}} - \widehat{\lambda}_0^2/\mathrm{Var}_n(\mu_i)$. Since $R^2_{\text{ols}} \in [0,1]$ for the OLS regression with an intercept, it follows that $R^2_{\text{pen}} \leq 1 - \widehat{\lambda}_0^2/\mathrm{Var}_n(\mu_i) = 1 - \text{Wedge}$ and $R^2_{\text{pen}} \geq -\widehat{\lambda}_0^2/\mathrm{Var}_n(\mu_i) = -\text{Wedge}$. Moreover, $R^2_{\text{pen}} < 0$ iff $R^2_{\text{ols}} < \widehat{\lambda}_0^2/\mathrm{Var}_n(\mu_i)$, equivalently iff $\widehat{\lambda}_0^2 > \mathrm{Var}_n(\mu_i)\, R^2_{\text{ols}}$. □

A.7 Proof of Corollary 2

Proof. Define

$$\widehat{\nu}_i = \mu_i - \widetilde{\boldsymbol{\lambda}}'\widehat{\boldsymbol{\beta}}_i, \qquad \widehat{\alpha}_i = \mu_i - \widehat{\boldsymbol{\lambda}}'\widehat{\boldsymbol{\beta}}_i.$$

By Lemma 1,

$$\widehat{\boldsymbol{\lambda}} - \widetilde{\boldsymbol{\lambda}} = -\widehat{\lambda}_0\, \boldsymbol{u}.$$

Hence

$$\widehat{\alpha}_i = \mu_i - \widetilde{\boldsymbol{\lambda}}'\widehat{\boldsymbol{\beta}}_i + \widehat{\lambda}_0\, \widehat{\boldsymbol{\beta}}_i'\boldsymbol{u} = \widehat{\nu}_i + \widehat{\lambda}_0\, \widehat{\boldsymbol{\beta}}_i'\boldsymbol{u}.$$

By the normal equations for the restricted CSR,

$$\mathbb{E}_n(\widehat{\boldsymbol{\beta}}_i\widehat{\nu}_i) = \mathbf{0}.$$

Therefore

$$\begin{aligned}
\mathbb{E}_n(\widehat{\alpha}_i^2) &= \mathbb{E}_n(\widehat{\nu}_i^2) + 2\widehat{\lambda}_0\, \boldsymbol{u}'\mathbb{E}_n(\widehat{\boldsymbol{\beta}}_i\widehat{\nu}_i) + \widehat{\lambda}_0^2\, \mathbb{E}_n\big[(\widehat{\boldsymbol{\beta}}_i'\boldsymbol{u})^2\big] \\
&= \mathbb{E}_n(\widehat{\nu}_i^2) + \widehat{\lambda}_0^2\, \boldsymbol{u}'\boldsymbol{M}\boldsymbol{u} \\
&= \mathbb{E}_n(\widehat{\nu}_i^2) + s\, \widehat{\lambda}_0^2.
\end{aligned}$$

Dividing by $\mathrm{Var}_n(\mu_i)$ yields

$$R^2_{\mathrm{ols}*} - R^2_{\mathrm{pen}} = s\, \frac{\widehat{\lambda}_0^2}{\mathrm{Var}_n(\mu_i)} \geq 0.$$

Finally, $R^2_{\mathrm{ols}*} = R^2_{\mathrm{pen}}$ iff $s\widehat{\lambda}_0^2 = 0$, i.e. iff $\widehat{\lambda}_0 = 0$ or $s = 0$. Since $\boldsymbol{M} \succ 0$, $s = \boldsymbol{m}'\boldsymbol{M}^{-1}\boldsymbol{m} = 0$ iff $\boldsymbol{m} = \mathbf{0}$. □

A.8 Proof of Corollary 3

Proof. By Corollary 2,

$$R^2_{\mathrm{pen}} < 0 < R^2_{\mathrm{ols}*} \quad \Longleftrightarrow \quad 0 < R^2_{\mathrm{ols}*} < s\frac{\widehat{\lambda}_0^2}{\mathrm{Var}_n(\mu_i)}.$$

Substituting Proposition 3 yields

$$R^2_{\mathrm{ols}*} < \frac{s}{(1-s)^2}\frac{\mathbb{E}_n(\widehat{\nu}_i)^2}{\mathrm{Var}_n(\mu_i)}.$$

Since

$$R^2_{\mathrm{ols}*} = 1 - \frac{\mathbb{E}_n(\widehat{\nu}_i^2)}{\mathrm{Var}_n(\mu_i)},$$

the condition $0 < R^2_{\text{ols}*} < s\widehat{\lambda}_0^2/\text{Var}_n(\mu_i)$ is equivalent to

$$\mathbb{E}_n(\widehat{\nu}_i^2) < \text{Var}_n(\mu_i) < \mathbb{E}_n(\widehat{\nu}_i^2) + s\widehat{\lambda}_0^2.$$

□

A.9 Proof of Proposition 5

Proof. The result follows immediately from $\widehat{\alpha}_i = \widehat{\omega}_i + \widehat{\lambda}_0$ and $\mathbb{E}_n(\widehat{\omega}_i) = 0$. □

A.10 Proof of Proposition 6

Proof. We have

$$\boldsymbol{\Sigma}\,\widehat{\boldsymbol{\lambda}} = (\boldsymbol{M} - \boldsymbol{m}\boldsymbol{m}')\widehat{\boldsymbol{\lambda}} = (\boldsymbol{M}\widehat{\boldsymbol{\lambda}} + \boldsymbol{m}\,\widehat{\lambda}_0) - \boldsymbol{m}(\widehat{\lambda}_0 + \boldsymbol{m}'\widehat{\boldsymbol{\lambda}}) = \boldsymbol{v} - \boldsymbol{m}\,\overline{\mu},$$

where the last equality follows from the normal equations for the unrestricted CSR, $\boldsymbol{v} = \boldsymbol{M}\widehat{\boldsymbol{\lambda}} + \boldsymbol{m}\,\widehat{\lambda}_0$. Using the normal equations for the restricted CSR, $\boldsymbol{v} = \boldsymbol{M}\widetilde{\boldsymbol{\lambda}}$, yields

$$\boldsymbol{\Sigma}\,\widehat{\boldsymbol{\lambda}} = \boldsymbol{M}\widetilde{\boldsymbol{\lambda}} - \boldsymbol{m}\,\overline{\mu}.$$

Premultiplying by $\boldsymbol{\Sigma}^{-1}$ gives

$$\widehat{\boldsymbol{\lambda}} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{M}\widetilde{\boldsymbol{\lambda}} - \boldsymbol{m}\overline{\mu}).$$

Finally, using Lemma 1 and Proposition 3 yields:

$$\widehat{\boldsymbol{\lambda}} = \widetilde{\boldsymbol{\lambda}} + \frac{\boldsymbol{m}'\widetilde{\boldsymbol{\lambda}} - \overline{\mu}}{1 - s}\,\boldsymbol{u}.$$

□

A.11 Proof of Corollary 4

Proof. From (14), $\widehat{\boldsymbol{\lambda}} - \widetilde{\boldsymbol{\lambda}} = -\widehat{\lambda}_0\,\boldsymbol{u}$. Under $\boldsymbol{M} \succ 0$, $\boldsymbol{u} = \boldsymbol{M}^{-1}\boldsymbol{m} = \boldsymbol{0}$ if and only if $\boldsymbol{m} = \boldsymbol{0}$. Hence $\widehat{\boldsymbol{\lambda}} = \widetilde{\boldsymbol{\lambda}}$ if and only if either $\widehat{\lambda}_0 = 0$ or $\boldsymbol{m} = \boldsymbol{0}$. By Proposition 3, $\widehat{\lambda}_0 = 0$ is equivalent to $\mathbb{E}_n(\widehat{\nu}_i) = 0$ in the restricted CSR. □

A.12 Proof of Corollary 5

Proof. Fix $j \in \{1, \dots, K\}$. From (14), $\widehat{\lambda}_j = \widetilde{\lambda}_j + \Delta_j$ with $\Delta_j = -\widehat{\lambda}_0 u_j$.

(i) Attenuation without sign change. If $\widetilde{\lambda}_j > 0$, attenuation means $0 < \widehat{\lambda}_j < \widetilde{\lambda}_j$, i.e., $0 < \widetilde{\lambda}_j + \Delta_j < \widetilde{\lambda}_j$, which is equivalent to $-\widetilde{\lambda}_j < \Delta_j < 0$. If $\widetilde{\lambda}_j < 0$, attenuation means $\widetilde{\lambda}_j < \widehat{\lambda}_j < 0$, i.e., $\widetilde{\lambda}_j < \widetilde{\lambda}_j + \Delta_j < 0$, which is equivalent to $0 < \Delta_j < -\widetilde{\lambda}_j$.

(ii) Sign flip. A sign flip occurs iff $\widehat{\lambda}_j \widetilde{\lambda}_j < 0$, equivalently iff $(\widetilde{\lambda}_j + \Delta_j)\widetilde{\lambda}_j < 0$. Solving this inequality yields $\Delta_j < -\widetilde{\lambda}_j$ when $\widetilde{\lambda}_j > 0$ and $\Delta_j > -\widetilde{\lambda}_j$ when $\widetilde{\lambda}_j < 0$. □

A.13 Proof of Corollary 6

Proof. When $K = 1$, we have $\boldsymbol{u} = m/M$, $s = m^2/M$, and

$$\boldsymbol{\Sigma} = \sigma^2 = M - m^2 > 0.$$

Lemma 1 gives

$$\widehat{\lambda} - \widetilde{\lambda} = -\widehat{\lambda}_0 \frac{m}{M},$$

Proposition 3 yields

$$\widehat{\lambda}_0 = \frac{\overline{\mu} - m\widetilde{\lambda}}{1 - m^2/M},$$

and Proposition 6 gives

$$\sigma^2 \widehat{\lambda} = M\widetilde{\lambda} - m\overline{\mu}.$$

Therefore

$$\widehat{\lambda} = \frac{M\widetilde{\lambda} - m\overline{\mu}}{\sigma^2}, \qquad \widehat{\lambda} - \widetilde{\lambda} = \frac{m(\widetilde{\lambda} m - \overline{\mu})}{\sigma^2}.$$

Assume $\widetilde{\lambda} > 0$ and $m > 0$. A sign flip ($\widehat{\lambda} < 0$) occurs iff

$$M\widetilde{\lambda} - m\overline{\mu} < 0 \quad \Longleftrightarrow \quad \widetilde{\lambda} < \frac{\overline{\mu} m}{M}.$$

Attenuation without sign change ($0 < \widehat{\lambda} < \widetilde{\lambda}$) occurs iff

$$0 < \frac{M\widetilde{\lambda} - m\overline{\mu}}{\sigma^2} < \widetilde{\lambda},$$

which is equivalent to

$$\frac{\overline{\mu}m}{M} < \widetilde{\lambda} < \frac{\overline{\mu}}{m}.$$

The remaining cases follow by symmetry. □

A.14 Proof of Corollary 7

Proof. By the Frisch–Waugh–Lovell theorem, the slope coefficients from regressing $\boldsymbol{\mu}$ on $[\mathbf{1}, \widehat{\boldsymbol{\beta}}]$ equal the OLS coefficients from regressing $\boldsymbol{M}_1\boldsymbol{\mu}$ on $\boldsymbol{M}_1\widehat{\boldsymbol{\beta}}$. These coefficients are unique if and only if $\boldsymbol{M}_1\widehat{\boldsymbol{\beta}}$ has full column rank, equivalently $\widehat{\boldsymbol{\beta}}'\boldsymbol{M}_1\widehat{\boldsymbol{\beta}}$ is nonsingular. □

A.15 Proof of Lemma 2

Proof. The proof is the GLS analogue of the proof of Lemma 1. The restricted and unrestricted GLS normal equations imply

$$\boldsymbol{M}_{\text{gls}}\widetilde{\boldsymbol{\lambda}}_{\text{gls}} = \boldsymbol{v}_{\text{gls}}, \qquad \boldsymbol{M}_{\text{gls}}\widehat{\boldsymbol{\lambda}}_{\text{gls}} + \boldsymbol{m}_{\text{gls}}\widehat{\lambda}_{0,\text{gls}} = \boldsymbol{v}_{\text{gls}}.$$

Subtracting gives

$$\boldsymbol{M}_{\text{gls}}(\widehat{\boldsymbol{\lambda}}_{\text{gls}} - \widetilde{\boldsymbol{\lambda}}_{\text{gls}}) = -\boldsymbol{m}_{\text{gls}}\widehat{\lambda}_{0,\text{gls}}.$$

Premultiplying by $\boldsymbol{M}_{\text{gls}}^{-1}$ yields

$$\widehat{\boldsymbol{\lambda}}_{\text{gls}} - \widetilde{\boldsymbol{\lambda}}_{\text{gls}} = -\widehat{\lambda}_{0,\text{gls}}\boldsymbol{M}_{\text{gls}}^{-1}\boldsymbol{m}_{\text{gls}} = -\widehat{\lambda}_{0,\text{gls}}\boldsymbol{u}_{\text{gls}}.$$

□

A.16 Proof of Proposition 7

Proof. The proof follows the proof of Proposition 3 after replacing ordinary cross-sectional moments by $\boldsymbol{C}$-weighted moments. The unrestricted intercept normal equation is

$$\overline{\mu}_{\text{gls}} = \widehat{\lambda}_{0,\text{gls}} + \boldsymbol{m}'_{\text{gls}}\widehat{\boldsymbol{\lambda}}_{\text{gls}}.$$

Using Lemma 2,

$$\overline{\mu}_{\text{gls}} = \widehat{\lambda}_{0,\text{gls}} + \boldsymbol{m}'_{\text{gls}}(\widetilde{\boldsymbol{\lambda}}_{\text{gls}} - \widehat{\lambda}_{0,\text{gls}}\boldsymbol{u}_{\text{gls}}) = \boldsymbol{m}'_{\text{gls}}\widetilde{\boldsymbol{\lambda}}_{\text{gls}} + (1 - s_{\text{gls}})\widehat{\lambda}_{0,\text{gls}}.$$

Averaging the restricted GLS CSR gives

$$\overline{\mu}_{\text{gls}} = \boldsymbol{m}_{\text{gls}}'\widetilde{\boldsymbol{\lambda}}_{\text{gls}} + \mathbb{E}_{n,\text{gls}}(\widehat{\nu}_{i,\text{gls}}).$$

Equating the two expressions for $\overline{\mu}_{\text{gls}}$ gives the stated identity. □

A.17 Proof of Proposition 8

Proof. The proof is the GLS analogue of the proof of Proposition 4. Since $\widehat{\alpha}_{i,\text{gls}} = \widehat{\omega}_{i,\text{gls}} + \widehat{\lambda}_{0,\text{gls}}$ and $\mathbb{E}_{n,\text{gls}}(\widehat{\omega}_{i,\text{gls}}) = 0$,

$$\mathbb{E}_{n,\text{gls}}(\widehat{\alpha}_{i,\text{gls}}^2) = \mathbb{E}_{n,\text{gls}}(\widehat{\omega}_{i,\text{gls}}^2) + \widehat{\lambda}_{0,\text{gls}}^2.$$

Dividing by $\text{Var}_{n,\text{gls}}(\mu_i)$ yields the identity for $R^2_{\text{pen,gls}}$. □

A.18 Proof of Corollary 8

Proof. The proof is the GLS analogue of the proof of Corollary 2. By Lemma 2,

$$\widehat{\alpha}_{i,\text{gls}} = \mu_i - \widehat{\boldsymbol{\lambda}}_{\text{gls}}'\widehat{\boldsymbol{\beta}}_i = \widehat{\nu}_{i,\text{gls}} + \widehat{\lambda}_{0,\text{gls}}\widehat{\boldsymbol{\beta}}_i'\boldsymbol{u}_{\text{gls}}.$$

By the restricted GLS normal equations, $\mathbb{E}_{n,\text{gls}}(\widehat{\boldsymbol{\beta}}_i\widehat{\nu}_{i,\text{gls}}) = \mathbf{0}$. Therefore,

$$\mathbb{E}_{n,\text{gls}}(\widehat{\alpha}_{i,\text{gls}}^2) = \mathbb{E}_{n,\text{gls}}(\widehat{\nu}_{i,\text{gls}}^2) + \widehat{\lambda}_{0,\text{gls}}^2\boldsymbol{u}_{\text{gls}}'\boldsymbol{M}_{\text{gls}}\boldsymbol{u}_{\text{gls}} = \mathbb{E}_{n,\text{gls}}(\widehat{\nu}_{i,\text{gls}}^2) + s_{\text{gls}}\widehat{\lambda}_{0,\text{gls}}^2.$$

Dividing by $\text{Var}_{n,\text{gls}}(\mu_i)$ gives the result. □

B Internet Appendix: Definition of Factors

In this appendix, we describe the construction of the factors. Starting with the quarterly factors, consumption and wage data are obtained from the National Income and Product Accounts (NIPA) database. In line with the macro-finance literature, consumption is the sum of personal consumption expenditures on nondurable goods and services. Following Jagannathan and Wang (1996), aggregate labor income represents total personal income minus dividend income. We deflate consumption and labor income by both total population and the Consumer Price Index (CPI) in order to compute both per capita real consumption and per capita real labor income.[42]

[42] For convenience, we use the same price deflator for both macro variables.

We use the standard "end-of-period" timing assumption, in which the consumption growth over the period from t to $t+1$ is matched with asset returns calculated over the same time interval. The quarterly default spread (def) represents the yield spread between Moody's Baa- and Aaa-rated corporate bonds from the St. Louis Fed data library. The conditioning variables my, sw, and cay follow the construction in Lustig and Van Nieuwerburgh (2005), Santos and Veronesi (2006), and Lettau and Ludvigson (2001), respectively. The data on both LEV and IC are obtained from the corresponding authors' web pages. Finally, we employ the quarterly market factor from Lu Zhang's q-data library.

Turning to the monthly factors, the data on RM, SMB, HML, RMW, CMA, and UMD are from Kenneth French's website. The data on the default spread (DS), term spread (TS), and aggregate dividend-to-price ratio (DY) are retrieved from Amit Goyal's webpage and correspond to the variables employed in Goyal, Welch, and Zafirov (2024). TB is the monthly return on the 3-month T-bill from the St. Louis Fed.[43] Following Hahn and Lee (2006) and Maio and Santa-Clara (2012), ΔTS, ΔDS, ΔDY, and ΔTB represent the first differences in the original variables. MP denotes the log growth in the industrial production index, while the inflation rate (INF) is the log growth in the CPI index. EI represents the fitted value of an AR(1) regression applied to INF, while UI represents the respective residual. The data on both industrial production and CPI are from the St. Louis Fed.

C Internet Appendix: Additional Results

This appendix tabulates the summary statistics for portfolio risk premia (Table A.1) and the risk price estimates for the cross-section of 17 portfolios (Tables A.2 and A.3).

[43] Specifically, $TB = TB^*/12$, where TB^* denotes the original (annualized) T-bill rate.

Table A.1: Descriptive Statistics for Asset Risk Premia

This table reports cross-sectional descriptive statistics for the test portfolio risk premia. The portfolios represent deciles sorted on investment (INV10), operating profitability (OP10), momentum (M10), and book-to-market (BM10). "All" refers to a cross-section including all 40 portfolios. The sample is 1970:I to 2017:III (Panel A) or 1970:01 to 2017:09 (Panel B).

	Mean (%)	Stdev. (%)	Min. (%)	Max. (%)
Panel A: Quarterly Risk Premia				
INV10	1.944	0.438	1.051	2.557
OP10	1.687	0.326	1.047	2.149
M10	1.688	0.886	−0.332	3.146
BM10	2.084	0.448	1.337	2.938
All	1.851	0.571	−0.332	3.146
Panel B: Monthly Risk Premia				
INV10	0.622	0.151	0.297	0.832
OP10	0.537	0.119	0.287	0.692
M10	0.534	0.312	−0.188	1.019
BM10	0.668	0.144	0.412	0.928
All	0.590	0.198	−0.188	1.019

Table A.2: Quarterly Risk Price Estimates: 17 Portfolios

This table presents the quarterly risk price estimates. The method is the standard two-step regression approach. The testing portfolios are the first two and last two deciles for portfolios sorted on investment (INV10), operating profitability (OP10), momentum (M10), and book-to-market (BM10), plus the market factor (for a total of 17 portfolios). λ_0 denotes the intercept estimate (in %). λ denote the risk price estimates (in %), while the numbers in parentheses represent the respective Shanken t-ratios (with * denoting statistical significance at the 5% level). The risk price estimates and t-ratios displayed after the comma are associated with the corresponding CSR with no intercept. $\mathbb{E}_n(\widehat{\nu}_i)$ denotes the average pricing errors (in %) of the restricted CSR. For a description of the different models and factors, see Section 3. The sample is 1970:I to 2017:III.

	λ_0	λ_{RM}	λ_Y	λ_{def}	λ_{RMsw}	λ_{Ysw}	λ_{RMcay}	$\mathbb{E}_n(\widehat{\nu}_i)$
CAPM	5.88	$-3.64, 1.52$						0.12
	(5.74*)	$(-3.10^*, 2.38^*)$						
JW2	3.98	$-1.80, 1.89$	$0.82, 1.79$					0.04
	(3.52*)	$(-1.42, 2.88^*)$	$(2.08^*, 2.71^*)$					
JW3	3.92	$-1.74, 1.93$	$0.77, 1.53$	$-0.10, -0.24$				0.04
	(3.18*)	$(-1.27, 2.96^*)$	$(2.24^*, 2.47^*)$	$(-0.35, -0.65)$				
SV2	6.16	$-3.88, 1.52$			$-0.09, 0.12$			0.11
	(5.30*)	$(-3.04^*, 2.36^*)$			$(-1.21, 1.91)$			
SV3	4.02	$-1.91, 1.82$	$0.70, 1.69$			$0.02, 0.02$		0.04
	(3.37*)	$(-1.44, 2.71^*)$	$(1.98^*, 2.94^*)$			$(0.88, 0.57)$		
SV4	4.39	$-2.25, 1.82$	$0.62, 1.68$		$-0.03, 0.10$	$0.02, 0.02$		0.04
	(3.51*)	$(-1.63, 2.71^*)$	$(1.97^*, 2.92^*)$		$(-0.33, 0.89)$	$(0.90, 0.52)$		
LL2	5.63	$-3.38, 1.62$					$0.03, 0.11$	0.11
	(5.92*)	$(-3.05^*, 2.53^*)$					$(0.64, 1.91)$	
	λ_0	λ_{RM}	λ_{LEV}	λ_{IC}	λ_C	λ_{my}	λ_{Cmy}	$\mathbb{E}_n(\widehat{\nu}_i)$
AEM1	0.83		$13.18, 21.03$					0.24
	(0.70)		$(3.05^*, 1.70)$					
AEM2	3.54	$-1.49, 1.67$	$8.79, 15.60$					0.04
	(2.12*)	$(-0.87, 2.55^*)$	$(1.98^*, 4.21^*)$					
HKM1	4.33			$-0.04, 0.03$				0.13
	(4.29*)			$(-2.11^*, 2.37^*)$				
HKM2	5.91	$-3.71, 1.49$		$-0.03, 0.03$				0.12
	(5.76*)	$(-3.18^*, 2.33^*)$		$(-1.31, 1.70)$				
CCAPM	-0.78				$1.50, 1.07$			-0.04
	(-0.30)				$(1.81, 1.66)$			
LVN3	1.41				$0.68, 1.14$	$-21.57, -18.62$	$0.02, -0.01$	0.02
	(0.72)				$(0.87, 2.31^*)$	$(-1.77, -1.45)$	$(0.22, -0.04)$	

Table A.3: Monthly Risk Price Estimates: 17 Portfolios

This table presents the monthly risk price estimates. The method is the standard two-step regression approach. The testing portfolios are the first two and last two deciles for portfolios sorted on investment (INV10), operating profitability (OP10), momentum (M10), and book-to-market (BM10), plus the market factor (for a total of 17 portfolios). λ_0 denotes the intercept estimate (in %). λ denote the risk price estimates (in %), while the numbers in parentheses represent the respective Shanken t-ratios (with * denoting statistical significance at the 5% level). The risk price estimates and t-ratios displayed after the comma are associated with the corresponding CSR with no intercept. $\mathbb{E}_n(\widehat{\nu}_i)$ denotes the average pricing errors (in %) of the restricted CSR. For a description of the different models and factors, see Section 3. The sample is 1970:01 to 2017:09.

	λ_0	λ_{RM}	$\lambda_{\Delta TS}$	$\lambda_{\Delta DS}$	$\lambda_{\Delta DY}$	$\lambda_{\Delta TB}$	λ_{SMB}	λ_{HML}	$\mathbb{E}_n(\widehat{\nu}_i)$
CAPM	2.14	$-1.42, 0.49$							0.03
	(5.88*)	(−3.58*, 2.53*)							
HL3	2.92	$-2.30, 0.56$	$-0.00, -0.10$	$-0.22, 0.06$					0.03
	(3.26*)	(−2.51*, 2.91*)	(−0.02, −0.87)	(−1.71, 1.25)					
HL5	2.54	$-1.92, 0.54$	$-0.44, -1.02$	$0.01, 0.08$			$0.73, 0.12$	$0.42, 0.53$	0.01
	(3.12*)	(−2.31*, 2.67*)	(−1.81, −1.90)	(0.10, 0.67)			(1.75, 0.24)	(2.66*, 2.50*)	
P5	3.55	$-2.88, 0.59$	$0.06, 0.31$	$-0.10, -0.19$	$3.05, -0.67$	$0.01, -0.06$			0.01
	(3.82*)	(−3.09*, 2.88*)	(0.40, 1.04)	(−1.50, −1.23)	(3.00*, −2.23*)	(0.69, −1.84)			
P7	1.84	$-1.24, 0.56$	$-0.46, -0.59$	$-0.00, -0.01$	$1.31, -0.55$	$0.02, 0.00$	$0.73, 0.42$	$0.47, 0.61$	0.00
	(1.95)	(−1.31, 2.82*)	(−1.86, −1.73)	(−0.03, −0.13)	(1.29, −1.98*)	(1.01, 0.17)	(1.79, 0.92)	(2.58*, 2.85*)	
	λ_0	λ_{RM}	λ_{SMB}	λ_{HML}	λ_{MP}	λ_{RMW}	λ_{CMA}	λ_{UMD}	$\mathbb{E}_n(\widehat{\nu}_i)$
LZ1	0.75				$2.21, -2.30$				0.38
	(1.13)				(1.61, −0.53)				
LZ4	3.27	$-2.54, 0.68$	$0.51, -1.02$	$0.40, 0.48$	$1.28, 2.31$				0.02
	(2.50*)	(−1.98*, 3.22*)	(0.98, −1.52)	(2.30*, 2.14*)	(1.71, 1.44)				
FF6	1.73	$-1.13, 0.56$	$0.72, 0.08$	$0.38, 0.42$		$0.33, 0.33$	$0.28, 0.29$	$0.63, 0.65$	0.00
	(2.15*)	(−1.39, 2.97*)	(2.03*, 0.38)	(2.75*, 3.05*)		(2.78*, 2.81*)	(2.87*, 2.97*)	(3.26*, 3.38*)	
	λ_0	λ_{RM}	λ_{MP}	λ_{TS}	λ_{DS}	λ_{UI}	λ_{EI}		$\mathbb{E}_n(\widehat{\nu}_i)$
CRR5	1.09		$1.09, 0.34$	$-0.90, 2.18$	$-0.42, -0.45$	$-0.08, -0.51$	$0.19, 0.22$		0.04
	(2.49*)		(1.57, 0.31)	(−0.85, 0.93)	(−1.23, −0.86)	(−0.53, −1.95)	(1.27, 0.94)		
CRR6	1.88	$-1.15, 0.61$	$0.30, -1.15$	$-0.35, -0.50$	$1.07, 0.88$	$-0.03, -0.25$	$0.11, 0.28$		0.01
	(2.57*)	(−1.59, 3.04*)	(0.33, −1.00)	(−1.05, −1.25)	(1.72, 1.14)	(−0.25, −1.47)	(0.95, 1.54)		

D Internet Appendix: Quarterly Risk Price Estimates for Single Portfolio Groups

In this appendix, we present the quarterly risk price estimates for the single-anomaly estimations, whose results appear in Tables A.4, A.5, A.6, and A.7 below. One key difference relative to both All and All* is that several intercept estimates are insignificant. For example, in the estimation with BM10 none of the estimates of λ_0 is statistically significant despite the large magnitudes in the majority of the cases (i.e., $\widehat{\lambda}_0$ around or above 1% per quarter for 11 of the 13 models being estimated). On the other hand, eight of the 13 intercept estimates are also insignificant in the estimation with M10. One common misconception is that the intercept estimates "can be ignored if statistically insignificant". This argument conflates statistical significance with economic significance.[44] If anything, an insignificant estimate of λ_0 provides a strong statistical argument for ignoring the results from the unrestricted CSR and focusing exclusively on the restricted counterpart.

We can also see that the sign flip in the market risk price estimates is not as extreme as in the estimation with the 40 or 17 portfolios, since the sign does not actually reverse for five of the nine models (containing such a factor) in the estimations with M10 or BM10. Yet, the corresponding positive estimates of λ_{RM} are always insignificant, which compares with strongly positive estimates for the counterparts in the restricted CSR. Thus, and in a similar vein to the two augmented estimations, there is not a single case where the market risk price is significantly positive under the unrestricted CSR.

Notably, there is one pattern in the estimations with OP10 or BM10 that is even more extreme than in the estimation with the 40 or 17 portfolios: the risk price estimates associated with the unrestricted CSR are insignificant in all cases. In comparison, all the risk price estimates associated with RM (as mentioned above), LEV, and IC are significant under the restricted CSR for BM10. In the estimation with OP10, it turns out that the restricted estimates of λ_{RM} (all nine models), λ_{RMsw} (within SV2), λ_{IC} (in HKM1), λ_{C} (within LVN3), and λ_{RMcay} are all significant. This illustrates the substantial effect of the intercept on the statistical significance of the risk price

[44] In fact, to provide a comparison, researchers include all cross-sectional residuals in the computation of the cross-sectional R^2 or $RMSE$ irrespective of whether all of these residuals are individually statistically significant or not. Naturally, the same holds for the intercept, which represents the average pricing error.

estimates.

In a similar vein to the augmented estimations, AEM1 produces the largest positive mean pricing error ($\mathbb{E}_n(\widehat{\nu}_i) = 0.22\%$) in the estimation with BM10. Even with a "moderate" $s = 0.83$, Eq. (8) implies an intercept of about $0.2153\%/(1 - 0.8291) \approx 1.26\%$. This in turn leads to an economically large risk price shift (about 10 percentage points in magnitude) given the high value of $u = 8.01$: $7.97\% - 18.06\% \approx -10.09\% = -1.26\% \times 8.0109$. HKM1 when estimated over M10, represents another relevant case: the relatively large $\mathbb{E}_n(\widehat{\nu}_i) = 0.20\%$, combined with a high value of s (0.96), translates into a large intercept estimate (4.80%). This in turn leads to a sign flip in the risk price (from 0.03% to -0.05%, with both estimates being significant at the 5% level).

At the other end of the spectrum, AEM2 produces the smallest $\mathbb{E}_n(\widehat{\nu}_i)$ (around 0%), and thus the smallest intercept (only 0.06%), among the 13 models (in all single-anomaly estimations). Therefore, and in line with both Proposition 6 and Corollary 4, this is one of the models with the lowest gaps in risk prices: 1.81% versus 1.87% for λ_{RM} and 8.46% versus 8.50% for λ_{LEV}. A similar pattern holds for HKM1 (over BM10): the relatively small negative intercept estimate of -0.48% (due to a marginally negative $\mathbb{E}_n(\widehat{\nu}_i)$ of -0.01%) combined with a small $u \approx 0.02$ yields a gap in the risk price for IC of only $0.05\% - 0.04\% \approx 0.01\% = -(-0.48\%) \times 0.0176$.

Table A.4: Quarterly Risk Price Estimates: INV10

This table presents the quarterly risk price estimates. The method is the standard two-step regression approach. The testing portfolios are deciles sorted on investment (INV10). λ_0 denotes the intercept estimate (in %). λ denote the risk price estimates (in %), while the numbers in parentheses represent the respective Shanken t-ratios (with * denoting statistical significance at the 5% level). The risk price estimates and t-ratios displayed after the comma are associated with the corresponding CSR with no intercept. $\mathbb{E}_n(\widehat{\nu}_i)$ denotes the average pricing errors (in %) of the restricted CSR. For a description of the different models and factors, see Section 3. The sample is 1970:I to 2017:III.

	λ_0	λ_{RM}	λ_Y	λ_{def}	λ_{RMsw}	λ_{Ysw}	λ_{RMcay}	$\mathbb{E}_n(\widehat{\nu}_i)$
CAPM	3.30	$-1.36, 1.88$						0.07
	(3.95*)	$(-1.31, 2.97^*)$						
JW2	5.02	$-3.11, 1.89$	$-1.40, 0.33$					0.06
	(2.57*)	$(-1.55, 2.99^*)$	$(-1.74, 0.91)$					
JW3	5.04	$-3.14, 1.89$	$-1.16, 0.52$	$0.20, 0.09$				0.06
	(2.82*)	$(-1.68, 2.98^*)$	$(-1.98^*, 1.21)$	$(0.70, 0.44)$				
SV2	3.29	$-1.45, 1.78$			$0.10, 0.13$			0.07
	(3.72*)	$(-1.33, 2.80^*)$			$(0.94, 1.13)$			
SV3	5.26	$-3.37, 2.00$	$-1.42, -0.23$			$0.01, -0.07$		0.04
	(2.19*)	$(-1.36, 3.09^*)$	$(-1.66, -0.32)$			$(0.23, -1.94)$		
SV4	5.39	$-3.57, 2.05$	$-1.38, -0.27$		$0.06, -0.15$	$0.01, -0.07$		0.04
	(2.10*)	$(-1.31, 3.07^*)$	$(-1.68, -0.38)$		$(0.26, -0.61)$	$(0.38, -1.88)$		
LL2	3.26	$-1.45, 1.74$					$-0.10, -0.11$	0.07
	(3.43*)	$(-1.27, 2.74^*)$					$(-1.41, -1.44)$	
	λ_0	λ_{RM}	λ_{LEV}	λ_{IC}	λ_C	λ_{my}	λ_{Cmy}	$\mathbb{E}_n(\widehat{\nu}_i)$
AEM1	0.82		$14.30, 23.80$					0.08
	(0.64)		$(1.88, 1.58)$					
AEM2	0.83	$1.03, 1.83$	$14.29, 17.22$					0.01
	(0.56)	$(0.64, 2.88^*)$	$(2.02^*, 2.72^*)$					
HKM1	3.06			$-0.02, 0.04$				0.04
	(3.11*)			$(-0.90, 2.96^*)$				
HKM2	1.62	$0.21, 1.76$		$0.06, 0.10$				0.02
	(1.48)	$(0.17, 2.78^*)$		$(1.60, 3.52^*)$				
CCAPM	2.03				$-0.05, 1.01$			0.04
	(2.29*)				$(-0.19, 1.89)$			
LVN3	2.93				$-0.53, 0.75$	$-16.42, -16.77$	$-0.06, -0.07$	0.04
	(2.39*)				$(-1.15, 1.75)$	$(-1.76, -1.75)$	$(-0.84, -0.94)$	

Table A.5: Quarterly Risk Price Estimates: OP10

This table presents the quarterly risk price estimates. The method is the standard two-step regression approach. The testing portfolios are deciles sorted on operating profitability (OP10). λ_0 denotes the intercept estimate (in %). λ denote the risk price estimates (in %), while the numbers in parentheses represent the respective Shanken t-ratios (with * denoting statistical significance at the 5% level). The risk price estimates and t-ratios displayed after the comma are associated with the corresponding CSR with no intercept. $\mathbb{E}_n(\widehat{\nu}_i)$ denotes the average pricing errors (in %) of the restricted CSR. For a description of the different models and factors, see Section 3. The sample is 1970:I to 2017:III.

	λ_0	λ_{RM}	λ_Y	λ_{def}	λ_{RMsw}	λ_{Ysw}	λ_{RMcay}	$\mathbb{E}_n(\widehat{\nu}_i)$
CAPM	3.79	$-2.07, 1.60$						0.06
	(3.55*)	$(-1.66, 2.53^*)$						
JW2	3.42	$-1.69, 1.74$	$0.21, 1.49$					0.02
	(2.83*)	$(-1.23, 2.75^*)$	$(0.57, 1.94)$					
JW3	2.91	$-1.17, 1.78$	$0.28, 1.13$	$-0.19, -0.56$				0.01
	(1.95)	$(-0.71, 2.81^*)$	$(0.70, 1.51)$	$(-0.92, -1.70)$				
SV2	3.70	$-1.98, 1.69$			$-0.00, 0.17$			0.03
	(3.51*)	$(-1.61, 2.68^*)$			$(-0.07, 2.63^*)$			
SV3	3.21	$-1.49, 1.73$	$0.25, 1.43$			$0.02, 0.03$		0.02
	(2.36*)	$(-0.99, 2.71^*)$	$(0.57, 1.81)$			$(1.00, 0.89)$		
SV4	3.40	$-1.69, 1.73$	$0.40, 1.38$		$-0.01, 0.08$	$0.02, 0.03$		0.02
	(2.26*)	$(-1.03, 2.72^*)$	$(0.78, 1.91)$		$(-0.10, 0.83)$	$(1.22, 0.94)$		
LL2	4.21	$-2.49, 1.65$					$0.03, -0.12$	0.03
	(3.80*)	$(-1.95, 2.61^*)$					$(0.73, -2.16^*)$	
	λ_0	λ_{RM}	λ_{LEV}	λ_{IC}	λ_C	λ_{my}	λ_{Cmy}	$\mathbb{E}_n(\widehat{\nu}_i)$
AEM1	2.35		$-8.90, 21.60$					0.07
	(3.03*)		$(-0.92, 1.30)$					
AEM2	3.74	$-2.02, 1.59$	$3.06, 10.27$					0.05
	(3.36*)	$(-1.57, 2.51^*)$	$(0.52, 1.43)$					
HKM1	3.59			$-0.03, 0.03$				0.07
	(3.76*)			$(-1.69, 2.49^*)$				
HKM2	3.45	$-1.68, 1.79$		$-0.04, -0.04$				0.04
	(3.25*)	$(-1.35, 2.83^*)$		$(-1.53, -1.41)$				
CCAPM	1.58				$0.06, 0.91$			0.04
	(1.82)				$(0.25, 1.71)$			
LVN3	1.91				$0.15, 0.93$	$9.13, 6.46$	$0.09, 0.05$	0.03
	(1.76)				$(0.43, 2.12^*)$	$(0.94, 0.52)$	$(1.20, 0.53)$	

Table A.6: Quarterly Risk Price Estimates: M10

This table presents the quarterly risk price estimates. The method is the standard two-step regression approach. The testing portfolios are deciles sorted on momentum (M10). λ_0 denotes the intercept estimate (in %). λ denote the risk price estimates (in %), while the numbers in parentheses represent the respective Shanken t-ratios (with * denoting statistical significance at the 5% level). The risk price estimates and t-ratios displayed after the comma are associated with the corresponding CSR with no intercept. $\mathbb{E}_n(\widehat{\nu}_i)$ denotes the average pricing errors (in %) of the restricted CSR. For a description of the different models and factors, see Section 3. The sample is 1970:I to 2017:III.

	λ_0	λ_{RM}	λ_Y	λ_{def}	λ_{RMsw}	λ_{Ysw}	λ_{RMcay}	$\mathbb{E}_n(\widehat{\nu}_i)$
CAPM	4.53	$-2.73, 1.46$						0.17
	(4.40*)	$(-2.20^*, 2.24^*)$						
JW2	2.04	$0.27, 2.40$	$1.66, 2.28$					0.04
	(1.13)	$(0.14, 3.41^*)$	$(1.88, 1.92)$					
JW3	1.53	$0.74, 2.33$	$0.97, 1.18$	$-0.39, -0.50$				0.03
	(0.96)	$(0.43, 3.51^*)$	$(1.91, 1.82)$	$(-1.10, -1.32)$				
SV2	6.00	$-3.73, 1.42$			$-0.38, 0.05$			0.16
	(2.81*)	$(-1.78, 2.19^*)$			$(-1.60, 0.46)$			
SV3	1.37	$0.82, 2.22$	$1.71, 2.08$			$0.03, 0.04$		0.02
	(0.62)	$(0.36, 3.14^*)$	$(1.67, 1.99^*)$			$(0.76, 0.94)$		
SV4	1.61	$0.60, 2.13$	$1.64, 2.11$		$0.05, 0.20$	$0.03, 0.04$		0.01
	(0.70)	$(0.26, 2.95^*)$	$(2.11^*, 1.97^*)$		$(0.19, 0.81)$	$(1.10, 1.02)$		
LL2	3.55	$-1.34, 2.11$					$0.16, 0.22$	0.12
	(3.03*)	$(-1.02, 3.22^*)$					$(2.11^*, 2.37^*)$	
	λ_0	λ_{RM}	λ_{LEV}	λ_{IC}	λ_C	λ_{my}	λ_{Cmy}	$\mathbb{E}_n(\widehat{\nu}_i)$
AEM1	0.52		$16.97, 22.91$					0.12
	(0.31)		$(1.53, 1.63)$					
AEM2	-4.33	$6.00, 1.74$	$35.96, 18.82$					-0.01
	(-0.70)	$(1.00, 2.61^*)$	$(1.31, 2.76^*)$					
HKM1	4.80			$-0.05, 0.03$				0.20
	(4.34*)			$(-2.31^*, 2.17^*)$				
HKM2	4.20	$-2.12, 1.83$		$-0.08, -0.05$				0.15
	(3.60*)	$(-1.62, 2.83^*)$		$(-2.09^*, -1.41)$				
CCAPM	-1.05				$1.86, 1.19$			-0.07
	(-0.30)				$(1.27, 1.60)$			
LVN3	2.20				$-0.05, 1.10$	$-39.72, -23.76$	$-0.11, -0.06$	0.02
	(0.48)				$(-0.02, 1.63)$	$(-0.76, -0.86)$	$(-0.57, -0.55)$	

Table A.7: Quarterly Risk Price Estimates: BM10

This table presents the quarterly risk price estimates. The method is the standard two-step regression approach. The testing portfolios are deciles sorted on book-to-market (BM10). λ_0 denotes the intercept estimate (in %). λ denote the risk price estimates (in %), while the numbers in parentheses represent the respective Shanken t-ratios (with * denoting statistical significance at the 5% level). The risk price estimates and t-ratios displayed after the comma are associated with the corresponding CSR with no intercept. $\mathbb{E}_n(\widehat{\nu}_i)$ denotes the average pricing errors (in %) of the restricted CSR. For a description of the different models and factors, see Section 3. The sample is 1970:I to 2017:III.

	λ_0	λ_{RM}	λ_Y	λ_{def}	λ_{RMsw}	λ_{Ysw}	λ_{RMcay}	$\mathbb{E}_n(\widehat{\nu}_i)$
CAPM	1.58	0.52, 2.15						0.01
	(1.24)	(0.37, 3.31*)						
JW2	2.82	−0.79, 2.14	0.80, 0.25					0.01
	(1.57)	(−0.42, 3.29*)	(1.92, 0.69)					
JW3	4.33	−2.38, 2.14	0.21, −0.16	0.67, 0.35				0.02
	(1.24)	(−0.69, 3.24*)	(0.26, −0.26)	(0.78, 0.71)				
SV2	1.58	0.57, 2.19			−0.06, −0.06			0.01
	(1.21)	(0.40, 3.26*)			(−0.38, −0.32)			
SV3	2.57	−0.53, 2.16	0.75, 0.22			−0.00, −0.03		0.01
	(1.26)	(−0.25, 3.23*)	(1.55, 0.45)			(−0.19, −1.27)		
SV4	2.44	−0.25, 2.32	0.90, 0.40		−0.26, −0.31	−0.02, −0.05		0.01
	(0.92)	(−0.09, 2.88*)	(1.24, 0.56)		(−0.76, −0.74)	(−0.65, −1.17)		
LL2	1.44	0.26, 1.74					−0.32, −0.32	0.01
	(0.57)	(0.10, 2.63*)					(−1.34, −1.32)	
	λ_0	λ_{RM}	λ_{LEV}	λ_{IC}	λ_C	λ_{my}	λ_{Cmy}	$\mathbb{E}_n(\widehat{\nu}_i)$
AEM1	1.26		7.97, 18.06					0.22
	(1.55)		(1.80, 2.04*)					
AEM2	0.06	1.81, 1.87	8.46, 8.50					0.00
	(0.04)	(1.02, 2.93*)	(1.82, 1.99*)					
HKM1	−0.48			0.05, 0.04				−0.01
	(−0.38)			(1.94, 3.31*)				
HKM2	1.15	0.68, 1.86		0.03, 0.05				0.01
	(0.87)	(0.46, 2.91*)		(1.61, 3.10*)				
CCAPM	1.26				0.45, 1.10			0.03
	(1.62)				(1.26, 1.92)			
LVN3	0.94				0.61, 1.18	−10.95, −12.85	0.00, 0.04	0.02
	(0.96)				(1.42, 1.61)	(−1.09, −0.95)	(0.02, 0.31)	

E Internet Appendix: Bootstrap Simulation

E.1 Bootstrap Algorithm

We describe the bootstrap simulation used to assess the statistical significance of both R^2_{ols} and R^2_{pen}. We use the CCAPM to illustrate the procedure. The case of the restricted CSR and $R^2_{\text{ols}*}$ arises by setting $\lambda_0 = 0$. This simulation is similar to that employed in Maio and Min (2026), with a slightly different return generating process in the case of the unrestricted CSR.

1. First, we estimate the time-series regressions to obtain the factor loadings,

$$R_{i,t} = a_i + \beta_{i,c}\Delta c_t + \xi_{i,t},$$

and in a second step, the expected return-beta representation is estimated by an OLS CSR,

$$\mu_i = \lambda_0 + \lambda_c\widehat{\beta}_{i,c} + \omega_i,$$

where the intercept is excluded in the case of the restricted CSR. We compute and save the sample explanatory ratios $(R^2_{\text{ols}}, R^2_{\text{pen}})$. We also save the CSR residuals $(\widehat{\omega}_i)$. Furthermore, we run the following auxiliary time-series regressions,

$$R_{i,t} = \mu_i + \varsigma_{i,t},$$

and save the corresponding residuals $\varsigma_{i,t}$.

2. In each replication $b = 1, ..., 10000$, we construct pseudo-samples for both the time-series residuals (of each test asset) and the factor (of size T) by drawing with replacement,

$$\{\varsigma^b_{i,t}\}, i = 1, ..., N,$$
$$\{\Delta c^b_t\}, t = s^b_1, s^b_2, ..., s^b_T,$$

where the time indices $s^b_1, s^b_2, ..., s^b_T$ are created randomly from the original time sequence $1, ..., T$. Note that the time sequence is the same for all variables, thus we allow for a possible correlation between the test returns and the factor.

3. For each replication, the pseudo asset excess returns are constructed by imposing the null that the factor risk price is zero:

$$R_{i,t}^b = \widehat{\omega}_i + \varsigma_{i,t}^b.$$

4. In each replication, we estimate the CCAPM by the two-step procedure, but using the artificial data (rather than the original data):

$$\begin{aligned} R_{i,t}^b &= a_i^b + \beta_{i,c}^b \Delta c_t^b + \xi_{i,t}^b, \\ \mu_i^b &= \lambda_0^b + \lambda_c^b \widehat{\beta}_{i,c}^b + \omega_i^b, \end{aligned}$$

with the intercept being omitted in the case of the restricted CSR. We compute and save the pseudo cross-sectional R^2, leading to an empirical distribution of these statistics: $R^2_{\text{ols,b}}, R^2_{\text{pen,b}}, b = 1, ..., 10000$.

5. Finally, the empirical p-values associated with R^2_{ols} and R^2_{pen} are computed as

$$\begin{aligned} p(R^2_{\text{ols}}) &= \frac{\sum_{b=1}^{10000} \mathbb{I}(R^2_{\text{ols,b}} \geq R^2_{\text{ols}})}{10000}, \\ p(R^2_{\text{pen}}) &= \frac{\sum_{b=1}^{10000} \mathbb{I}(R^2_{\text{pen,b}} \geq R^2_{\text{pen}})}{10000}, \end{aligned}$$

where $\mathbb{I}(x)$ denotes the indicator function (which takes the value of one if x is true, and zero otherwise). $p(R^2_{\text{ols}*})$ is computed in a similar way.

E.2 Inspecting the Bootstrap

Monte Carlo experiments are conducted to enable us to assess the power and size of the bootstrap simulations associated with the three cross-sectional R^2: R^2_{pen}, $R^2_{\text{ols}*}$, and R^2_{ols}.[45] We consider All* as test assets and FF6 as the benchmark model.[46]

The CSR for FF6,

$$\begin{aligned} \mu_i &= \lambda_0 + \lambda_{RM}\widehat{\beta}_{i,RM} + \lambda_{SMB}\widehat{\beta}_{i,SMB} + \lambda_{HML}\widehat{\beta}_{i,HML} + \lambda_{RMW}\widehat{\beta}_{i,RMW} \\ &\quad + \lambda_{CMA}\widehat{\beta}_{i,CMA} + \lambda_{UMD}\widehat{\beta}_{i,UMD} + \omega_i, i = 1, .., 16, \end{aligned} \tag{E.2}$$

[45] Maio and Min (2026) conduct a similar analysis for an alternative R^2 metric under the restricted CSR.

[46] In these experiments, we exclude RM from the test assets in All*. This enables consistency between the OLS and GLS Monte Carlo results (see Appendix I for further details).

is estimated over the original sample (with and without intercept). In each of the 1000 pseudo samples, we implement the bootstrap simulation (with 1000 replications) and calculate the empirical p-value for the simulated R^2. Based on the distribution of the pseudo bootstrap p-values, we compute the fraction of these 1000 draws where the null, $R^2_{\text{pen}} = 0$, $R^2_{\text{ols}} = 0$, or $R^2_{\text{ols}*} = 0$, is rejected at the 5% level (i.e., the empirical p-value is below 5%).

The Monte Carlo simulation consists of several steps. First, we compute the residuals from the auxiliary regression,

$$R_{i,t} = \mu_i + \varsigma_{i,t}, \tag{E.3}$$

which decomposes the total excess return into the expected (μ_i) and unexpected excess return ($\varsigma_{i,t}$). Then, we simulate (jointly) the factors and the unexpected returns (with 1000 draws) by assuming that these variables are jointly normally distributed. We use the same sample size as the original data ($T = 573$), thus producing pseudo series for both the factors and residual portfolio returns, $\left\{RM_t^m, SMB_t^m, HML_t^m, RMW_t^m, CMA_t^m, UMD_t^m, \varsigma_{i,t}^m\right\}, i = 1, ..., 16, m = 1, ..., 1000.$

Different return generating processes are employed to evaluate the power and size of the bootstrap. To assess power, the pseudo excess return for portfolio i (in each draw) is defined as

$$R_{i,t}^m = \widehat{\lambda}_{RM}\widehat{\beta}_{i,RM} + \widehat{\lambda}_{SMB}\widehat{\beta}_{i,SMB} + \widehat{\lambda}_{HML}\widehat{\beta}_{i,HML} + \widehat{\lambda}_{RMW}\widehat{\beta}_{i,RMW} + \widehat{\lambda}_{CMA}\widehat{\beta}_{i,CMA} + \widehat{\lambda}_{UMD}\widehat{\beta}_{i,UMD} + \varsigma_{i,t}^m, \tag{E.4}$$

where the risk price estimates ($\widehat{\lambda}$) are obtained from (E.2). This return generating process is consistent with FF6 being true in the population. We obtain rejection rates (across the 1000 replications) around 100% for both R^2_{ols} and $R^2_{\text{ols}*}$, but only 18% for R^2_{pen}. This indicates a very weak power in the case of R^2_{pen}, which means that the statistical inference associated with the appropriate R^2 metric is highly unreliable when the CSR includes an intercept.

To assess the size of the bootstrap-based tests, the pseudo excess return for portfolio i is given by

$$R_{i,t}^m = \widehat{\omega}_i + \varsigma_{i,t}^m, \tag{E.5}$$

where $\widehat{\omega}_i$ denote the CSR residuals from (E.2). This return generating process is consistent with the null that the factors in FF6 have no pricing power in the population, as all risk prices are zero. We obtain a rejection rate of 5.2%, 5.1%, and 5.7% for R^2_{ols}, R^2_{pen}, and $R^2_{\text{ols}*}$, respectively. These

results indicate a marginal over-rejection of the null (of a zero fit) when relying on $R^2_{\text{ols}*}$.

If anything, these simulations provide yet another argument for focusing mainly on the restricted CSR, while largely discounting the results from the unrestricted counterpart when it delivers a positive R^2_{pen}. The reason is that the bootstrap-based inference associated with such a metric might be quite misleading in terms of evaluating "good" models.

F Internet Appendix: The Cross-Sectional Intercept and Factor Measurement Error

A natural concern is that a nonzero estimate of the CSR intercept $\widehat{\lambda}_0$ in (1) might reflect measurement error in the factor(s), rather than omitted priced components. This appendix clarifies what factor measurement error can and cannot do. The key point is that, under classical factor measurement error, the population cross-section of excess returns remains linear through the origin under correct specification (i.e., when $\alpha_i = 0$). Measurement error rescales the relevant risk price(s) but does not generate a nonzero intercept. A nonzero intercept can arise only if the measurement-error component itself carries systematic exposures across assets, in which case it is observationally equivalent to an omitted factor component and falls under the missing-factor logic of Subsection 2.1.[47]

For transparency, consider a single-factor model that is potentially misspecified. Let h_t denote the latent (true) factor and suppose expected excess returns satisfy

$$\mathrm{E}(R_{i,t}) \;=\; \lambda_h\,\beta_{i,h} \;+\; \alpha_i, \qquad \beta_{i,h} \;=\; \frac{\mathrm{Cov}(R_{i,t}, h_t)}{\mathrm{Var}(h_t)}, \qquad i = 1, \ldots, n, \tag{F.6}$$

where λ_h is the (true) factor risk price and α_i is a structural pricing error (model misspecification), distinct from the CSR residual $\widehat{\omega}_i$ defined in (1). The econometrician observes a noisy proxy:

$$f_t \;=\; h_t \;+\; \eta_t, \qquad \mathrm{E}(\eta_t) = 0, \qquad \mathrm{Cov}(h_t, \eta_t) = 0. \tag{F.7}$$

Let $\beta_{i,f} = \mathrm{Cov}(R_{i,t}, f_t)/\mathrm{Var}(f_t)$ denote the population OLS beta of asset i with respect to the proxy factor f_t.

[47] To isolate the role of factor mismeasurement, we work in population throughout this appendix and abstract from finite-T first-pass estimation noise. The latter affects inference (e.g., Shanken (1992)), but does not change the population restriction that the CSR intercept is zero for excess returns under correct model specification.

Under classical measurement error, the noise component is orthogonal to asset returns, $\mathrm{Cov}(R_{i,t}, \eta_t) = 0$ for all i. Then

$$\beta_{i,f} \;=\; \frac{\mathrm{Cov}(R_{i,t}, h_t)}{\mathrm{Var}(h_t) + \mathrm{Var}(\eta_t)} \;=\; \kappa\, \beta_{i,h}, \qquad \kappa \;=\; \frac{\mathrm{Var}(h_t)}{\mathrm{Var}(h_t) + \mathrm{Var}(\eta_t)} \in (0,1). \tag{F.8}$$

Substituting (F.8) into (F.6) yields

$$\mathrm{E}(R_{i,t}) \;=\; \lambda_f\, \beta_{i,f} \;+\; \alpha_i, \qquad \lambda_f = \frac{\lambda_h}{\kappa} \;=\; \lambda_h \left[1 + \frac{\mathrm{Var}(\eta_t)}{\mathrm{Var}(h_t)}\right], \tag{F.9}$$

so classical measurement error changes the slope (the effective risk price) but does not add a constant term. To connect this to the second-pass CSR, recall that $\mathbb{E}_n(\cdot) = n^{-1}\sum_{i=1}^{n}(\cdot)$ denotes cross-sectional averages, and consider the population CSR (best linear projection)

$$\mathrm{E}(R_{i,t}) = \lambda_0^{\mathrm{pop}} + b^{\mathrm{pop}}\beta_{i,f} + \omega_i^{\mathrm{pop}}, \qquad \mathbb{E}_n(\omega_i^{\mathrm{pop}}) = 0, \qquad \mathbb{E}_n(\beta_{i,f}\,\omega_i^{\mathrm{pop}}) = 0, \tag{F.10}$$

where ω_i^{pop} denotes the population CSR residual and $\mathrm{Var}_n(\beta_{i,f}) > 0$ is assumed so that $(\lambda_0^{\mathrm{pop}}, b^{\mathrm{pop}})$ are well defined. If the factor model is correctly specified ($\alpha_i = 0$), then (F.9) implies $\mathrm{E}(R_{i,t}) = \lambda_f\beta_{i,f}$ for all i, so $\lambda_0^{\mathrm{pop}} = 0$ and $b^{\mathrm{pop}} = \lambda_f$, and therefore $\omega_i^{\mathrm{pop}} = 0$. More generally, when an intercept is included, the orthogonality conditions in (F.10) imply that any nonzero intercept reflects the unspanned mean component of the structural pricing errors α_i (after partialling out $\beta_{i,f}$), not factor noise.

When can measurement error itself generate an intercept? Only when the "error" behaves like an omitted factor. If $\mathrm{Cov}(R_{i,t}, \eta_t) \neq 0$ for some assets (while maintaining $\mathrm{Cov}(h_t, \eta_t) = 0$ from (F.7)), the proxy beta $\beta_{i,f}$ mixes exposure to h_t and exposure to η_t. Define $\beta_{i,\eta} = \mathrm{Cov}(R_{i,t}, \eta_t)/\,\mathrm{Var}(\eta_t)$. A simple covariance decomposition gives:

$$\beta_{i,h} \;=\; \pi\, \beta_{i,f} \;-\; \phi_i, \qquad \pi = \frac{\mathrm{Var}(h_t) + \mathrm{Var}(\eta_t)}{\mathrm{Var}(h_t)} \;>\; 1, \qquad \phi_i = \beta_{i,\eta}\,\frac{\mathrm{Var}(\eta_t)}{\mathrm{Var}(h_t)}. \tag{F.11}$$

Combining (F.6) and (F.11) implies

$$\mathrm{E}(R_{i,t}) \;=\; (\pi\lambda_h)\,\beta_{i,f} \;-\; \lambda_h\,\phi_i \;+\; \alpha_i. \tag{F.12}$$

Note that $(\pi\lambda_h) = \lambda_f$ since $\pi = 1/\kappa$. The term ϕ_i is proportional to the asset's loading on the measurement-error component η_t. In an unrestricted CSR of $\mathrm{E}(R_{i,t})$ on $(1, \beta_{i,f})$, the intercept is therefore driven by the component of $-\lambda_h\phi_i + \alpha_i$ that loads on the constant after partialling out $\beta_{i,f}$. Let $\phi_i = a_0 + a_1\beta_{i,f} + \xi_i$ and $\alpha_i = c_0 + c_1\beta_{i,f} + \zeta_i$ denote the cross-sectional projections of ϕ_i and α_i on $(1, \beta_{i,f})$, with

$$\mathbb{E}_n(\xi_i) = \mathbb{E}_n(\zeta_i) = 0, \qquad \mathbb{E}_n(\beta_{i,f}\xi_i) = \mathbb{E}_n(\beta_{i,f}\zeta_i) = 0.$$

Then the unrestricted population CSR implied by (F.12) has intercept

$$\lambda_0^{\mathrm{pop}} = -\lambda_h a_0 + c_0, \qquad c_0 = \mathbb{E}_n(\alpha_i) - \frac{\mathrm{Cov}_n(\alpha_i, \beta_{i,f})}{\mathrm{Var}_n(\beta_{i,f})}\,\mathbb{E}_n(\beta_{i,f}),$$

so a mean-zero α_i contributes to the intercept unless α_i is also cross-sectionally uncorrelated with $\beta_{i,f}$ or $\mathbb{E}_n(\beta_{i,f}) = 0$. To isolate the measurement-error channel, suppose $c_0 = 0$. Note that $c_0 = 0$ is weaker than the absence of mispricing ($\alpha_i = 0$ for all i) since it removes only the constant component of the pricing errors after projection on $(1, \beta_{i,f})$, which is the only component that enters the intercept. With $\lambda_h > 0$, the channel then yields a positive intercept if and only if $a_0 < 0$, that is, when the measurement-error loadings have a negative unspanned mean after projection on the proxy beta. In particular, if, in addition, ϕ_i is cross-sectionally uncorrelated with $\beta_{i,f}$, then $a_0 = \mathbb{E}_n(\phi_i)$ and

$$\lambda_0^{\mathrm{pop}} = -\lambda_h\,\mathbb{E}_n(\phi_i).$$

Moreover, averaging (F.12) across i gives the identity

$$\overline{\mu} = \mathbb{E}_n\big(\mathrm{E}(R_{i,t})\big) = \lambda_f\,\mathbb{E}_n(\beta_{i,f}) - \lambda_h\,\mathbb{E}_n(\phi_i) + \mathbb{E}_n(\alpha_i),$$

where $\overline{\mu}$ denotes the cross-sectional average expected excess return. Since $\lambda_0^{\mathrm{pop}} = \overline{\mu} - b^{\mathrm{pop}}\,\mathbb{E}_n(\beta_{i,f})$ in (F.10), it follows that if $\mathbb{E}_n(\alpha_i) = 0$ and $b^{\mathrm{pop}} = \lambda_f$ (e.g., if $-\lambda_h\phi_i + \alpha_i$ is cross-sectionally uncorrelated with $\beta_{i,f}$), then

$$\lambda_0^{\mathrm{pop}} = \overline{\mu} - \lambda_f\,\mathbb{E}_n(\beta_{i,f}) = -\lambda_h\,\mathbb{E}_n(\phi_i) = -\lambda_h\,\mathbb{E}_n(\beta_{i,\eta})\,\frac{\mathrm{Var}(\eta_t)}{\mathrm{Var}(h_t)}.$$

With $\lambda_h > 0$, this channel yields a positive intercept only if $\mathbb{E}_n(\beta_{i,\eta}) < 0$, with magnitude $\lambda_h \, |\mathbb{E}_n(\beta_{i,\eta})| \, \mathrm{Var}(\eta_t)/\mathrm{Var}(h_t)$ under the same conditions, so an economically large positive intercept further requires a sufficiently negative average loading on the measurement-error component, a large variance ratio, or both. Both features strain the interpretation of η_t as innocuous noise in a proxy for h_t. A large variance ratio means the noise component is quantitatively important, while a negative average $\beta_{i,\eta}$ indicates systematic cross-sectional exposure to η_t. Once η_t has such systematic exposures, it is empirically indistinguishable from an omitted factor component, and the resulting intercept is exactly the type of misspecification captured by the omitted-factor algebra in Subsection 2.1.

To sum up, classical factor measurement error alone rescales estimated risk prices but does not generate a nonzero CSR intercept. Economically large positive intercepts require nonclassical structure in which the measurement-error component carries systematic cross-sectional exposure that is not absorbed by the observed proxy beta. For $\lambda_h > 0$, this requires a negative unspanned mean loading on the measurement-error component, while the magnitude is amplified by $\mathrm{Var}(\eta_t)/\mathrm{Var}(h_t)$. Such a case is best interpreted as an omitted (or otherwise unmodeled) factor component rather than innocuous measurement noise.

G Internet Appendix: Asymptotic Inference for R^2_{pen}

As an illustrative alternative to the bootstrap-based inference used in the paper, this appendix develops large-T inference for R^2_{pen}. Our focus is the economically relevant null $H_0 : R^2_{\text{pen}} = 0$. We restrict attention to cases in which this null is an interior restriction, which holds when $\lambda_0 \neq 0$.

Let $\boldsymbol{r}_t$ be the $n \times 1$ vector of excess returns on the test assets and $\boldsymbol{f}_t$ the $K \times 1$ factor vector, with n and K fixed and $T \to \infty$. Define the primitive moment vector, $\boldsymbol{\psi} = \big[\boldsymbol{\mu}_0', \ \mathrm{vec}(\boldsymbol{\Sigma}_{rf})', \ \mathrm{vec}(\boldsymbol{\Sigma}_f)'\big]'$, with $\boldsymbol{\mu}_0 = \mathrm{E}(\boldsymbol{r}_t)$, $\boldsymbol{\Sigma}_{rf} = \mathrm{Cov}(\boldsymbol{r}_t, \boldsymbol{f}_t)$, and $\boldsymbol{\Sigma}_f = \mathrm{Var}(\boldsymbol{f}_t)$, and let $\widehat{\boldsymbol{\psi}}$ denote the corresponding sample analogue. We assume a standard time-series CLT,

$$\sqrt{T}\big(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}\big) \ \Rightarrow \ \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \tag{G.13}$$

where $\boldsymbol{\Omega}$ is the long-run covariance matrix (Assumptions A.1–A.2 below).

Let $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{rf}\boldsymbol{\Sigma}_f^{-1}$ denote the population beta matrix. Let $Q_0 = \frac{1}{n}\boldsymbol{\mu}_0'\boldsymbol{M}_1\boldsymbol{\mu}_0$ denote the CS

variance of expected returns, where $\boldsymbol{M}_1 = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'$. Let $Q_{\text{pen}} = \frac{1}{n}\boldsymbol{\alpha}'\boldsymbol{\alpha}$ denote the penalized numerator, where $\boldsymbol{\alpha} = \boldsymbol{\mu}_0 - \boldsymbol{\beta}\boldsymbol{\lambda} = \lambda_0\mathbf{1} + \boldsymbol{\omega}$ is the slope-only pricing-error vector from the unrestricted CSR. The population ratio is $R^2_{\text{pen}} = 1 - Q_{\text{pen}}/Q_0$. Unlike R^2_{ols}, R^2_{pen} can be negative, which makes $R^2_{\text{pen}} = 0$ generically an interior point. More precisely, the lower bound is $-\lambda_0^2/Q_0$, so $R^2_{\text{pen}} = 0$ is a boundary point only when $\lambda_0 = 0$. The following proposition presents the asymptotic distribution.

Proposition G.1. *Maintain Assumptions A.1–A.2 below. Then, if $R^2_{\text{pen}} = 0$ and $\lambda_0 \neq 0$, then $H_0 : R^2_{\text{pen}} = 0$ is an interior restriction and*

$$\sqrt{T}\,\widehat{R}^2_{\text{pen}} \;\Rightarrow\; \mathcal{N}(0, V_{\text{pen}}), \qquad V_{\text{pen}} = \boldsymbol{D}_{\text{pen}}\,\boldsymbol{\Omega}\,\boldsymbol{D}'_{\text{pen}},$$

where $\boldsymbol{D}_{\text{pen}} = \partial g_{\text{pen}}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}'$ is evaluated at the true $\boldsymbol{\psi}$ satisfying $R^2_{\text{pen}} = 0$. Assume additionally that $V_{\text{pen}} > 0$ so that Wald inference is nondegenerate. A feasible Wald statistic is

$$\mathcal{T}_{\text{pen}} = \frac{\widehat{R}^2_{\text{pen}}}{\sqrt{\widehat{V}_{\text{pen}}/T}}, \qquad \widehat{V}_{\text{pen}} = \widehat{\boldsymbol{D}}_{\text{pen}}\,\widehat{\boldsymbol{\Omega}}\,\widehat{\boldsymbol{D}}'_{\text{pen}}, \tag{G.14}$$

where $\widehat{\boldsymbol{\Omega}}$ is a HAC estimator of $\boldsymbol{\Omega}$ and $\widehat{\boldsymbol{D}}_{\text{pen}}$ is the plug-in estimate of $\boldsymbol{D}_{\text{pen}}$ using the closed-form null gradients reported in the next appendix. Under the maintained interior-null conditions, $\mathcal{T}_{\text{pen}} \Rightarrow \mathcal{N}(0,1)$. For example, $\widehat{\boldsymbol{\Omega}}$ may be a Newey–West HAC estimator with truncation lag $L_T \to \infty$ and $L_T/T \to 0$ as $T \to \infty$, under the weak dependence conditions in Assumption A.1.

For the one-sided alternative $H_1 : R^2_{\text{pen}} > 0$, reject when $\mathcal{T}_{\text{pen}} > z_{1-\varrho}$.

If we rely on a spectral density matrix $\widehat{\boldsymbol{\Omega}}$ using zero lags, we obtain asymptotic p-values (for $H_0 : R^2_{\text{pen}} = 0$) of 0.48 and 0.30 for AEM1 (estimation with M10) and AEM2 (BM10), respectively. Using a Newey–West estimator (Newey and West (1987)) with one lag yields identical p-values. Thus, according to the asymptotic inference, the sample R^2_{pen} of 81% for AEM2 is not significant even at the 10% level, something that is at odds with the bootstrap-based inference. These results should, however, be interpreted with considerable caution, given the relatively small $T = 191$ in the quarterly estimations. In fact, untabulated Monte Carlo simulations show that the asymptotic test for $H_0 : R^2_{\text{pen}} = 0$ against $H_1 : R^2_{\text{pen}} > 0$ has very low power (rejection rate under the alternative of positive fit of only 2.6%). Accordingly, we regard this asymptotic test as illustrative rather than as a recommended inference procedure at these sample sizes, and we rely on the bootstrap described

in Appendix E for the inference reported in the paper.

H Internet Appendix: Proof of Proposition G.1

This appendix provides the proof of Proposition G.1. Throughout, the cross-sectional dimension n and the number of factors K are fixed while $T \to \infty$.

H.1 Time-Series Sampling Framework and Primitive Moments

Let $\boldsymbol{r}_t = (R_{1,t}, \ldots, R_{n,t})'$ denote the $n \times 1$ vector of excess returns on the n test assets and let $\boldsymbol{f}_t$ denote the $K \times 1$ factor vector. Define the population moments

$$\boldsymbol{\mu}_0 = \mathrm{E}(\boldsymbol{r}_t) \in \mathbb{R}^n, \qquad \boldsymbol{\mu}_f = \mathrm{E}(\boldsymbol{f}_t) \in \mathbb{R}^K, \qquad \boldsymbol{\Sigma}_{rf} = \mathrm{Cov}(\boldsymbol{r}_t, \boldsymbol{f}_t) \in \mathbb{R}^{n \times K}, \qquad \boldsymbol{\Sigma}_f = \mathrm{Var}(\boldsymbol{f}_t) \in \mathbb{R}^{K \times K},$$

and the population beta matrix (with a first-pass intercept) as

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{rf} \boldsymbol{\Sigma}_f^{-1} \in \mathbb{R}^{n \times K}. \tag{H.15}$$

Since $\boldsymbol{\Sigma}_f$ is symmetric positive definite, $\boldsymbol{\Sigma}_f^{-1}$ exists and is itself symmetric, so $(\boldsymbol{\Sigma}_f^{-1})' = \boldsymbol{\Sigma}_f^{-1}$.

Sample analogues are

$$\boldsymbol{\mu} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{r}_t, \qquad \overline{\boldsymbol{f}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{f}_t,$$

$$\widehat{\boldsymbol{\Sigma}}_{rf} = \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{r}_t - \boldsymbol{\mu})(\boldsymbol{f}_t - \overline{\boldsymbol{f}})', \qquad \widehat{\boldsymbol{\Sigma}}_f = \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{f}_t - \overline{\boldsymbol{f}})(\boldsymbol{f}_t - \overline{\boldsymbol{f}})',$$

and

$$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\Sigma}}_{rf} \widehat{\boldsymbol{\Sigma}}_f^{-1}. \tag{H.16}$$

Define the stacked vector of primitive moments and its sample analogue by

$$\boldsymbol{\psi} = \begin{bmatrix} \boldsymbol{\mu}_0 \\ \mathrm{vec}(\boldsymbol{\Sigma}_{rf}) \\ \mathrm{vec}(\boldsymbol{\Sigma}_f) \end{bmatrix} \in \mathbb{R}^{n+nK+K^2}, \qquad \widehat{\boldsymbol{\psi}} = \begin{bmatrix} \boldsymbol{\mu} \\ \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_{rf}) \\ \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_f) \end{bmatrix}. \tag{H.17}$$

Assumption A.1 (Weak dependence and CLT for primitive moments). $\{(\boldsymbol{r}_t, \boldsymbol{f}_t)\}_{t \geq 1}$ is strictly stationary and ergodic, with $\mathrm{E} \|\boldsymbol{r}_t\|^{4+\delta} < \infty$ and $\mathrm{E} \|\boldsymbol{f}_t\|^{4+\delta} < \infty$ for some $\delta > 0$, and

satisfies a strong mixing condition with mixing coefficients $\alpha(\ell) = O(\ell^{-a})$ for some $a > 2(4+\delta)/\delta$. Define the mean-zero influence vector

$$\boldsymbol{z}_t = \begin{bmatrix} \boldsymbol{r}_t - \boldsymbol{\mu}_0 \\ \mathrm{vec}\Big((\boldsymbol{r}_t - \boldsymbol{\mu}_0)(\boldsymbol{f}_t - \boldsymbol{\mu}_f)' - \boldsymbol{\Sigma}_{rf}\Big) \\ \mathrm{vec}\Big((\boldsymbol{f}_t - \boldsymbol{\mu}_f)(\boldsymbol{f}_t - \boldsymbol{\mu}_f)' - \boldsymbol{\Sigma}_f\Big) \end{bmatrix}. \tag{H.18}$$

Then the linear representation and CLT hold:

$$\sqrt{T}\left(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}\right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{z}_t + o_p(1) \;\Rightarrow\; \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}), \qquad \boldsymbol{\Omega} = \sum_{\ell=-\infty}^{\infty} \mathrm{E}\left(\boldsymbol{z}_t \boldsymbol{z}_{t-\ell}'\right). \tag{H.19}$$

Because $\mathrm{vec}(\boldsymbol{\Sigma}_f)$ stacks symmetric entries twice, $\boldsymbol{\Omega}$ may be singular; this does not affect the scalar variances $V_{\mathrm{pen}} = \boldsymbol{D}_{\mathrm{pen}} \boldsymbol{\Omega} \boldsymbol{D}_{\mathrm{pen}}'$ reported below.

Assumption A.2 (Identification and nondegeneracy). $\boldsymbol{\Sigma}_f$ is nonsingular and the unrestricted second-pass design matrix $\boldsymbol{W} = [\boldsymbol{1} : \boldsymbol{\beta}] \in \mathbb{R}^{n \times (K+1)}$ has full column rank. In particular, $n \geq K+1$ and $\boldsymbol{\beta}$ has full column rank K. The cross-sectional variance of expected returns is strictly positive:

$$Q_0 = \frac{1}{n} \boldsymbol{\mu}_0' \boldsymbol{M}_1 \boldsymbol{\mu}_0 > 0, \qquad \boldsymbol{M}_1 = \boldsymbol{I}_n - \frac{1}{n} \boldsymbol{1}\boldsymbol{1}'. \tag{H.20}$$

Finally, the Frisch–Waugh–Lovell matrix $\boldsymbol{A} = \boldsymbol{\beta}' \boldsymbol{M}_1 \boldsymbol{\beta}$ is nonsingular.

H.2 Second-Pass CSRs and the Two R^2 Measures

Let $\widehat{\boldsymbol{W}} = [\boldsymbol{1} : \widehat{\boldsymbol{\beta}}]$. The unrestricted (with-intercept) CSR is

$$\boldsymbol{\mu} = \widehat{\lambda}_0 \boldsymbol{1} + \widehat{\boldsymbol{\beta}} \widehat{\boldsymbol{\lambda}} + \widehat{\boldsymbol{\omega}}, \qquad \begin{bmatrix} \widehat{\lambda}_0 \\ \widehat{\boldsymbol{\lambda}} \end{bmatrix} = (\widehat{\boldsymbol{W}}' \widehat{\boldsymbol{W}})^{-1} \widehat{\boldsymbol{W}}' \boldsymbol{\mu}. \tag{H.21}$$

The paper's penalized pricing-error vector is the slope-only pricing error from the unrestricted CSR:

$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{\mu} - \widehat{\boldsymbol{\beta}} \widehat{\boldsymbol{\lambda}} = \widehat{\lambda}_0 \boldsymbol{1} + \widehat{\boldsymbol{\omega}}. \tag{H.22}$$

Define the common denominator

$$\widehat{Q}_0 = \frac{1}{n} \boldsymbol{\mu}' \boldsymbol{M}_1 \boldsymbol{\mu}, \tag{H.23}$$

and the two numerator quadratic forms

$$\widehat{Q} = \frac{1}{n}\widehat{\boldsymbol{\omega}}'\widehat{\boldsymbol{\omega}}, \qquad \widehat{Q}_{\text{pen}} = \frac{1}{n}\widehat{\boldsymbol{\alpha}}'\widehat{\boldsymbol{\alpha}}.$$

The corresponding explanatory ratios are

$$\widehat{R}^2_{\text{ols}} = 1 - \frac{\widehat{Q}}{\widehat{Q}_0}, \qquad \widehat{R}^2_{\text{pen}} = 1 - \frac{\widehat{Q}_{\text{pen}}}{\widehat{Q}_0}. \tag{H.24}$$

Wedge identity for the penalized numerator. Because the unrestricted CSR includes an intercept, $\mathbf{1}'\widehat{\boldsymbol{\omega}} = 0$ holds exactly. Therefore

$$\widehat{Q}_{\text{pen}} = \frac{1}{n}(\widehat{\lambda}_0\mathbf{1} + \widehat{\boldsymbol{\omega}})'(\widehat{\lambda}_0\mathbf{1} + \widehat{\boldsymbol{\omega}}) = \widehat{Q} + \widehat{\lambda}_0^2, \qquad \widehat{R}^2_{\text{pen}} = \widehat{R}^2_{\text{ols}} - \frac{\widehat{\lambda}_0^2}{\widehat{Q}_0}. \tag{H.25}$$

Population counterparts satisfy the analogous identities:

$$Q_{\text{pen}} = Q + \lambda_0^2, \qquad R^2_{\text{pen}} = R^2_{\text{ols}} - \frac{\lambda_0^2}{Q_0}.$$

Population CSR objects (used below). Let $\boldsymbol{W} = [\mathbf{1} : \boldsymbol{\beta}]$ and write the population unrestricted CSR decomposition as

$$\boldsymbol{\mu}_0 = \lambda_0\mathbf{1} + \boldsymbol{\beta}\boldsymbol{\lambda} + \boldsymbol{\omega}, \qquad \mathbf{1}'\boldsymbol{\omega} = 0, \qquad \boldsymbol{\beta}'\boldsymbol{\omega} = \mathbf{0},$$

where the Frisch–Waugh–Lovell formulas are

$$\boldsymbol{\lambda} = (\boldsymbol{\beta}'\boldsymbol{M}_1\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{M}_1\boldsymbol{\mu}_0, \qquad \lambda_0 = \frac{1}{n}\mathbf{1}'\boldsymbol{\mu}_0 - \boldsymbol{m}'\boldsymbol{\lambda}, \qquad \boldsymbol{m} = \frac{1}{n}\boldsymbol{\beta}'\mathbf{1}.$$

Finally define the population slope-only pricing error $\boldsymbol{\alpha} = \boldsymbol{\mu}_0 - \boldsymbol{\beta}\boldsymbol{\lambda} = \lambda_0\mathbf{1} + \boldsymbol{\omega}$ and the population quadratic forms

$$Q = \frac{1}{n}\boldsymbol{\omega}'\boldsymbol{\omega}, \qquad Q_{\text{pen}} = \frac{1}{n}\boldsymbol{\alpha}'\boldsymbol{\alpha}, \qquad R^2_{\text{pen}} = 1 - \frac{Q_{\text{pen}}}{Q_0}.$$

H.3 Wald Inference at $R^2_{\text{pen}} = 0$: Null Gradients, Delta-Method Variance, and HAC Estimation

Consider the null $H_0 : R^2_{\text{pen}} = 0$. Under Assumptions A.1–A.2, the CLT in (H.19) and the delta method yield an asymptotic normal approximation for $\widehat{R}^2_{\text{pen}} = g_{\text{pen}}(\widehat{\boldsymbol{\psi}})$ provided the corresponding null is an interior restriction. The key implementation objects are the gradient $\boldsymbol{D}_{\text{pen}} = \partial g_{\text{pen}}(\boldsymbol{\psi})/\partial \boldsymbol{\psi}'$ and the long-run covariance matrix $\boldsymbol{\Omega}$.

H.3.1 Gradients with respect to $(\boldsymbol{\mu}, \boldsymbol{\beta})$ and the Chain Rule to $\boldsymbol{\psi}$

For any scalar functional $g(\boldsymbol{\mu}, \boldsymbol{\beta})$, define gradients $\boldsymbol{G}_\mu(g) \in \mathbb{R}^n$ and $\boldsymbol{G}_\beta(g) \in \mathbb{R}^{n\times K}$ by

$$dg = \boldsymbol{G}_\mu(g)'\, d\boldsymbol{\mu} + \operatorname{tr}\big[\boldsymbol{G}_\beta(g)'\, d\boldsymbol{\beta}\big]. \tag{H.26}$$

From $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{rf}\boldsymbol{\Sigma}_f^{-1}$,

$$d\boldsymbol{\beta} = d\boldsymbol{\Sigma}_{rf}\, \boldsymbol{\Sigma}_f^{-1} - \boldsymbol{\beta}\, (d\boldsymbol{\Sigma}_f)\, \boldsymbol{\Sigma}_f^{-1}. \tag{H.27}$$

Substituting (H.27) into (H.26) and using the cyclic property of the trace yields

$$dg = \boldsymbol{G}_\mu(g)'\, d\boldsymbol{\mu} + \operatorname{tr}\big[\boldsymbol{G}_{\Sigma_{rf}}(g)'\, d\boldsymbol{\Sigma}_{rf}\big] + \operatorname{tr}\big[\boldsymbol{G}_{\Sigma_f}(g)'\, d\boldsymbol{\Sigma}_f\big], \tag{H.28}$$

where

$$\boldsymbol{G}_{\Sigma_{rf}}(g) = \boldsymbol{G}_\beta(g)\boldsymbol{\Sigma}_f^{-1} \in \mathbb{R}^{n\times K}, \qquad \boldsymbol{G}_{\Sigma_f}(g) = -\boldsymbol{\beta}'\boldsymbol{G}_\beta(g)\boldsymbol{\Sigma}_f^{-1} \in \mathbb{R}^{K\times K}. \tag{H.29}$$

Vectorizing (H.28) yields the gradient of g with respect to the primitive moments $\boldsymbol{\psi}$:

$$\boldsymbol{D}(g) = \frac{\partial g}{\partial \boldsymbol{\psi}'} = \Big[\boldsymbol{G}_\mu(g)' \quad \operatorname{vec}\big(\boldsymbol{G}_{\Sigma_{rf}}(g)\big)' \quad \operatorname{vec}\big(\boldsymbol{G}_{\Sigma_f}(g)\big)'\Big] \in \mathbb{R}^{1\times(n+nK+K^2)}. \tag{H.30}$$

We write $\boldsymbol{D}_{\text{pen}} = \boldsymbol{D}(R^2_{\text{pen}})$ and $V_{\text{pen}} = \boldsymbol{D}_{\text{pen}}\boldsymbol{\Omega}\boldsymbol{D}_{\text{pen}}'$.

H.3.2 Auxiliary Derivatives for λ_0

Define $\boldsymbol{A} = \boldsymbol{\beta}'\boldsymbol{M}_1\boldsymbol{\beta}$, $\boldsymbol{m} = \frac{1}{n}\boldsymbol{\beta}'\mathbf{1}$, and $\boldsymbol{\kappa} = \boldsymbol{A}^{-1}\boldsymbol{m}$. The gradients of the intercept λ_0 with respect to $(\boldsymbol{\mu}, \boldsymbol{\beta})$ satisfy

$$\boldsymbol{g}_\mu = \frac{1}{n}\mathbf{1} - \boldsymbol{M}_1\boldsymbol{\beta}\,\boldsymbol{\kappa}, \qquad d\lambda_0 = \boldsymbol{g}_\mu' d\boldsymbol{\mu} + \operatorname{tr}\big(\boldsymbol{G}_0' d\boldsymbol{\beta}\big), \tag{H.31}$$

where

$$\boldsymbol{G}_0 = -\frac{1}{n}\mathbf{1}\,\boldsymbol{\lambda}' - \boldsymbol{\omega}\,\boldsymbol{\kappa}' + (\boldsymbol{M}_1\boldsymbol{\beta})\boldsymbol{\kappa}\boldsymbol{\lambda}'. \tag{H.32}$$

(Here $\boldsymbol{g}_\mu$ is the gradient of the scalar λ_0; it is distinct from the notation $\boldsymbol{G}_\mu(\cdot)$ in (H.26).)

H.3.3 Closed-Form Null Gradients for R^2_{pen}

At the null $R^2_{\text{pen}} = 0$, the gradients in (H.26) simplify to a closed-form expression: if $R^2_{\text{pen}} = 0$ (equivalently $Q + \lambda_0^2 = Q_0$), then

$$\boldsymbol{G}_\mu(R^2_{\text{pen}})\Big|_{R^2_{\text{pen}}=0} = -\frac{2}{nQ_0}\boldsymbol{\omega} - \frac{2\lambda_0}{Q_0}\boldsymbol{g}_\mu + \frac{2}{nQ_0}\boldsymbol{M}_1\boldsymbol{\mu}_0, \qquad \boldsymbol{G}_\beta(R^2_{\text{pen}})\Big|_{R^2_{\text{pen}}=0} = \frac{2}{nQ_0}\boldsymbol{\omega}\,\boldsymbol{\lambda}' - \frac{2\lambda_0}{Q_0}\boldsymbol{G}_0. \tag{H.33}$$

H.3.4 Proof of Proposition G.1 and Feasible Implementation

Under Assumptions A.1–A.2, the CLT in (H.19) holds for $\widehat{\boldsymbol{\psi}}$. Write $R^2_{\text{pen}} = g_{\text{pen}}(\boldsymbol{\psi})$. At the relevant interior nulls, g_{pen} is differentiable, and the delta method gives

$$\sqrt{T}\big(\widehat{R}^2_{\text{pen}} - R^2_{\text{pen}}\big) \;\Rightarrow\; \mathcal{N}(0, V_{\text{pen}}), \qquad V_{\text{pen}} = \boldsymbol{D}_{\text{pen}}\,\boldsymbol{\Omega}\,\boldsymbol{D}'_{\text{pen}}, \qquad \boldsymbol{D}_{\text{pen}} = \frac{\partial g_{\text{pen}}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}'}.$$

When $R^2_{\text{pen}} = 0$, this is the distribution of $\sqrt{T}\widehat{R}^2_{\text{pen}}$. The null is an interior restriction when $\lambda_0 \neq 0$. The required derivatives are obtained from the chain-rule formula (H.30), using the null gradients in (H.33). If $V_{\text{pen}} > 0$, the Wald statistic in Proposition G.1 is nondegenerate. This proves Proposition G.1.

Feasible HAC estimation. Let $\widehat{\boldsymbol{z}}_t$ be the sample analogue of (H.18) obtained by replacing population objects $(\boldsymbol{\mu}_0, \boldsymbol{\mu}_f, \boldsymbol{\Sigma}_{rf}, \boldsymbol{\Sigma}_f)$ with their sample analogues $(\boldsymbol{\mu}, \overline{\boldsymbol{f}}, \widehat{\boldsymbol{\Sigma}}_{rf}, \widehat{\boldsymbol{\Sigma}}_f)$. Given a bandwidth L, the Newey–West estimator is

$$\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Gamma}}_0 + \sum_{\ell=1}^{L} w_\ell\big(\widehat{\boldsymbol{\Gamma}}_\ell + \widehat{\boldsymbol{\Gamma}}'_\ell\big), \qquad \widehat{\boldsymbol{\Gamma}}_\ell = \frac{1}{T}\sum_{t=\ell+1}^{T} \widehat{\boldsymbol{z}}_t\,\widehat{\boldsymbol{z}}'_{t-\ell}, \qquad w_\ell = 1 - \frac{\ell}{L+1}.$$

Form $\widehat{\boldsymbol{D}}_{\text{pen}}$ by plugging sample analogues into (H.30), using the null gradients (H.33). Then $\widehat{V}_{\text{pen}} = \widehat{\boldsymbol{D}}_{\text{pen}}\,\widehat{\boldsymbol{\Omega}}\,\widehat{\boldsymbol{D}}'_{\text{pen}}$.

Wald test at $R^2_{\text{pen}} = 0$. The feasible Wald statistic is

$$\mathcal{T}_{\text{pen}} = \frac{\widehat{R}^2_{\text{pen}}}{\sqrt{\widehat{V}_{\text{pen}}/T}} \;\Rightarrow\; \mathcal{N}(0,1) \quad \text{under } H_0 : R^2_{\text{pen}} = 0.$$

The two-sided test rejects at asymptotic size ϱ when $|\mathcal{T}_{\text{pen}}| > z_{1-\varrho/2}$; the one-sided alternative $H_1 : R^2_{\text{pen}} > 0$ rejects when $\mathcal{T}_{\text{pen}} > z_{1-\varrho}$.

Feasible computation

1. Compute $\boldsymbol{\mu}$, $\overline{\boldsymbol{f}}$, $\widehat{\boldsymbol{\Sigma}}_{rf}$, $\widehat{\boldsymbol{\Sigma}}_f$, and $\widehat{\boldsymbol{\beta}}$ in (H.16), and set $\widehat{Q}_0 = (1/n)\boldsymbol{\mu}'\boldsymbol{M}_1\boldsymbol{\mu}$.
2. Compute $(\widehat{\lambda}_0, \widehat{\boldsymbol{\lambda}})$ and $\widehat{\boldsymbol{\omega}}$ from (H.21).
3. Compute $\widehat{R}^2_{\text{pen}}$ from (H.24).
4. Compute the auxiliary quantities for the null gradients: $\widehat{\boldsymbol{m}} = (1/n)\widehat{\boldsymbol{\beta}}'\mathbf{1}$, $\widehat{\boldsymbol{A}} = \widehat{\boldsymbol{\beta}}'\boldsymbol{M}_1\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\kappa}} = \widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{m}}$, $\widehat{\boldsymbol{g}}_\mu = \frac{1}{n}\mathbf{1} - \boldsymbol{M}_1\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\kappa}}$, and $\widehat{\boldsymbol{G}}_0 = -\frac{1}{n}\mathbf{1}\,\widehat{\boldsymbol{\lambda}}' - \widehat{\boldsymbol{\omega}}\widehat{\boldsymbol{\kappa}}' + (\boldsymbol{M}_1\widehat{\boldsymbol{\beta}})\widehat{\boldsymbol{\kappa}}\widehat{\boldsymbol{\lambda}}'$.
5. Form $\widehat{\boldsymbol{D}}_{\text{pen}}$ using (H.30) with gradients (H.33) (replace population objects by sample analogues).
6. Construct $\widehat{\boldsymbol{z}}_t$, estimate $\widehat{\boldsymbol{\Omega}}$ by Newey–West, compute $\widehat{V}_{\text{pen}} = \widehat{\boldsymbol{D}}_{\text{pen}}\,\widehat{\boldsymbol{\Omega}}\,\widehat{\boldsymbol{D}}'_{\text{pen}}$, and report $\mathcal{T}_{\text{pen}} = \widehat{R}^2_{\text{pen}}/\sqrt{\widehat{V}_{\text{pen}}/T}$.

I GLS CSR: Additional Evidence

This appendix collects the identification analysis for the GLS CSR and reports the GLS risk price estimates. The identification subsection states the GLS rank and determinant conditions. The second subsection contains the GLS empirical results.

I.1 GLS Identification

Proposition I.1 (GLS analogue of Corollary 7)**.** *Let $q = \mathbf{1}'\boldsymbol{C}\mathbf{1}$. When $\boldsymbol{M}_{\text{gls}}$ is nonsingular,*

$$\det(\boldsymbol{W}'\boldsymbol{C}\boldsymbol{W}) = q^{K+1}\det(\boldsymbol{M}_{\text{gls}})(1 - s_{\text{gls}}) = q^{K+1}\det(\boldsymbol{\Sigma}_{\text{gls}}).$$

Hence the unrestricted GLS CSR is identified if and only if $\boldsymbol{\Sigma}_{\text{gls}}$ is positive definite, while the restricted GLS CSR requires only $\boldsymbol{M}_{\text{gls}}$ positive definite.

Proof. Since $\boldsymbol{C} \succ 0$, every matrix of the form $\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X} = (\boldsymbol{C}^{1/2}\boldsymbol{X})'(\boldsymbol{C}^{1/2}\boldsymbol{X})$ is positive semidefinite, hence nonsingular if and only if positive definite. With $q = \mathbf{1}'\boldsymbol{C}\mathbf{1} > 0$, the definitions of the GLS moments give

$$\frac{1}{q}\boldsymbol{W}'\boldsymbol{C}\boldsymbol{W} = \begin{bmatrix} 1 & \boldsymbol{m}_{\text{gls}}' \\ \boldsymbol{m}_{\text{gls}} & \boldsymbol{M}_{\text{gls}} \end{bmatrix}.$$

The Schur complement of the $(1,1)$ entry is $\boldsymbol{M}_{\text{gls}} - \boldsymbol{m}_{\text{gls}}\boldsymbol{m}_{\text{gls}}' = \boldsymbol{\Sigma}_{\text{gls}}$, so

$$\det(\boldsymbol{W}'\boldsymbol{C}\boldsymbol{W}) = q^{K+1} \det \boldsymbol{\Sigma}_{\text{gls}}.$$

Computing the same determinant by completing the $\boldsymbol{M}_{\text{gls}}$ block instead gives, when $\boldsymbol{M}_{\text{gls}}$ is nonsingular,

$$\det(\boldsymbol{W}'\boldsymbol{C}\boldsymbol{W}) = q^{K+1} \det(\boldsymbol{M}_{\text{gls}})(1 - \boldsymbol{m}_{\text{gls}}'\boldsymbol{M}_{\text{gls}}^{-1}\boldsymbol{m}_{\text{gls}}) = q^{K+1} \det(\boldsymbol{M}_{\text{gls}})(1 - s_{\text{gls}}).$$

The same Schur factorization, whose pivot $(1,1)$ entry is positive, yields $\boldsymbol{W}'\boldsymbol{C}\boldsymbol{W} \succ 0 \iff \boldsymbol{\Sigma}_{\text{gls}} \succ 0$. Hence the unrestricted GLS CSR is identified if and only if $\boldsymbol{\Sigma}_{\text{gls}}$ is positive definite. The restricted GLS CSR has design $\widehat{\boldsymbol{\beta}}$ and matrix $\widehat{\boldsymbol{\beta}}'\boldsymbol{C}\widehat{\boldsymbol{\beta}} = q\boldsymbol{M}_{\text{gls}}$, which is nonsingular if and only if $\boldsymbol{M}_{\text{gls}}$ is positive definite. □

The same conclusion can be stated as a rank condition. Since $\boldsymbol{C} \succ 0$, premultiplication by $\boldsymbol{C}^{1/2}$ preserves exact linear dependence among the columns of $\boldsymbol{W}$, so $\operatorname{rank}(\boldsymbol{W}'\boldsymbol{C}\boldsymbol{W}) = \operatorname{rank}(\boldsymbol{W})$. Hence the identification conditions in Proposition I.1 are equivalent to $\boldsymbol{W} = [\mathbf{1} : \widehat{\boldsymbol{\beta}}]$ having full column rank in the unrestricted GLS CSR and $\widehat{\boldsymbol{\beta}}$ having full column rank in the restricted GLS CSR.

The unrestricted GLS CSR can become ill conditioned for two distinct reasons. First, $\boldsymbol{M}_{\text{gls}}$ can be nearly singular, reflecting weak or redundant beta directions. Second, even when $\boldsymbol{M}_{\text{gls}}$ is well conditioned, s_{gls} can be close to one, reflecting near collinearity between the constant vector and the beta span in the $\boldsymbol{C}$-weighted geometry. The latter is the GLS analogue of the intercept-amplification channel studied in the main text. These two problems can occur together, but neither implies the other.

The rank tests in Kleibergen and Zhan (2020) are formulated using the first-pass beta system. Their relevance is not limited to an OLS second pass, since the same full-rank condition is required

for the GLS risk price estimator. What differs is the diagnostic object. The KZ rank test assesses whether the beta matrix, or the augmented matrix $(\mathbf{1} : \widehat{\boldsymbol{\beta}})$ when an intercept is included, has sufficient rank to identify risk premia. The scalar s_{gls} instead measures how close the constant vector is to the beta span under the GLS weighting matrix. Therefore, a failure of beta rank does not imply that s_{gls} is close to one; in that case s_{gls} is undefined without first reducing the beta matrix to a full-rank set of columns. Conversely, s_{gls} can be close to one even when the beta matrix is full rank.

I.2 Empirical Results

Tables A.8 and A.9 below report the GLS quarterly and monthly risk price estimates, respectively, for the cross-section of 40 portfolios (All). Tables A.10 and A.11 present the corresponding pricing performance metrics at the quarterly and monthly frequency, respectively; to keep the focus, we report fit results only for All and All*. In the estimation with All*, the market factor is excluded from the menu of test assets, given that most models contain RM as a factor.[48] To test for $R^2_{\text{gls}} = 0$, $R^2_{\text{pen,gls}} = 0$, or $R^2_{\text{gls}*} = 0$, we rely on the bootstrap-based inference described in Appendix E. Using Monte Carlo experiments similar to those described in Appendix E, we obtain rejection rates (under the alternative that FF6 is true) around 100% for both R^2_{gls} and $R^2_{\text{gls}*}$, but only 20% for $R^2_{\text{pen,gls}}$. Under the null that the model is false, we obtain rejection rates of 4.6% for both R^2_{gls} and $R^2_{\text{pen,gls}}$ and 5.0% for $R^2_{\text{gls}*}$. These results are in line with those obtained for the OLS counterparts and suggest that considerable caution must be used when conducting inference on $R^2_{\text{pen,gls}}$.

[48] If the market factor is included in the cross-section, the first-pass regression for the market test asset is fitted exactly in any model containing RM as a factor, so the corresponding residual variance is zero, $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ is singular, and the GLS weighting matrix $\boldsymbol{C} = \widehat{\boldsymbol{\Sigma}}_\varepsilon^{-1}$ is undefined.

Table A.8: Quarterly Risk Price Estimates: GLS

This table presents the quarterly GLS risk price estimates. The method is the standard two-step regression approach. The testing portfolios are deciles sorted on investment (INV10), operating profitability (OP10), momentum (M10), and book-to-market (BM10). λ_0 denotes the intercept estimate (in %). λ denote the risk price estimates (in %), while the numbers in parentheses represent the respective Shanken t-ratios (with * denoting statistical significance at the 5% level). The risk price estimates and t-ratios displayed after the comma are associated with the corresponding CSR with no intercept. For a description of the different models and factors, see Section 3. The sample is 1970:I to 2017:III.

	λ_0	λ_{RM}	λ_Y	λ_{def}	λ_{RMsw}	λ_{Ysw}	λ_{RMcay}
CAPM	3.70	$-1.95, 1.75$					
	(6.23*)	$(-2.25^*, 2.78^*)$					
JW2	3.60	$-1.84, 1.76$	$0.15, 0.31$				
	(5.91*)	$(-2.10^*, 2.79^*)$	$(1.04, 2.13^*)$				
JW3	3.38	$-1.62, 1.76$	$0.15, 0.29$	$-0.10, -0.21$			
	(5.23*)	$(-1.79, 2.79^*)$	$(1.04, 1.88)$	$(-1.38, -2.77^*)$			
SV2	3.69	$-1.94, 1.75$			$0.01, 0.04$		
	(6.19*)	$(-2.23^*, 2.78^*)$			$(0.40, 1.26)$		
SV3	3.63	$-1.88, 1.76$	$0.18, 0.33$			$0.01, 0.00$	
	(5.83*)	$(-2.12^*, 2.78^*)$	$(1.23, 2.23^*)$			$(1.04, 0.87)$	
SV4	3.62	$-1.87, 1.75$	$0.18, 0.32$		$0.02, 0.04$	$0.01, 0.00$	
	(5.83*)	$(-2.11^*, 2.78^*)$	$(1.17, 2.15^*)$		$(0.52, 1.23)$	$(1.04, 0.87)$	
LL2	3.70	$-1.95, 1.75$					$-0.01, 0.00$
	(6.23*)	$(-2.25^*, 2.78^*)$					$(-0.29, 0.05)$
	λ_0	λ_{RM}	λ_{LEV}	λ_{IC}	λ_C	λ_{my}	λ_{Cmy}
AEM1	2.31		$5.18, 8.76$				
	(4.63*)		$(2.63^*, 4.41^*)$				
AEM2	3.14	$-1.38, 1.76$	$4.55, 8.32$				
	(4.69*)	$(-1.50, 2.79^*)$	$(2.27^*, 4.18^*)$				
HKM1	3.44			$-0.03, 0.02$			
	(6.31*)			$(-1.95, 2.38^*)$			
HKM2	3.71	$-1.96, 1.75$		$-0.02, 0.02$			
	(6.23*)	$(-2.25^*, 2.78^*)$		$(-1.84, 1.70)$			
CCAPM	2.81				$-0.00, 0.20$		
	(6.29*)				$(-0.03, 1.89)$		
LVN3	2.75				$-0.02, 0.18$	$-3.72, -4.11$	$0.00, 0.02$
	(5.79*)				$(-0.20, 1.58)$	$(-1.52, -1.63)$	$(0.15, 1.04)$

Table A.9: Monthly Risk Price Estimates: GLS

This table presents the monthly GLS risk price estimates. The method is the standard two-step regression approach. The testing portfolios are deciles sorted on investment (INV10), operating profitability (OP10), momentum (M10), and book-to-market (BM10). λ_0 denotes the intercept estimate (in %). λ denote the risk price estimates (in %), while the numbers in parentheses represent the respective Shanken t-ratios (with * denoting statistical significance at the 5% level). The risk price estimates and t-ratios displayed after the comma are associated with the corresponding CSR with no intercept. For a description of the different models and factors, see Section 3. The sample is 1970:01 to 2017:09.

	λ_0	λ_{RM}	$\lambda_{\Delta TS}$	$\lambda_{\Delta DS}$	$\lambda_{\Delta DY}$	$\lambda_{\Delta TB}$	λ_{SMB}	λ_{HML}
CAPM	1.06	$-0.49, 0.55$						
	(4.91^*)	$(-1.73, 2.93^*)$						
HL3	1.17	$-0.60, 0.56$	$0.14, 0.09$	$-0.01, 0.02$				
	(4.84^*)	$(-1.98^*, 2.94^*)$	$(2.07^*, 1.39)$	$(-0.66, 0.90)$				
HL5	1.48	$-0.91, 0.55$	$0.08, 0.02$	$-0.00, 0.01$			$0.39, -0.06$	$0.30, 0.35$
	(4.31^*)	$(-2.34^*, 2.92^*)$	$(1.13, 0.24)$	$(-0.17, 0.76)$			$(2.03^*, -0.36)$	$(2.32^*, 2.76^*)$
P5	1.31	$-0.74, 0.56$	$0.13, 0.13$	$-0.01, 0.00$	$0.76, -0.68$	$-0.01, -0.01$		
	(4.09^*)	$(-2.00^*, 2.94^*)$	$(1.93, 1.83)$	$(-0.61, 0.13)$	$(1.89, -3.51^*)$	$(-0.89, -2.10^*)$		
P7	1.48	$-0.91, 0.55$	$0.08, 0.03$	$-0.00, 0.01$	$0.82, -0.71$	$-0.00, -0.01$	$0.49, 0.05$	$0.29, 0.35$
	(4.10^*)	$(-2.26^*, 2.91^*)$	$(0.96, 0.41)$	$(-0.06, 0.71)$	$(1.94, -3.54^*)$	$(-0.44, -1.09)$	$(2.38^*, 0.29)$	$(2.24^*, 2.73^*)$
	λ_0	λ_{RM}	λ_{SMB}	λ_{HML}	λ_{MP}	λ_{RMW}	λ_{CMA}	λ_{UMD}
LZ1	0.79				$0.03, -0.05$			
	(5.46^*)				$(0.27, -0.46)$			
LZ4	1.50	$-0.93, 0.55$	$0.39, -0.07$	$0.29, 0.36$	$0.09, -0.08$			
	(4.30^*)	$(-2.37^*, 2.92^*)$	$(2.02^*, -0.42)$	$(2.29^*, 2.79^*)$	$(0.78, -0.69)$			
FF6	0.80	$-0.24, 0.55$	$0.36, 0.13$	$0.30, 0.34$		$0.32, 0.33$	$0.28, 0.28$	$0.54, 0.54$
	(2.10^*)	$(-0.57, 2.90^*)$	$(1.81, 0.79)$	$(2.35^*, 2.62^*)$		$(3.04^*, 3.12^*)$	$(3.20^*, 3.12^*)$	$(2.90^*, 2.89^*)$
	λ_0	λ_{RM}	λ_{MP}	λ_{TS}	λ_{DS}	λ_{UI}	λ_{EI}	
CRR5	0.68		$0.02, -0.01$	$-0.08, 0.08$	$-0.10, -0.15$	$-0.05, -0.11$	$0.02, 0.00$	
	(4.01^*)		$(0.13, -0.10)$	$(-0.31, 0.28)$	$(-1.37, -1.89)$	$(-1.34, -3.05^*)$	$(0.66, 0.05)$	
CRR6	0.88	$-0.32, 0.55$	$-0.01, -0.10$	$-0.09, -0.15$	$0.03, -0.02$	$-0.04, -0.09$	$0.02, 0.01$	
	(3.31^*)	$(-0.99, 2.91^*)$	$(-0.06, -0.36)$	$(-1.18, -1.89)$	$(0.21, -0.13)$	$(-1.15, -2.52^*)$	$(0.68, 0.30)$	

Table A.10: GLS Pricing Performance in Quarterly Asset Pricing Tests

This table presents the evaluation results for the asset pricing models at the quarterly frequency. The method is the standard two-step regression approach with a GLS CSR. The testing portfolios are deciles sorted on investment (INV10), operating profitability (OP10), momentum (M10), and book-to-market (BM10). For a description of the different models and factors, see Section 3. Panel A contains the results for the augmented estimation including all 40 portfolios (INV10+OP10+M10+BM10), while Panel B contains the results for a restricted cross-section containing the first two and last two deciles in each group (for a total of 16 portfolios). $\text{Wedge}_{\text{gls}}$ equals $\lambda_0^2/\operatorname{Var}_{n,\text{gls}}(\mu_i)$, where λ_0 denotes the intercept estimate and $\operatorname{Var}_{n,\text{gls}}(\mu_i)$ is the cross-sectional variance of portfolio risk premia. s_{gls} is the cross-sectional projection of one on the factor loadings. $R^2_{\text{pen,gls}}$ denotes the cross-sectional R^2, while R^2_{gls} denotes a similar metric, except that it is based on pricing errors that exclude the intercept. $R^2_{\text{gls}*}$ denotes the performance metric associated with the CSR without intercept. The values in parentheses represent bootstrap p-values to test the null of a zero fit ($R^2_{\text{gls}} = 0$ or $R^2_{\text{pen,gls}} = 0$) or, in the case of $R^2_{\text{gls}*}$, the null of no explanatory power (zero risk prices). The sample is 1970:I to 2017:III.

	Intercept						**No intercept**	
	$\text{Wedge}_{\text{gls}}$	s_{gls}	R^2_{gls}	p	$R^2_{\text{pen,gls}}$	p	$R^2_{\text{gls}*}$	p
Panel A: All								
CAPM	245.31	1.00	0.11	(0.05)	−245.21		−0.28	
JW2	230.30	1.00	0.13	(0.08)	−230.17		−0.23	
JW3	200.45	1.00	0.15	(0.13)	−200.29		−0.13	
SV2	245.68	1.00	0.12	(0.14)	−245.56		−0.27	
SV3	239.11	1.00	0.14	(0.14)	−238.97		−0.22	
SV4	239.03	1.00	0.15	(0.24)	−238.88		−0.21	
LL2	249.03	1.00	0.11	(0.15)	−248.92		−0.28	
AEM1	0.31	0.20	0.10	(0.07)	−0.21		−0.15	
AEM2	168.66	1.00	0.21	(0.02)	−168.45		−0.03	
HKM1	1.45	0.72	0.08	(0.09)	−1.38		−0.33	
HKM2	252.23	1.00	0.11	(0.14)	−252.12		−0.28	
CCAPM	0.46	0.12	0.00	(0.98)	−0.46		−0.41	
LVN3	0.44	0.14	0.05	(0.63)	−0.38		−0.32	
Panel B: All*								
CAPM	63.66	0.99	0.16	(0.03)	−63.50		−0.33	
JW2	57.77	0.99	0.19	(0.08)	−57.58		−0.23	
JW3	27.32	0.99	0.33	(0.07)	−26.99		0.18	(0.21)
SV2	58.84	0.99	0.20	(0.06)	−58.64		−0.23	
SV3	60.15	0.99	0.20	(0.20)	−59.96		−0.23	
SV4	54.27	0.99	0.23	(0.27)	−54.04		−0.14	
LL2	64.88	0.99	0.18	(0.08)	−64.71		−0.32	
AEM1	0.20	0.36	0.24	(0.01)	0.04	(0.17)	0.11	(0.02)
AEM2	22.90	1.00	0.34	(0.01)	−22.56		0.22	(0.02)
HKM1	2.31	0.79	0.11	(0.06)	−2.20		−0.38	
HKM2	65.60	0.99	0.16	(0.12)	−65.44		−0.33	
CCAPM	0.42	0.33	0.03	(0.46)	−0.39		−0.25	
LVN3	0.52	0.36	0.15	(0.44)	−0.37		−0.19	

Table A.11: GLS Pricing Performance in Monthly Asset Pricing Tests

This table presents the evaluation results for the asset pricing models at the monthly frequency. The method is the standard two-step regression approach with a GLS CSR. The testing portfolios are deciles sorted on investment (INV10), operating profitability (OP10), momentum (M10), and book-to-market (BM10). For a description of the different models and factors, see Section 3. Panel A contains the results for the augmented estimation including all 40 portfolios (INV10+OP10+M10+BM10), while Panel B contains the results for a restricted cross-section containing the first two and last two deciles in each group (for a total of 16 portfolios). $\text{Wedge}_{\text{gls}}$ equals $\lambda_0^2/\operatorname{Var}_{n,\text{gls}}(\mu_i)$, where λ_0 denotes the intercept estimate and $\operatorname{Var}_{n,\text{gls}}(\mu_i)$ is the cross-sectional variance of portfolio risk premia. s_{gls} is the cross-sectional projection of one on the factor loadings. $R^2_{\text{pen,gls}}$ denotes the cross-sectional R^2, while R^2_{gls} denotes a similar metric, except that it is based on pricing errors that exclude the intercept. $R^2_{\text{gls}*}$ denotes the performance metric associated with the CSR without intercept. The values in parentheses represent bootstrap p-values to test the null of a zero fit ($R^2_{\text{gls}} = 0$) or, in the case of $R^2_{\text{gls}*}$, the null of no explanatory power (zero risk prices). The sample is 1970:01 to 2017:09.

	Intercept					**No intercept**	
	$\text{Wedge}_{\text{gls}}$	s_{gls}	R^2_{gls}	p	$R^2_{\text{pen,gls}}$	$R^2_{\text{gls}*}$	p
Panel A: All							
CAPM	161.57	1.00	0.06	(0.07)	−161.52	−0.20	
HL3	198.79	1.00	0.12	(0.09)	−198.67	−0.17	
HL5	151.11	1.00	0.63	(0.00)	−150.48	0.53	(0.01)
P5	255.13	1.00	0.13	(0.26)	−255.00	−0.07	
P7	155.85	1.00	0.64	(0.00)	−155.21	0.55	(0.01)
LZ1	0.34	0.02	0.00	(0.83)	−0.33	−0.33	
LZ4	155.83	1.00	0.63	(0.00)	−155.21	0.53	(0.01)
CRR5	0.26	0.26	0.07	(0.66)	−0.19	−0.12	
CRR6	112.98	1.00	0.11	(0.53)	−112.87	−0.02	
FF6	26.05	1.00	0.87	(0.00)	−25.18	0.86	(0.00)
Panel B: All*							
CAPM	65.49	0.99	0.15	(0.02)	−65.34	−0.28	
HL3	90.93	1.00	0.23	(0.17)	−90.70	−0.22	
HL5	92.21	1.00	0.75	(0.01)	−91.46	0.57	(0.02)
P5	180.24	1.00	0.33	(0.23)	−179.91	0.02	(0.93)
P7	62.50	1.00	0.76	(0.02)	−61.73	0.69	(0.02)
LZ1	0.67	0.10	0.13	(0.21)	−0.54	−0.47	
LZ4	179.26	1.00	0.84	(0.00)	−178.43	0.55	(0.01)
CRR5	0.45	0.27	0.42	(0.13)	−0.03	0.09	(0.30)
CRR6	80.70	1.00	0.54	(0.09)	−80.15	0.24	(0.47)
FF6	27.85	1.00	0.92	(0.00)	−26.93	0.90	(0.00)

Internet Appendix References

Goyal, A., I. Welch, and A. Zafirov, 2024, A comprehensive 2022 look at the empirical performance of equity premium prediction, *Review of Financial Studies* 37, 3490–3557.

Hahn, J., and H. Lee, 2006, Yield spreads as alternative risk factors for size and book-to-market, *Journal of Financial and Quantitative Analysis* 41, 245–269.

Jagannathan, R., and Z. Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Kleibergen, F., and Z. Zhan, 2020, Robust inference for consumption-based asset pricing, *Journal of Finance* 75, 507–550.

Lettau, M., and S.C. Ludvigson, 2001, Resurrecting the (C)CAPM: A cross sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Lustig, H.N., and S.G. Van Nieuwerburgh, 2005, Housing collateral, consumption insurance, and risk premia: An empirical perspective, *Journal of Finance* 60, 1167–1219.

Maio, P., and B.-K. Min, 2026, Revisiting Lettau and Ludvigson (2001): Does *cay* really matter for cross-sectional risk premia? *Critical Finance Review* (forthcoming).

Maio, P., and P. Santa-Clara, 2012, Multifactor models and their consistency with the ICAPM, *Journal of Financial Economics* 106, 586–613.

Newey, W.K., and K.D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Santos, T., and P. Veronesi, 2006, Labor income and predictable stock returns, *Review of Financial Studies* 19, 1–44.

Shanken, J., 1992, On the estimation of beta pricing models, *Review of Financial Studies* 5, 1–34.